# Polygraph

**Special Edition**

**Chart Interpretation**

# The Utah Numerical Scoring System

## Brian G. Bell[1], David C. Raskin[2], Charles R. Honts[3], & John C. Kircher[1]

## Abstract

The Utah method for numerically evaluating polygraph charts is a highly reliable and valid method for scoring specific-incident, comparison-question tests. For respiration, electrodermal activity (skin conductance or skin resistance), relative blood pressure (cardiograph), and peripheral vasomotor activity (finger plethysmograph), a score from +3 to -3 is assigned for each presentation of a relevant question. The reaction to the relevant question is compared to the reaction to a nearby comparison (control) question. A positive score is assigned when the psychophysiological reaction is greater to the comparison question than to the relevant question, a negative score is assigned when the reaction is greater to the relevant question, and a zero is assigned when the responses to the relevant and comparison questions are approximately equal. Scores are based on the criteria described in the present report. Common artifacts that may affect numerical evaluations are discussed, as are limitations of this scoring system.

Key words: comparison-question tests, detection of deception, numerical scoring, polygraph

This report describes the Utah method for numerically evaluating polygraph charts from specific-incident, comparison-question tests. The development of the Utah method was preceded by numerical scoring techniques introduced by Backster (1969) and the U.S. Army (Weaver, 1980; 1985). Although these early scoring systems represented major improvements over global approaches to chart evaluation, many of their scoring rules had no scientific basis and had not been validated by scientific research. The Backster system in particular had been shown to be biased against truthful subjects, (Raskin, 1986) and consisted of many complex scoring rules that made it difficult to evaluate polygraph charts reliably. The Utah system was developed to simplify the scoring process, reduce bias, and improve the accuracy of decisions. It consists of relatively few rules that may be applied with considerable consistency by different numerical evaluators after a brief period of training.

The reliability of the Utah scoring system has been evaluated in several laboratory experiments at the University of Utah. The results of five such studies are summarized in Table 1. On average, the interrater reliability of the Utah system exceeded .90, as measured by the correlations between total numerical scores assigned by two or more evaluators. The percent agreement on decisions exceeded 95% when both numerical evaluators reached a definite decision. Similar reliabilities between raters who use the Utah system have been obtained

in field studies. For example, the interrater correlation was .94 in a field study by Honts and Raskin (1988). These reliabilities far exceed standards of acceptable interrater reliability for psychological tests as established by the American Psychological Association (1985).

**Table 1.  Reliability of the Utah System of Numerical Scoring in Laboratory Studies**

| Study | Agreement on Decisions Between Original Examiner and Independent  Evaluator* | Correlation Between Numerical Scores of Independent Evaluator and Original Examiner |
|---|---|---|
| Podlesny & Raskin  (1978) | 100% | .97 |
| Rovner et al. (1979) | 95% | .97 |
| Kircher & Raskin (1988) | 99% | .97 |
| Honts et al. (1994) | 96% | .92 |
| Horowitz et al. (1997) | .98** | .92 |

* Includes only cases in which both examiners made a decision (excludes inconclusives)

** Only Kappa was reported in this study.

The validity of the Utah system of numerical evaluation has also been established. Table 2 presents decision accuracies from several laboratory experiments. Excluding inconclusive outcomes, the overall percentage of correct decisions was 91% for guilty subjects and was 89% for innocent subjects.

The results from field studies with the Utah system are consistent with those reported in Table 2 (Honts & Raskin, 1988; Raskin, 1976; Raskin, Kircher, Honts, & Horowitz, 1988). In one field study, two numerical evaluators independently evaluated the polygraph charts using the Utah system (Raskin, 1976). Their decisions were 100% correct for both guilty and innocent suspects. In another study, decisions were 92% correct for guilty suspects and 100% correct for innocent suspects (Honts & Raskin, 1988).

## Overview of the Utah Scoring System

The Utah scoring system, when used with the probable-lie and directed-lie comparison question tests, assigns numerical scores by assessing differences between relevant and comparison questions. Scores are assigned on a 7-point scale that ranges from -3 to +3. The reaction to a relevant question is compared to the reaction produced by a temporally adjacent, comparison question. If a relevant question was presented between two comparison questions, its reaction is compared to the reaction to the comparison question that produced the stronger physiological response.

For each channel, the relative size of the reactions to the comparison and relevant questions is evaluated and quantified. Positive scores are assigned when the physiological reaction to the comparison question was greater than the reaction to the relevant question.  Negative scores are assigned when the reaction to the relevant question was greater, and zero is assigned when reactions to the relevant and comparison questions are not noticeably different. In general, a noticeable difference between the reactions to the comparison and relevant questions is assigned a score of 1. A strong, clear difference between

**Table 2. Validity of the Utah System of Numerical Scoring in Laboratory Studies**

| | Guilty | | | | | Innocent | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Study | N | Correct | Incorrect | Inconclusive | % Correct* | N | Correct | Incorrect | Inconclusive | % Correct* |
| Raskin & Hare (1978) | 24 | 88% | 0% | 12% | 100% | 24 | 75% | 4% | 21% | 95% |
| Podlesny & Raskin (1978) | 20 | 70% | 15% | 15% | 82% | 20 | 90% | 5% | 5% | 95% |
| Rovner et al. (1979)[a] | 24 | 88% | 0% | 12% | 100% | 24 | 88% | 8% | 4% | 92% |
| Kircher & Raskin (1988) | 50 | 88% | 6% | 6% | 94% | 50 | 86% | 6% | 8% | 93% |
| Honts et al. (1994) | 20 | 70% | 20% | 10% | 78% | 20 | 75% | 10% | 15% | 88% |
| Horowitz et al. (1997)[b] | 15 | 53% | 20% | 27% | 73% | 15 | 80% | 13% | 7% | 86% |

* The percent correct was calculated by dividing the number correct by the sum of the number correct and the number incorrect.

[a] Excludes 24 countermeasure-trained subjects.
[b] Excludes 90 subjects given relevant-irrelevant and directed lie tests.

the reactions is assigned a 2. A score of 3 is assigned when there is a dramatic difference between the reactions to the two questions, the tracing is stable, and the stronger response is the largest on the chart for that physiological measure.

In a single-issue test, all relevant questions must be answered truthfully or all must be answered deceptively. In this case, the scores for all presentations of relevant questions are summed. The subject is reported as deceptive if the total score is -6 or lower, truthful if the total is +6 or higher, and inconclusive if the total is between -6 and +6.

For mixed issue tests, such as the Modified General Question Test, some relevant questions can be answered truthfully while others can be answered deceptively. In this case, a separate total is obtained for each relevant question. When the total score for a single relevant question is -3 or lower, the subject's answer to that question is considered deceptive. When the total score for a single relevant question is +3 or higher, the subject's answer is considered truthful. When the total score for a question is between -3 and +3, the outcome is considered inconclusive. However, if the total score for all relevant questions combined is at least +6 or -6, and the total scores for each relevant question are in the same direction (all positive or all negative), the subject is considered truthful or deceptive, respectively, to each relevant question.

# Scoring Criteria

A total of ten scoring criteria are used to assess the relative strength of physiological reactions to relevant and comparison questions. The criteria change depending on the physiological measure being evaluated. Scores are assigned to respiration, electrodermal, cardiograph, and finger plethysmograph channels.

## Respiration

For a given relevant question, changes in respiration are evaluated first because deep breaths may affect how other channels are evaluated. In general, a reaction to a question is indicated by suppressed respiratory activity. The greater the suppression, the stronger the reaction. Suppression is indicated primarily by a reduction in the amplitude of at least two successive respiration cycles following question onset and brief periods of apnea (cessation of breathing). A rise in the respiration baseline, as indicated by a rise in the bottoms of at least two respiration cycles, is another criterion for scoring a reaction. An increase in cycle time (slowing of respiration rate) is also a criterion but is less heavily weighted than changes in amplitude, apnea, and baseline increase. Increases in respiratory activity, such as increased amplitude, speeding of respiration, and drops in respiration baseline, are not indications of a reaction and are not criteria for scoring.

Although thoracic and abdominal respiration are recorded on separate channels of the polygraph, only one numerical score is assigned that is based on a composite of both channels. Respiration reactions to the comparison and relevant questions are evaluated by noting the combined amount of reaction in both respiration channels for the relevant question and for the comparison question. A single numerical score is then assigned based on the difference in the composite reactions to the relevant and comparison questions. Thus, the numerical score for one relevant question may be based on observed changes in thoracic respiration, abdominal respiration, or a composite of both, depending on the total amount of change observed in the two channels.

## Electrodermal Activity

The electrodermal channel is evaluated next. Numerical scores for electrodermal activity are based mainly on changes in peak amplitude. The amplitude of a reaction is defined as the greatest difference between any low point and subsequent high point that occurs within the scoring window (described below). Amplitude may be measured by using the numerical scoring subprogram in the Stoelting Computerized Polygraph System (CPS) or a similar system. If only printed or inked charts are available, the amplitude is measured with a ruler to the nearest 0.5 millimeter. For each relevant question, a score of 1 is assigned if the amplitude of the reaction to the relevant or comparison question is twice as large as the amplitude of the reaction to the other question. A score of 2 is assigned when the amplitude of the reaction is three times as

large, and a score of 3 is assigned when the amplitude is four times as large. However, a score of 3 may be assigned only when the baseline is stable and the reaction is the largest on the chart. The baseline is considered unstable if there are many nonspecific electrodermal responses on the chart. Under those conditions, a score of 3 cannot be assigned.

The duration of the electrodermal reaction and its complexity (the number of waves or fluctuations that occur within the scoring window) are also considered when scoring the electrodermal channel. Reactions that have clearly longer duration or greater complexity may increase the score from 0 to 1 or from 1 to 2. The larger score may be assigned if the ratio of the amplitudes is at least 1.5:1 or 2.5:1, and the larger reaction has longer duration and/or is more complex. However, a score of 1 or 2 cannot be assigned if the reactions differ only in duration and/or complexity, and these criteria are not used to assign a score of 3.

**Cardiograph**

For the cardiograph channel, reactions are measured as rises in the baseline. The numerical score is based primarily on the largest rise in the baseline that occurs within the scoring window. Again, use of a computer-scoring algorithm, such as the CPS, or a ruler is recommended for making measurements of increases in the baseline. A minimum ratio of 1.5 to 1 is required for a score of 1. Measurements of baseline increases are made on the diastolic side of the waveform because the diastolic points show greater change than the systolic points and are easier to see. However, increases in the systolic points may be used if it is unclear whether to assign a 0 or 1 or to assign a 1 or 2 based on the diastolic points. The duration of the response is also considered. The rules described above for electrodermal reactions apply to the cardiograph; reactions with longer duration may increase the numerical score from 0 to 1 or from 1 to 2.

**Finger Plethysmograph**

Peripheral vasomotor activity is measured from a photoplethysmograph attached to the tip of the finger. Constriction of blood vessels in the finger produces a reduction (constriction) in the amplitude of finger pulses. Numerical scores are based on the duration and magnitude of reductions in finger pulse amplitude. Responses of longer duration and/or magnitude are assigned larger numerical scores. Unlike the cardiograph and electrodermal channels, scores of 1 or 2 may be assigned to this response system when there is little or no difference in the reduction of pulse amplitude, but there is a clear difference in the duration of the reactions.

## Scoring Windows

For all channels, the response is not scored unless it begins after question onset. However, the minimum latency for a response varies depending on the physiological measure. Respiratory and cardiograph reactions may be scored if they begin immediately after question onset. Electrodermal reactions are scored only if they begin at least 0.5 seconds after question onset, and finger pulse reactions are scored only if they begin at least 2 seconds after question onset. If a reaction begins prior to the minimum latency, a point of inflection or clear increase in slope that occurs after the minimum latency may be considered the beginning of the reaction. For all physiological measures, the reaction must begin no later than 5 seconds after the subject's answer, unless the subject characteristically has reactions that begin 5 to 8 seconds after answering. Such delayed reactions should be scored conservatively. Reactions that begin outside these scoring windows are not scored. The duration of a reaction that begins within the scoring window may be considered up to 20 seconds following question onset.

## Distributions of Numerical Scores

We measured the frequency with which we observed physiological changes that met the criteria for 'noticeable', 'clear', and 'dramatic' differences in a random sample of 25 innocent and 25 guilty subjects who participated in a previous experiment (Kircher & Raskin, 1988). Guilty subjects had committed a mock theft, innocent subject did not commit the theft, and all subjects had been promised a substantial monetary bonus if they could convince the polygraph examiner that they were innocent of the crime. The

charts were scored by the second author (DCR) who had no contact with the subjects and was unaware of the subjects' guilt or innocence. The overall accuracy of decisions was 95% for guilty subjects and 96% for innocent subjects. For each physiological component, each of three relevant questions was scored against the probable-lie comparison question that immediately preceded it on each of three charts. This provided a sample of 450 numerical scores for each physiological component (3 relevant questions X 3 charts X 50 subjects). The absolute values of these scores are presented in Figure 1 for each physiological measure.

**Figure 1.**
**Distributions of numerical scores for respiration, electrodermal, cardiograph, and finger plethysmograph.**
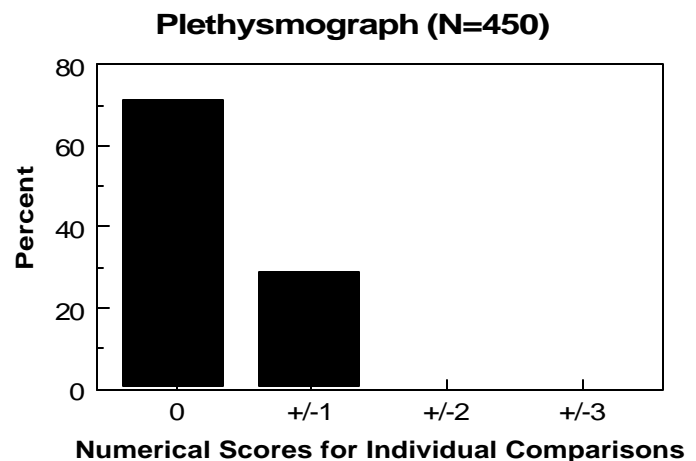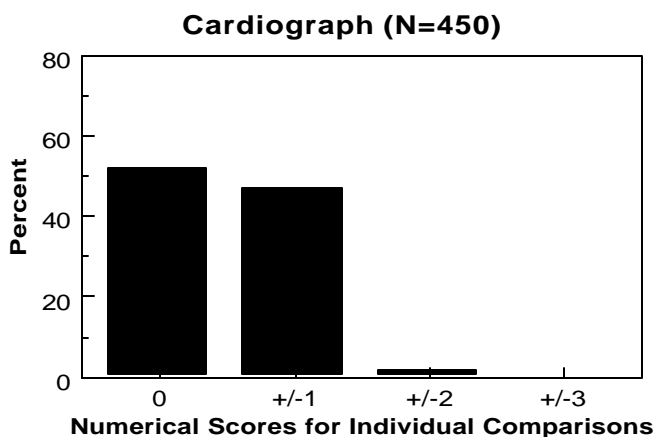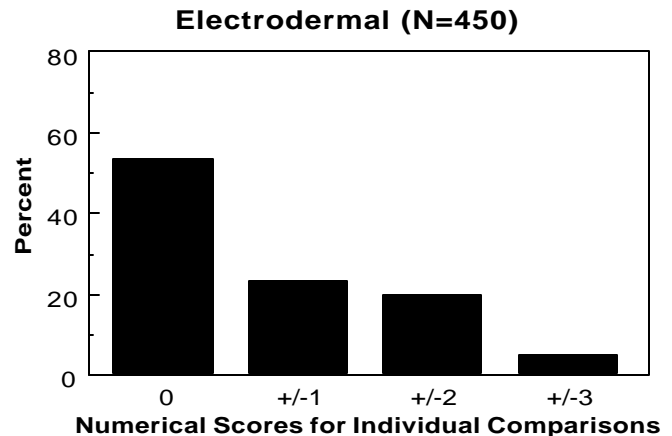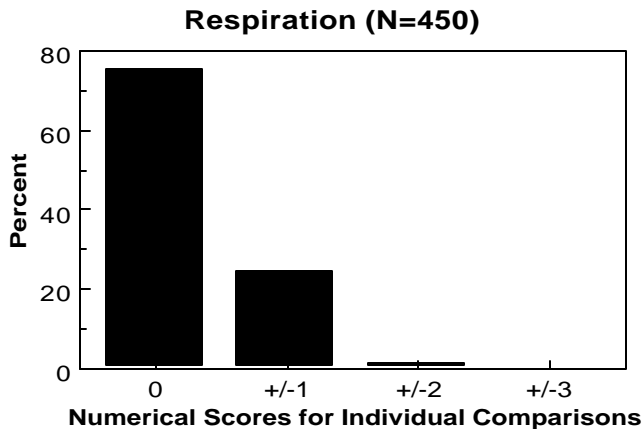
As shown in Figure 1, numerical scores of '0' were assigned considerably more often than any other value. It also may be seen that the frequency with which different numerical scores were assigned depended on the physiological measure. More than 70% of the numerical scores were '0' for the respiration and plethysmograph channels, whereas approximately 50% of the scores were '0' for skin conductance and cardiograph channels. These results suggest that, on average, decisions will be based largely on the numerical scores assigned to the electrodermal and cardiograph channels.

Although electrodermal and cardiograph channels had about the same number of scorable (nonzero) differences, numerical scores of 2 or 3 were more common for the electrodermal channel. Since more scores of 2 and 3 were assigned to the electrodermal channel, it had more influence than the cardiograph on the total numerical scores and the outcomes of the polygraph tests. Together, the data indicate that the Utah scoring rules give greater weight to electrodermal reactions than to cardiovascular, respiration, or plethysmograph reactions. The relative weights given the four physiological measures by the Utah scoring rules are remarkably consistent with the optimal combinations of weighted physiological measures that are generated by our Computerized Polygraph System (CPS) (Kircher & Raskin, 1991).

It should be noted that scores of 3 were extremely rare. The rules allow for the assignment of scores of 3 to any channel, but in this sample of 450 comparisons, not a single score of 3 was assigned to respiration, cardiograph, or plethysmograph channels.

## Artifacts

The quality of the tracing is considered when assigning scores. Artifacts, such as deep breaths, coughs, movements, and physiological abnormalities, affect how scores are assigned. To minimize the occurrence of artifacts, we instruct examinees to avoid movements and to breathe normally while the recordings are made. Nevertheless, artifacts may render the tracings unscorable.

If a deep breath occurs shortly before question onset, respiration should not be scored. If the deep breath is accompanied by physiological changes in other channels, the other channels may or may not be scored. If the reaction in the other channel began before the deep breath, then the portion preceding the deep breath may be used in scoring if it is larger than the reaction to the question to which it is compared. If it is smaller and is to a comparison question, then another comparison question may be used. The evaluator should also examine all of the charts for the subject and locate any other places where a deep breath occurred, especially at points where no question had been asked. If there is a similar physiological change at this point, then the reaction following the deep breath must not be used for scoring. If there is no reaction following the deep breath, then the reaction may be scored, but it should be scored conservatively.

If movements distort more than two successive pulses in the cardiovascular channels after question onset, the cardiovascular changes that occur after the movement should not be scored. If there is a reaction that precedes the artifact, it may be used for scoring if it is larger than the reaction to which it is compared. However, if only one or two pulses are distorted, it is usually possible to visually interpolate across the artifact and infer what the reaction would have been if the movement had not occurred. If multiple artifacts occur within the scoring window, it is usually not possible to score the response.

Physiological abnormalities, such as premature ventricular contractions (PVCs), may also render the cardiovascular reaction unscorable if they occur in the scoring window. PVCs are contractions of the left ventricle that occur before the left atrium has contracted and filled the left ventricle, causing very little blood to be pumped into the aorta. This is followed by a relatively long pause before the next ventricular contraction. During this pause, the drop in blood pressure produces a distinct downward deflection in the cardiovascular tracing. If two or three PVCs occur within the scoring window, the signal is usually so distorted that it is not possible to score it. However, a cardiovascular reaction

that occurred before the PVC may be scored. It is usually also possible to score the reaction if it contains only one PVC, although the subsequent rise in the tracing that is the recovery from the PVC should not be scored as a reaction.

# Limitations

The research that supports the use of the Utah system of numerical scoring has been limited to specific-incident examinations. It has not been validated for employment screening or periodic testing of employees with access to sensitive information. Furthermore, most of our research has focused on the probable-lie test. Use of the numerical scoring system with the directed-lie has also been validated (Honts & Raskin, 1988; Horowitz et al., 1997), but the relevant-irrelevant and other types of tests have received almost no attention.

Most of the laboratory research with the Utah system has used a single-issue test that contains three repetitions of neutral, comparison, and relevant questions in the question sequence. Other question sequences or mixed-issue tests have not been tested extensively in our laboratory, although two field studies included at least one numerical evaluator who used the Utah system with the Modified General Question Test format (Raskin, 1976; Raskin et. al., 1988), and the accuracy of those decisions was comparable to those we have observed in our laboratory experiments. Therefore, there is evidence that the validity of the Utah scoring technique generalizes across similar test formats.

# References

American Psychological Association (1985). *Standards for educational and psychological testing*. Washington, D.C.: American Psychological Association.

Backster, C. (1969). *Technique fundamentals of the tri-zone polygraph test*. New York: Backster Research Foundation.

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

Honts, C. R., Raskin, D. C. & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, 79, 252-259.

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

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

Kircher, J. C. & Raskin, D. C. (1991-98). Computerized Polygraph System (CPS) Versions 1.00-2.20. Scientific Assessment Technologies, Inc., 2532 Chadwick Street, Salt Lake City, UT 84106

Podlesny, J. A. & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, 15, 344-359.

Raskin, D. C., (1976). Reliability of chart interpretation and sources of errors in polygraph examinations. Report to the National Institute of Law Enforcement and Criminal Justice (Contract 75-NI-99-0001). Salt Lake City: University of Utah, Department of Psychology.

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

Raskin, D. C. (1986). The polygraph in 1986: Scientific, political, and legal issues surrounding application and acceptance of polygraph evidence. *Utah Law Review*, 1986, 29-74.

Raskin, D. C., Kircher, J. C., Honts, C. R., & Horowitz, S. W. (1988). A study of the validity of polygraph examinations in criminal investigation. Final report to the National Institute of Justice (Grant No. 85-IJ-CX-0040). Salt Lake City: University of Utah, Department of Psychology.

Rovner, L. I., Raskin, D.C., & Kircher, J. C. (1979). Effects of information and practice on detection of deception. *Psychophysiology*, 16, 197-198. (Abstract).

Weaver, R. S. (1980). The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. *Polygraph*, 9, 94-108.

Weaver, R. S. (1985). Effects of differing numerical chart evaluation systems on polygraph examination results. *Polygraph*, 14, 34-41.

**Table 2. Validity of the Utah System of Numerical Scoring in Laboratory Studies**

| Study | | Guilty | | | | | Innocent | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | N | Correct | Incorrect | Inconclusive | % Correct* | N | Correct | Incorrect | Inconclusive | % Correct* |
| Raskin & Hare (1978) | 24 | 88% | 0% | 12% | 100% | 24 | 75% | 4% | 21% | 95% |
| Podlesny & Raskin (1978) | 20 | 70% | 15% | 15% | 82% | 20 | 90% | 5% | 5% | 95% |
| Rovner et al. (1979)[a] | 24 | 88% | 0% | 12% | 100% | 24 | 88% | 8% | 4% | 92% |
| Kircher & Raskin (1988) | 50 | 88% | 6% | 6% | 94% | 50 | 86% | 6% | 8% | 93% |
| Honts et al. (1994) | 20 | 70% | 20% | 10% | 78% | 20 | 75% | 10% | 15% | 88% |
| Horowitz et al. (1997)[b] | 15 | 53% | 20% | 27% | 73% | 15 | 80% | 13% | 7% | 86% |

* The percent correct was calculated by dividing the number correct by the sum of the number correct and the number incorrect.

[a] Excludes 24 countermeasure-trained subjects.
[b] Excludes 90 subjects given relevant-irrelevant and directed lie tests.

# Manually Scoring Polygraph Charts Utilizing the Seven-Position Numerical Analysis Scale at the Department of Defense Polygraph Institute

## Jimmie Swinford

This documentation sets forth information on how the seven-position numerical analysis scale is taught/utilized at the Department of Defense Polygraph Institute (DoDPI). It lists the criteria currently used at DoDPI for manually scoring all three physiological parameters recorded on polygraph charts during the conduct of a psychophysiological detection of deception (PDD) examination. Finally, this document identifies the DoDPI scoring procedures and methods utilized for assigning values, ranging from +1, +2, +3 and 0 to -1, -2 and -3, for the respiratory, electrodermal activity and cardiovascular tracings of a comparison question test (CQT) format.

## Background Information

During the manual scoring process of test data analysis, the PDD examiner completes several steps before actually assigning values to recorded physiological responses. Initially, the examiner reviews the physiological data on the charts in an attempt to determine what an examinee's physiological tracing looks like while in a state of homeostasis or equilibrium. During this process, the PDD examiner also determines if there is any unwanted noise on the signal of interest when a scoreable question (comparison or relevant) was asked on the chart. If unwanted noise on the signal of interest is present, then a portion, or all, of the physiological activity recorded during that question may be unusable during the scoring process. However, just because unwanted noise appears in one of the physiological parameters does not mean that the other parameters in that analysis spot cannot be scored. For example, if an artifact appears in the respiratory tracings, they are not scored;

but the electrodermal and cardiovascular tracings may be scored if they have not been affected by the artifacted respiratory tracings. However, if all of the recorded physiological parameters appear to have been affected by unwanted noise, that particular question cannot be utilized during the scoring process.

Next, the examiner scores the charts using the appropriate analysis spots for that particular CQT format. After eliminating chart excerpts containing artifacts, recovery, and other unwanted noise, the examiner compares responses of the relevant question(s) to responses of the applicable comparison question(s) by individual physiological recorded parameters. Responses are only scored when there is no unwanted noise on the signal of interest at the time the stimulus was presented, and if the responses began within the response onset window (with latency exceptions). Depending on the type of response, respiratory responses end when recovery starts, the tracing returns to the prestimulus baseline or homeostasis returns. Electrodermal and cardiovascular responses end when the tracing either: (1) returns to the prestimulus baseline or (2) stabilizes at a new tonic level.

During the scoring process, physiological tracings in an analysis spot with no responses or comparable responses are assigned a value of "0". If there is any form of unwanted noise on the signal of interest that prevents that signal from being scored, a "⊘" (zero with a line through it) is placed on the score sheet. In the seven-position scale, magnitudes of plus and minus values, ranging from a +1, +2, or +3 to a -1, -2 or -3, are awarded according to specific analysis processes for each physiological activity.

---

The author is an instructor for the Department of Defense Polygraph Institute.

After values are assigned for each analysis spot for all the PDD charts collected for an examination, these values are tallied and decisions are rendered according to the specific cut-off scores necessary for a particular CQT format. Depending on the type of CQT or PDD examination, the examiner may render the following decisions:

| **Specific Issue CQT** | **Screening CQT** |
|---|---|
| 1. No Opinion | No Opinion |
| 2. No Deception Indicated (NDI) | No Significant Responses (NSR) |
| 3. Deception Indicated (DI) | Significant Responses (SR) |

## Types of Physiological Criteria Utilized in the Scoring Process

**Respiratory Tracing:**

This is the display of physiological activity indicative of an examinee's breathing pattern that is recorded by a pneumograph component. The respiratory tracing consists of inhalation and exhalation cycles. An examinee's breathing pattern and rate may vary due to their physical conditioning. Normally, during the data collection phase, the examiner will attach two recording sensors (pneumograph chest assembly) to the examinee via some type of device. Typically, the pneumograph chest assembly consists of a convoluted tube, return mechanism, anti-roll bars, beaded chain or velcro strips and rubber tubing for connection to the computer sensor box or analog instrument. One pneumograph chest assembly will be placed around the examinee's upper body area to record the thoracic breathing pattern. A second pneumograph chest assembly will be placed around the lower abdomen area to record the abdominal breathing pattern. When scoring the respiratory tracings, the PDD examiner will encounter the following five main categories of responses: (1) Changes in rate, (2) Changes in amplitude, (3) Change of baseline, (4) Loss of baseline and (5) Apnea. These five main categories of responses consist of the following 12 scoreable criteria:

**RESPIRATION RATE CHANGES**

**Criteria 1: Respiration Rate Decrease**



I    I - 4
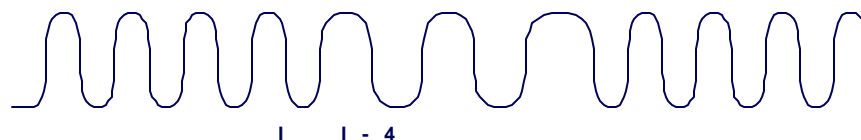
**RESPIRATION RATE CHANGES**

**Criteria 2: Respiration Rate Increase**

I I - 5
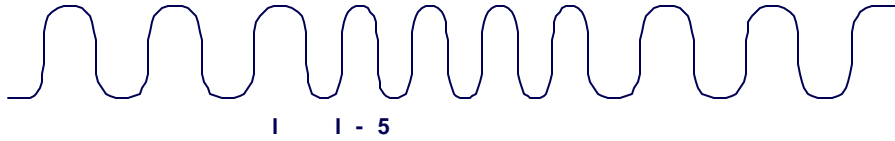
**RESPIRATION RATE CHANGES**

**Criteria 3: Respiration Inhalation/
Exhalation Ratio Change**

I I - 3
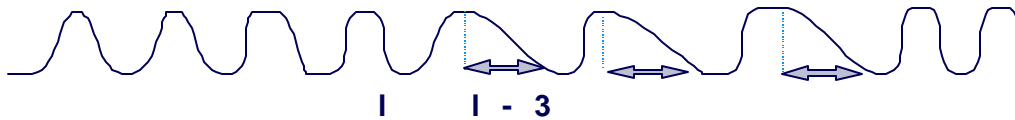
**RESPIRATION AMPLITUDE CHANGES**

**Criteria 4: Respiration Amplitude Increase**

I I - 7
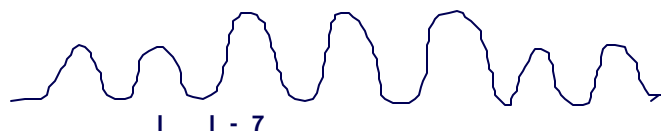
## RESPIRATION AMPLITUDE CHANGES

### Criteria 5: Respiration Amplitude Decrease/Suppression

**1 1 - 5**

## RESPIRATION AMPLITUDE CHANGES

### Criteria 6: Progressive Increase Followed by A Decrease

**I I-3**

## RESPIRATION AMPLITUDE CHANGES

### Criteria 7: Progressive Increase and Return to Homeostasis

**I I-5**
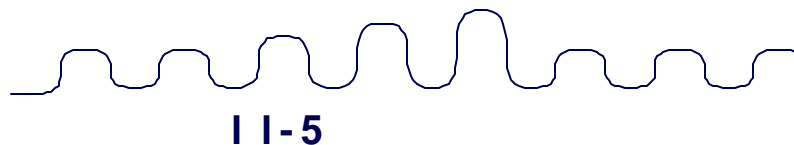
# RESPIRATION AMPLITUDE CHANGES

## Criteria 8: Progressive Decrease and Return to Homeostasis

I  I - 5

# RESPIRATION BASELINE CHANGE

## Criteria 9: Respiration Baseline Change (Temporary)

I  I- 5

I     I - 7
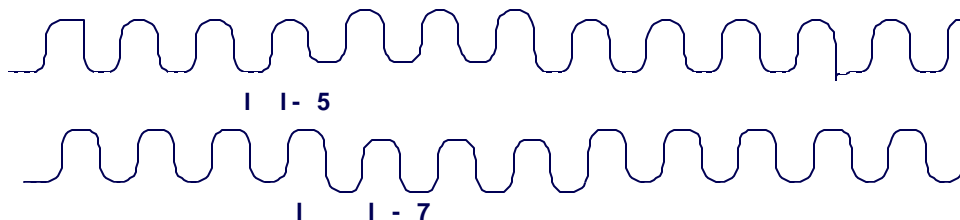
# RESPIRATION BASELINE LOSS

## Criteria 10: Respiration Baseline Loss (Permanent)

I     I - 4

I     I - 4

**R E S P I R A T I O N   A P N E A**

**C r i t e r i a   1 1 :   H o l d i n g**

I    I - 5

**C r i t e r i a   1 2 :   B l o c k i n g**

I    I - 7
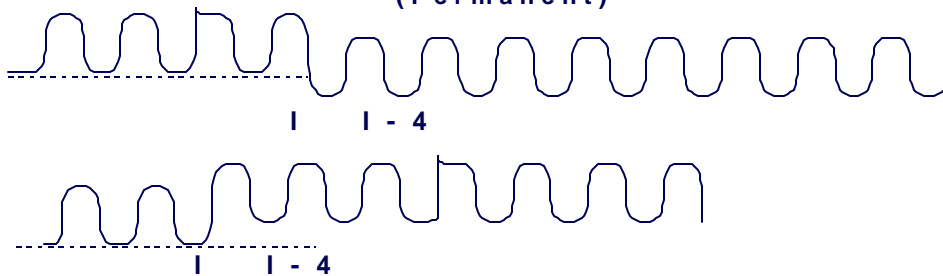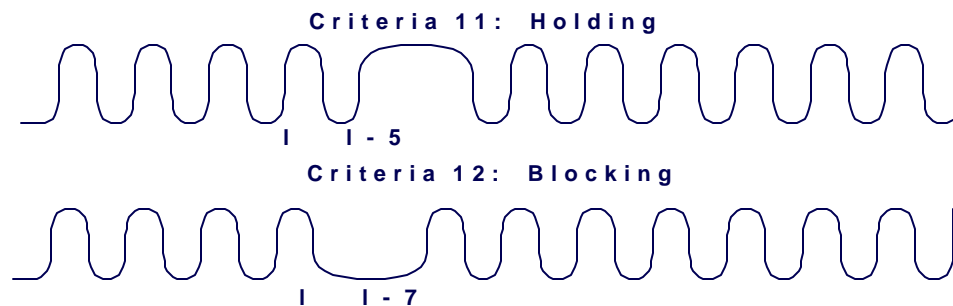
## Scoring the Respiratory Tracing

During the process of scoring the respiratory tracings, the examiner must accomplish several steps before being able to score physiological data. First, the examiner must decide if a physiological change (response) has occurred in a timely manner. To be considered timely, a response should occur within the response onset window (typically from stimulus onset until the examinee's answer or first complete respiratory cycle after the answer to a reviewed test question) (latency exceptions excluded). Even if a physiological change has occurred in a timely manner, the examiner must also decide if there was any type of unwanted noise (artifact, recovery, etc.) on the signal of interest at the time the stimulus (question) was applied. Finally, the examiner must then determine what constitutes a response to the stimulus and when compensatory action (recovery) begins for this response. In respiratory tracings, if there is any type of response to a presented stimulus, there will generally be some form of recovery (compensatory action) that will occur before an examinee's respiratory pattern returns to equilibrium or homeostasis. In scoring respiratory tracings, this is one of the more difficult determinations for a basic examiner student to make. It is imperative to make these distinctions as response can only be scored against response. An examiner cannot score response against any form of unwanted noise (i.e., artifact or recovery) or vice versa.

After making the above determinations, the examiner will award equal value for different response criteria. For example, if a comparison question has a change of baseline response and the relevant question has a decrease in amplitude response, then equal value will be assigned for these criteria in the scoring process. In this particular situation, a score of "0" would be appropriate unless one response lasted longer (more duration). Typically, in the seven-position scale, the examiner will utilize the following guidelines based on the number of observed physiological response criteria:

## Question Type

| Comparison | Relevant | Assigned Value |
|---|---|---|
| 0 criteria | 0 criteria | 0 |
| 1 criterion | 1 criterion | 0 (unless duration is a factor) |
| 1 criterion | 0 criteria | +1 |
| 1 criterion | multiple criteria | - 1 |
| Multiple criteria | 0 criteria | +2 |
| 0 criteria | multiple criteria | - 2 |
| 0 criteria | dramatically better | - 3 |

In deciding whether a physiological response for one of the comparative questions has multiple criteria present, the examiner cannot decide that a specific response has multiple criteria if the observed response would automatically cause another criterion to be present. For example, an increase or decrease in amplitude or apnea will generally cause a change in rate to occur. Even though there may appear to be multiple criteria present, it cannot be considered multiple criteria for scoring purposes. Typically, the following singular criterion will constitute multiple criteria when observed in respiratory tracings:

| Singular Criterion | | Multiple Criteria |
|---|---|---|
| Amplitude and/or Rate Changes | + | Apnea or baseline changes |
| Change/Loss of Baseline | + | Apnea or amplitude/rate change |
| Apnea | + | Change or loss of baseline |

In the seven-position scale, duration will generally allow assigning a value of only +/-1 (no more). Typically, duration is a factor only in those situations where one singular response criterion is compared against another singular response criterion. In these instances, if one response lasts longer (more duration), then a value of +/- 1 may be assigned for duration. However, if one comparative question has a singular response while the other comparative response exhibits multiple criteria, then duration is generally not a factor. As a rule, multiple response criteria will cause more response duration than a singular criterion.

In scoring the respiratory tracings, the examiner will rarely assign values higher than a "0" or +/-1. Occasionally, an examiner may assign a value of +/-2 based on physiological criteria observed in the comparative questions. Rarely, examiners will assign a value of +/-3.

To assign a +/-3 value, the physiological response in one question must be so dramatically better (multiple criteria for an extended period) than the response it is being compared against.

**Electrodermal Activity (EDA) Tracing**

The EDA tracing is the display of an examinee's physiological patterns of either skin resistance or skin conductance of an exosomatic recording obtained with a galvanograph component. During the data collection phase, the examiner will attach a sensor to the examinee called the EDA fingerplate electrode assembly. Normally, the fingerplate electrode assembly consists of two stainless steel plates, with velcro straps, and shielded cable for connection to the computer sensor box or analog instrument. Ideally, the fingerplates will be placed on two fingers of an examinee's non-dominant hand. During the data collection phase, once the sensor is attached to an examinee, an external (exosomatic) electrical signal is applied to the examinee's skin. The amount of resistance (skin resistance) that is encountered when the signal is applied or how freely the electrical signal travels (skin conductance) is recorded by the galvanograph component. The following criteria will be utilized when scoring an EDA tracing:

**E L E C T R O D E R M A L   T R A C I N G**

**C r i t e r i a   1 :   A m p l i t u d e   C h a n g e**

I     I - 4          I     I - 5          I     I - 6

**E L E C T R O D E R M A L   T R A C I N G**

**C r i t e r i a   2 :   C o m p l e x   R e s p o n s e**

I   I -5                    I   I -6
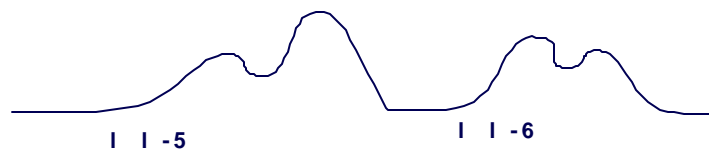
# ELECTRODERMAL TRACING

## Criteria 3: Response Duration



I I -5
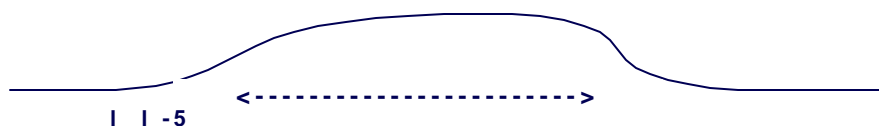
## Scoring the EDA Tracing

In the seven-position numerical analysis scale, assigned EDA values are based on a ratio scale. This is the only physiological parameter where a ratio method is used in scoring the responses. Because the EDA component is generally the most responsive of all the physiological parameters recorded by the polygraph, the unit of measurement for determining ratio is generally a vertical chart division. If the compared responses are like responses (i. e., both amplitude or both complex) and they are equal in amplitude and duration, then a score of "0" is assigned and the ratio method is unnecessary. However, if one of the compared responses has significantly more amplitude than the other, the examiner will utilize the ratio method in assigning values. To determine the ratio for appropriate responses, the smaller response is divided into the larger response (i. e., six chart divisions divided by two chart divisions would be a ratio of 3:1). Once the ratio of responses is determined, then values are assigned based on the following ratio formula:

| RATIO | SCORE |
|-------|-------|
| 4:1 | +3/-3 |
| 3:1 | +2/-2 |
| 2:1 | +1/-1 |
| 1:1 | 0 |

As an example, if there is a comparison question amplitude response of three chart divisions and a relevant question amplitude response of seven chart divisions, then the ratio would be 2.5:1 (7 divided by 3 equals 2.5). To determine the value for this response ratio, the examiner would utilize the ratio scale. In this case, the value would be a -1 since the larger response occurred in the relevant question. When using the ratio scale, the response must be at or above the appropriate ratio level to assign a specific value, (i. e., to assign a value of +/- 2, the response ratio must be at the 3:1 ratio level, etc.) In the above example, the response ratio was 2.5:1. Accordingly, since it was not at the 3:1 ratio level, the assigned score is a -1.



**Vs**

**(Comparison)**
**3 Chart Divisions**

**(Relevant)**
**7 Chart Divisions** =

**Ratio of 2.5:1 = Score of -1**
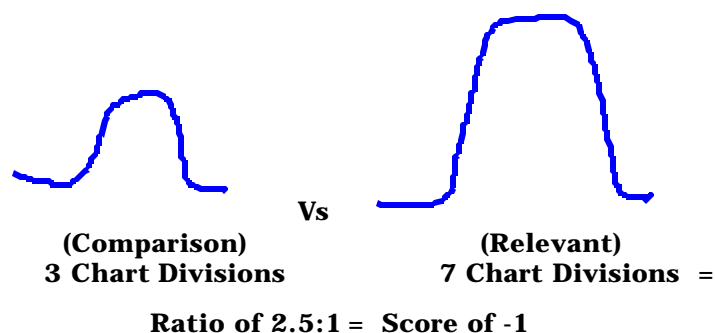
There is one exception to the above rule and that is when the ratio is less than 2:1. For instance, if there is a comparison question amplitude response of three chart divisions and a relevant question amplitude response of two chart divisions, the ratio is 1.5:1. If the ratio method were utilized, the assigned value would be a "0" since the ratio is less than 2:1. However, since the comparison response is 1.5 times larger in amplitude than the relevant question response, it cannot be ignored. As such, when the ratio is less than 2:1, the concept of "bigger is better" applies for the initial +1/-1 . In this instance, a score of +1 would be assigned, since the comparison question response is 1.5 times larger in amplitude than the relevant question. The "bigger is better" concept applies only to the initial +1 or -1. Once the compared responses reach a ratio of 2:1, this concept no longer applies.

**(Comparison)**
**3 Chart Divisions**

**Vs**

**(Relevant)**
**2 Chart Divisions**

**Ratio of 1.5:1    =    Score of +1        (Bigger is Better)**
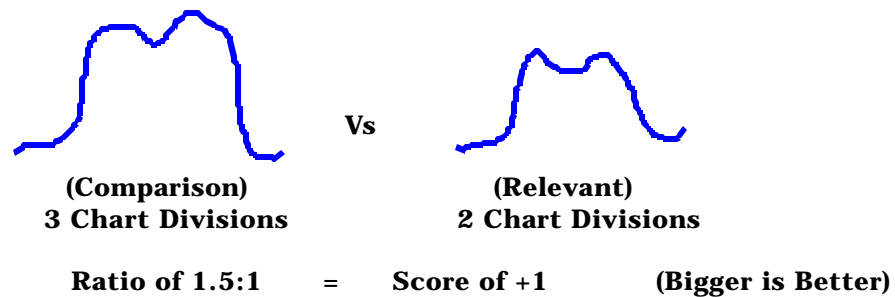
As indicated, all of the above concepts apply to "like responses", i. e., amplitude versus amplitude or complexity versus complexity. It would be great if examinees provided "like responses" all the time. However, as we all know, this does not always occur. As such, there must be provisions for comparing "unlike responses" (amplitude versus complexity) using the ratio method. In certain instances, complexity will allow assigning a value of +1 or -1 (no more).

When comparing an amplitude response to a complex response, complexity will only be considered when both responses are equal in amplitude or the ratio is less than 2:1. Once the ratio factor reaches a scale of 2:1, complexity is no longer a consideration and values are assigned for amplitude ratio. As an example, when there is a comparison question with an amplitude response of six chart divisions and a relevant question with a complex response of six chart divisions, the assigned value would be a -1.

**(Comparison)**
**6 Chart Divisions**

**Vs**

**(Relevant)**
**6 Chart Divisions**
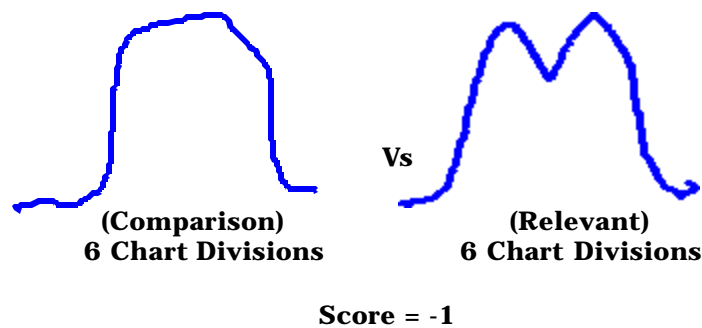
**Score = -1**

If the unlike compared responses are less than a 2:1 ratio in amplitude and the smaller response is complex, then the value will be a "0". If the larger response is complex, then complexity is no longer a consideration. For example:
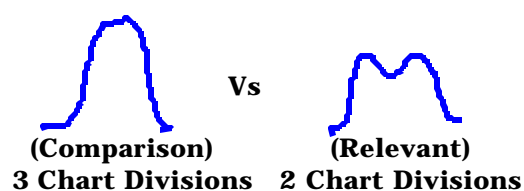
**Vs**

**(Comparison)** **(Relevant)**
**3 Chart Divisions** **2 Chart Divisions**

**Score = 0**

**Vs**

**(Comparison)** **(Relevant)**
**3 Chart Divisions** **2 Chart Divisions**
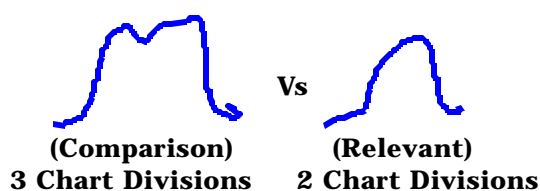
**Score = +1**

Once the comparative ratio reaches a scale of 2:1, complexity is no longer a factor for consideration. Values are then assigned for amplitude responses according to the ratio scale. For example:

**Vs**

**(Comparison)** **(Relevant)**
**4 Chart Divisions** **2 Chart Divisions**
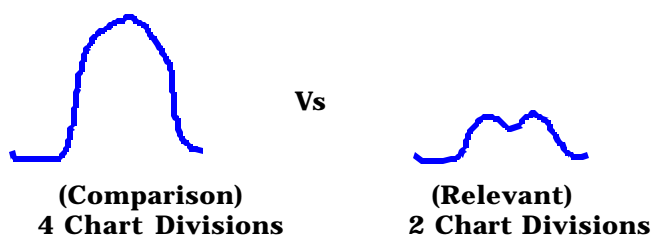
**Ratio of 2:1 = Score of +1**

**Other EDA Scoring Considerations:**

Duration in "like responses" will allow assigning a value of +1 or -1 only (no more). Duration is not a consideration in "unlike responses" (i. e., amplitude versus complexity) as complexity will generally have more duration just by the nature of the type of response. Therefore, in assigning a value for duration, the responses must be similar responses (amplitude versus amplitude or complexity versus complexity) and they must be equal or equivalent in amplitude:

**Vs**                         **or**                         **Vs**

**(Comparison)** **(Relevant)** **(Comparison)** **(Relevant)**
**2 Chart Divisions** **2 Chart Divisions** **2 Chart Divisions** **1.5 Chart Divisions**
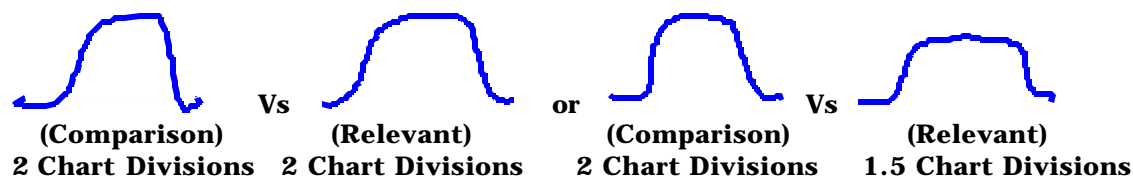
**Score = -1 (Duration)**                         **Score = 0**

**Something Against Nothing Concept**

If one of the questions (comparison or relevant) has a physiological response, while the other question in the same analysis spot has no response (preventing utilization of the ratio method), then the unit of measurement (chart division) is utilized for assigning values.

By accomplishing this, an examiner can always be consistent in assigning values for the responses being scored. For instance, if there are no responses in the comparison question, while the relevant question has two chart divisions of amplitude response, a value of -2 may be assigned (since the unit of measure is a chart division). Likewise, if the

reverse were true, then a value of +2 would have to be assigned to be consistent in

applying this principle.

**Vs**

**(Comparison)**
**(Nothing)**

**(Relevant)**
**2 Chart Divisions**

**Score = -2**
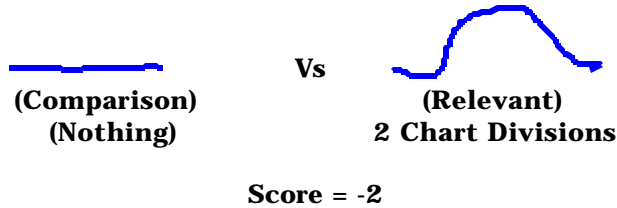
### Cardiovascular Tracing

This tracing displays the physiological patterns of an examinee's relative blood volume and pulse rate that are recorded by a cardiograph component. The contraction and relaxation of an examinee's heart will cause the polygraph system (analog or computerized) to record the systolic stroke (heart contraction), diastolic stroke (relaxation period of the heart) and a dicrotic notch, which appears during the diastolic stroke of the heart. During the data collection phase, the sensor attached to the examiner will normally be a cardiovascular blood pressure cuff assembly. Usually, this sensor consists of a rubber bladder, covered with a cloth sleeve and tightening component (velcro wrap), pump bulb assembly which includes a sphygmomanometer and associated rubber tubing for connecting the sensor to the computer sensor box or analog instrument. When scoring the cardiovascular tracing, the examiner will normally encounter four main categories of response. These are: (1) Changes in baseline, (2) Changes in amplitude, (3) Changes in rate and (4) Premature ventricular contractions. These four main categories of response consist of the following eight scoreable criteria:

**C A R D I O V A S C U L A R   T R A C I N G**

**C r i t e r i a   1 :   P h a s i c   I n c r e a s e   a n d
D e c r e a s e   i n   B a s e l i n e**

**I I - 8**
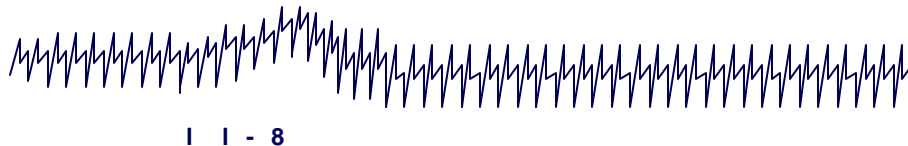
**C A R D I O V A S C U L A R   T R A C I N G**

**C r i t e r i a   2 :   T o n i c   I n c r e a s e   i n   B a s e l i n e**

**I      I - 9**
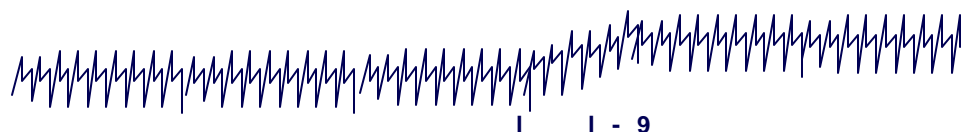
# CARDIOVASCULAR TRACING

## Criteria 3: Tonic Decrease in Baseline

I    I - 7
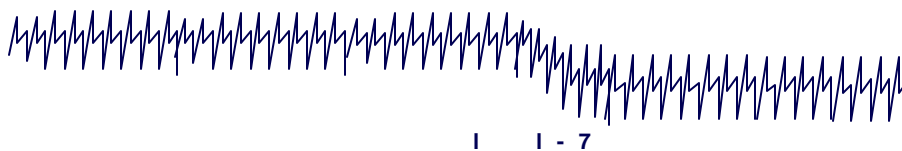
# CARDIOVASCULAR TRACING

## Criteria 4: Increase in Amplitude

I    I - 8
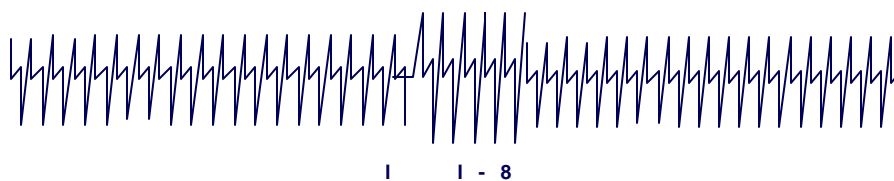
# CARDIOVASCULAR TRACING

## Criteria 5: Decrease in Amplitude

I    I - 7

# CARDIOVASCULAR TRACING

## Criteria 6: Increase in Rate

I    I - 8
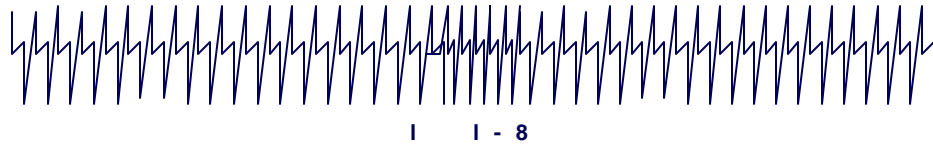
# CARDIOVASCULAR TRACING

## Criteria 7: Decrease in Rate

I    I - 7

# CARDIOVASCULAR TRACING

## Criteria 8: Premature Ventricular Contractions (PVCs)

I  I - 4          I  I - 5          I  I - 6

I    I - 7          I  I- 8          I    I - 9
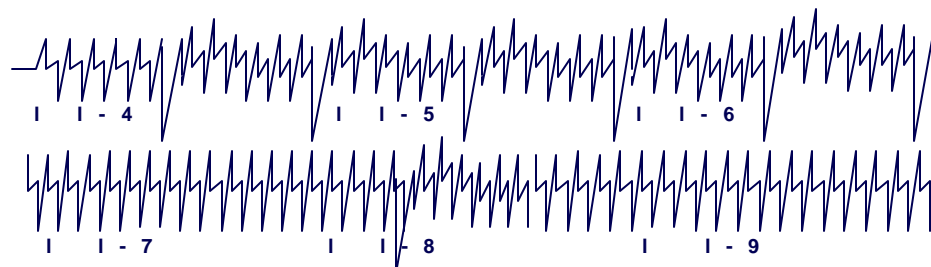
## Scoring the Cardiovascular Tracing

The primary criteria used in scoring the cardiovascular tracing will be phasic increases and returns to baseline (baseline arousal) (criterion #1). Generally, the most common physiological response observed during a conventional PDD examination is an increase (arousal) from the baseline level- usually beginning at or near stimulus onset and lasting for a few seconds with an eventual return to the prestimulus tonic level. However, for some examinees, this response may last up to 30 seconds or more. If a baseline arousal response returns to the prestimulus level, it is considered a phasic response. If there is baseline arousal without returning to the prestimulus level, it is considered a tonic response. If there are physiological changes following both the comparison and relevant questions within an analysis spot, greater weight is assigned to the question that evoked the greater change either in the amount of (degree) or duration (length) of baseline change.

Since one of the major factors in properly using the seven-position scale is consistency in applying the scoring principles, the unit of measurement for scoring a phasic or tonic response is generally a vertical chart division. In assigning values, any amount of visually discernible baseline arousal is awarded a value of +/- 1. To reach the +/- 2 level, the cardiovascular response must be at least two complete vertical chart divisions more than its comparative response. Likewise, for a +/- 3, the response must be at least three vertical chart divisions more than its comparative response.

In scoring a phasic response, duration of baseline arousal must be a consideration. Generally, duration of a cardiovascular response will allow assigning a value of +/- 1 only (no more). If one response has slightly more baseline arousal, while the other response has slightly less arousal, but with more duration, then a score of "0" would be appropriate. If both compared responses have the same amount of baseline arousal, but one response has more visually discernible duration, then a value of +/-1 is given to the response having more duration. Once a score of +/- 1 is assigned for baseline arousal in the cardiovascular tracing, then duration is no longer a consideration.

## Other Considerations in Scoring the Cardiovascular Tracing

If a score is given for a phasic change of baseline response, changes in amplitude, pulse rate and premature ventricular contractions in the cardiovascular tracing are generally not scored. Generally, these criteria are scored only when there is no baseline arousal of the cardiovascular tracing. When scored, values of +/- 1 (no more) are awarded for changes in amplitude, pulse rate and/or premature ventricular contractions in the cardiovascular. Likewise, a value of +/- 1 (no more) is awarded for a cardiovascular tracing having a tonic response (increase and/or decrease in baseline). Generally, in the comparison process, cardiovascular responses having unlike attributes (tonic against phasic; phasic response against a change in tracing amplitude with no baseline arousal; etc.) will result in a value of "0".

If both comparative responses have equal degrees and duration of baseline arousal, speed of arousal of the cardiovascular tracing from the baseline (if visually discernible) may allow assigning a value of +/- 1 (no more).

If a consistent response is exhibited to a particular question or a category of questions-comparisons only or relevants only- throughout the entire PDD examination, premature ventricular contractions may be scored. If consistency is established, a value of +/- 1 (no more) may be awarded for premature ventricular contractions; however, this cardiovascular criterion is seldom scored as consistency to a particular question or class of questions is rarely established.

## Glossary of Terms

**Analysis Spot** - The relevant and comparison question(s) that are actually evaluated during spot analysis. The number of appropriate comparison question(s) for each relevant question will vary depending on test format used [i.e., Test for Espionage and Sabotage (TES) format, Zone Comparison Test (ZCT),

and Modified General Question Test (MGQT)]. Regardless of the test format, each relevant question is always compared to the most appropriate comparison question on a tracing by tracing basis. If the test format allows a relevant question to be compared to more than one comparison question, then the comparison question with the greater response for that physiological tracing is used for comparison purposes.

**Artifact** - A change in an examinee's physiological pattern (activity) that is not attributable to a reviewed test question (stimulus) or recovery.

**Cardiovascular Tracing** - A display of physiological patterns of an examinee's relative blood volume and pulse rate that are recorded by a cardiograph component. The contraction and relaxation of an examinee's heart will cause the polygraph to record the systolic stroke (heart contraction), diastolic stroke (relaxation period of the heart) and a dicrotic notch, which appears during the diastolic stroke of the heart. The criteria used to evaluate this physiological tracing are changes in baseline, changes in amplitude and changes in rate.

**Comparison Question** - A question that is designed to produce a physiological response. During spot analysis, the physiological responses of comparison questions are compared to the physiological responses of relevant questions.

**Electrodermal Activity (EDA) Tracing** - The display of physiological patterns of either skin resistance or skin conductance obtained through exosomatic recording with a galvanograph component. When evaluating this component tracing, the criteria considered are changes in amplitude, complexity of response and duration of response.

**EDA Recovery Phase** - The physiological activity displayed in an EDA tracing that occurs between the highest peak and subsequent return to the prestimulus or newly established baseline. The EDA recovery phase begins once the tracing has reached its highest peak.

**EDA Rise Time** - The physiological activity displayed in an EDA tracing beginning with response onset and ending at the peak.

**Homeostasis** - A complex interactive regulatory system by which the body strives to maintain a state of internal equilibrium. During test data analysis, the examiner looks at the physiological tracings to ensure that the examinee is in a state of homeostasis before a scoreable test question is presented. If an examinee's physiological activity is not in a state of homeostasis (i.e., there is noise on the signal of interest) when a scoreable question is presented, then subsequent physiological activity should not be considered a response to that stimulus and cannot be scored.

**Psychophysiological Detection of Deception (PDD) Chart** - A graphic representation containing selected physiological data generated by an examinee during the data collection phase of a PDD examination.

**PDD Examination** - A process that encompasses all activities that take place between a PDD examiner and an examinee during a specific series of interactions. These interactions may include the pretest interview, use of a polygraph to collect physiological data from an examinee while presenting a series of tests (data collection phase), test data analysis phase, and the post-test interview phase, which may include interrogation of the examinee.

**PDD Examiner** - Someone who has successfully completed formal education and training in conducting PDD examinations and is certified by his or her agency to conduct such examinations.

**PDD Series** - Collection of PDD charts by presentation of reviewed test questions to an examinee the number of times required by a particular PDD testing format. A PDD examination may consist of any number of PDD series.

**PDD Test Data** - The signal of interest that may consist of artifact(s), recovery, other noise or examinee physiological response(s) to stimuli.

**PDD Test Data Analysis** - Analysis of the psychophysiological responses recorded on the PDD chart(s). For scoring purposes, only data that are timely with an applied stimulus (reviewed test question) and free of artifacts and noise on the signal of interest can be considered.

**Recovery (Returning to Homeostasis)** - A deviation in a PDD tracing attributable to a physiological phenomenon occurring as a compensatory action after a response or an artifact.

**Relevant Question** - A question that pertains directly to the matter under investigation or to the issue(s) for which the examinee is being tested.

**Respiratory Tracing** - The display of physiological patterns indicative of an examinee's breathing activity as recorded by the pneumograph component. The respiratory tracing consists of inhalation and exhalation strokes. An examinee's breathing pattern and rate may vary due to their physical conditioning. Evaluation criteria considered during the scoring process are changes in amplitude, apnea, changes in rate, changes in baseline, and loss of baseline.

**Response** - A physiological change that occurs following, and is attributable to, the presentation of an applied stimulus (i.e., reviewed test question). Responses are evaluated when they occur within the response onset window (latency exceptions) and there is no noise on the signal of interest at the time the stimulus is presented. A phasic response is a discrete (known origin) response to a specific stimulus that is generally seen as an upward movement from the baseline with subsequent return to the prestimulus (original) baseline. A tonic response is a discrete (known origin) response to a specific stimulus that is generally seen as a movement from the prestimulus baseline and establishment of a new baseline without returning to the prestimulus baseline.

**Response Amplitude** - The displayed physiological activity reflected in a PDD tracing occurring between response onset and response peak (highest level from prestimulus baseline).

**Response Duration** - The physiological activity (time) displayed between response onset and offset. Typically, this is the time from response onset until return to the prestimulus baseline (phasic response) or a newly established baseline (tonic response).

**Response Latency** - The time between stimulus onset and response onset.

**Response Onset** - The first indication of change from the prestimulus level of physiological activity to an applied stimulus (reviewed test question). To be utilized during test data analysis, unless latency is involved, response onset must occur within the response onset window to an applied stimulus (reviewed test question).

**Response Onset Window** - The period of time between stimulus onset (verbal) and an examinee's verbal response to that stimulus (assuming an examinee's verbal response occurs in a timely manner). Typically, to be considered during test data analysis, an examinee's physiological responses should occur during that period. However, if an examinee consistently exhibits response latencies that are outside of this response onset window, the response onset window may be increased to include an examinee's consistent late responding.

**Spot Analysis Concept** - The procedure wherein each component tracing is separately evaluated by comparing the response of a relevant question to the response of a comparison question.

**Stimulus Onset** - During data collection, this is the beginning of the presentation of the first word of a reviewed question.

**Tonic Level** - An examinee's level of physiological activity occurring prior to stimulus onset. This is sometimes referred to as the resting or baseline activity level. Tonic level describes a person's physiological activity when resting.

**Unwanted (Excessive) Noise on Signal of Interest** - Any noise (physiological activity) that should prevent a stimulus (scoreable test question) from being presented during the data collection phase. If an examiner asks a

scoreable question when there is unwanted noise on the signal of interest, this may prevent that question from being utilized during the scoring process. However, unwanted noise on one physiological tracing may not prevent other tracings in that same analysis spot from being evaluated. For example, unwanted noise on the respiratory tracings may prevent them from being evaluated. However, during the analysis process for that scoreable question, if the cardiovascular and/or EDA tracings are unaffected by the unwanted noise, they may be used for evaluation.

# References

Abrams, S. (1989).  The complete polygraph handbook.  Lexington, MA: Lexington Books.

Matte, J. A. (1980).  The art and science of the polygraph technique.  Springfield, IL:  Charles C. Thomas Publisher.

Matte, J. A. (1996).  Forensic psychophysiology using the polygraph:  Scientific truth  verification-lie detection.  Williamsville, NJ:  J.A.M. Publications.

Weinstein, D. A. (1994).  Anatomy and physiology for the forensic psychophysiologist.  Fort McClellan, AL:  Department of Defense Polygraph Institute.

# Development of Deception Criteria Prior to 1950

## Norman Ansley

## Abstract

This is a review of the literature published up to 1950 that contributed to the current list of physiological responses considered deception criteria. Even making allowances for differences in terminology, there are deception criteria in the current DoDPI list that had not been observed, or if observed, not described before 1950. An appendix describes Luria's motor movement technique and Wertheimer's word association test. As means of detecting deception, both were discontinued before 1950.

Key words: cardiovascular, deception criteria, electrodermal, motor movement, polygraph history, respiration, terminology, word association.

This is a review of the literature published up to 1950 that contributed to the current list of physiological responses considered deception criteria. That year marked the halfway point for the development of polygraph testing, as we know it in 1999. In 1950 that only formal polygraph training was at the Keeler Polygraph Institute, and most examiners were preceptor trained or self-taught. Most of the instruments were two-channel (cardiograph and pneumograph) mechanical units, although there were some with electrodermal units. The most widely used technique was relevant-irrelevant. A few examiners used one or the other of two published Control Question techniques, one published by Summers (1939), and the other by Inbau (1948). Among the many short-comings in 1950 was a lack of agreement on what constituted deception criteria. Add inadequate chart markings, and that independent analysis of someone else's charts was difficult, and the results were problematic.

In 1950, Charles M. Wilson, president of the International Society for the Detection of Deception (ISDD), was asked, "Should graphs be released or shown after the test?" Wilson's reply was printed in the *ISDD Bulletin*. He said that in his experience he never released an original record to anyone. He did not think making copies a good policy since possession of the record by an untrained operator represents the first step in the direction of perversion and quackery. Wilson said the charts mean nothing to anyone who was not present when the tests were run, and the only use to which they could be put was to cloud the issue (Wilson 1950).

If one examiner could not reliably read charts from another examiner, what did they know about chart interpretation in 1950? In this paper we list sixteen studies or reports which included something on deception criteria. The sixteen studies or texts did not discuss rank order scoring, only two had a form of numerical analysis, and computers were not yet useful machines. Taking the Department of Defense Polygraph Institute (DoDPI) list as state of the art for hand scoring in 1999, how many of the criteria had been identified by 1950? In the pneumograph tracing, DoDPI lists 12 items. Seven had been identified by 1950: I/E ratio change, amplitude increase, amplitude decrease/suppression, amplitude progressive decrease and return to homeostasis, respiration baseline change – temporary, baseline change – permanent, and apnea – blocking. By 1950, they had not yet observed respiration rate increase, rate decrease, respiration amplitude progressive increase followed by decrease, amplitude progressive increase and return

to homeostasis, and apnea – holding. Considering terminology, they might have seen the difference between holding and blocking but did not think the difference mattered. Some of today's examiners might have trouble recognizing DoDPI's more exact definitions of staircases up, staircases down, and staircases up and down. DoDPI lists three criteria under electrodermal, and two, amplitude and duration, were in the pre-1950 literature. Only the complex response, which some examiners call a saddle, was not mentioned. DoDPI includes eight deception criteria under cardiograph including premature ventricular contractions which were not listed prior to 1950, but some would say should not be listed now. Of the seven others on the list, six were known: phasic increase and decrease in baseline, tonic increase in baseline, tonic decrease in baseline, pulse rate increase, pulse rate decrease, and decrease in amplitude of the tracing. The one lacking in 1950 was the increase of the amplitude of the tracing. It would appear that the well-informed examiner of 1950 had enough deception criteria to decide most of his cases, but the more we go back in time, the less he had. The cumulative growth of a body of technical and scientific knowledge is a vital part of a profession. In the text that follows we will see the development of knowledge.

One wonders if the pioneers in instrumental detection of deception knew of Daniel Defoe's proposal to take the pulse of a suspected thief. One would think he was discussing a modern polygraph problem when he observed, "It may be true that this discovery by the pulsation of the blood cannot be brought to a certainty, and therefore it is not to be brought into evidence; but I insist, if it be duly and skillfully observed, it may be brought to be allowed for a just addition to other circumstances, especially if concurring with other just grounds of suspicion." (1730) (Moore 1955).

Cesare Lombroso (1911) mentions a case in which he used his recording hydrosphygmograph. His apparatus measured blood volume and pulse rate. He reports, "The same apathy persisted when he was spoken to of the robbery on the railroad, while there was an enormous depression – a fall of 14mm – when the Torelli theft was mentioned. I

concluded, that he had no part in the railway robbery, but he had certainly participated in the Torelli affair; and my conclusions were completely verified." Here we have a measure of a cardiovascular reaction, and a verified decision.

In 1914, Vittorio Benussi published the results of an experiment relating to the symptoms of lying in respiration. At the University of Graz in Leipzig, Benussi had subjects read aloud five statements, some of which were coded and not to be read as stated. Half of the items in the 80 experiments were to be lied about. Panels of witnesses made judgments as to when subjects were lying, and when they were telling the truth. Using a Marey pneumograph which recorded on a polygraph, Benussi measured the distance between the beginning and end (length) of each of three, four, or five cycles of breathing after the subject spoke. For each cycle of breathing Benussi measured the length (time) of the inspiration (I) and the length of the expiration (E), and calculated the ratio (I/E) for each of the cycles before and after the statement. He found that lying produced greater I/E ratios than truthfulness. Of the 80 experiments, I/E analysis resulted in one false positive error and one false negative error, for a total accuracy of 97.5%. The average panel accuracy was 56% for truthful and 58% for deceptive statements. This experiment attracted the attention of Marston, Larson, and others to the diagnostic value of a respiratory recording. Also, the I/E ratio has remained on the deception criteria lists of the DoD Polygraph Institute, The Maryland Institute of Criminal Justice, and other polygraph schools and courses.

John A. Larson (1923) had experience in hundreds of criminal cases as a basis for his description of deception criteria. Larson recorded a continuous cardiograph and pneumograph pattern on a smoked drum apparatus, and observed that the record of the innocent suspect will usually vary but slightly, if at all, from its normal. In describing some guilty test results, Larson describes repression in the pneumograph tracing, and the accompanying chart illustration shows a rise and fall in the cardiograph pattern of the confirmed deception. In another chart we see suppression, loss of baseline, and changes in the

I/E ratio and rhythm and regularity in the pneumograph and a rise and fall in the cardiograph tracing, but his text does not describe this illustration. Larson notes that the cardiac curve is usually more significant than the respiratory curve. In the description of a chart, Larson writes of the extreme blocking effect of deception. In one chart Larson described deception causing a drop in the blood pressure curve with the obliteration of the pulsations. In addition, there was an increase in frequency. Describing another chart segment with a lie, Larson notes in both the cardiac and respiratory curves there was repression. Larson states the following changes have been observed as the effect of deception. These changes may occur in both the cardiac and the respiratory curves or in one alone, more frequently in the cardiac action:

1. Increase in blood pressure – a rise.

2. Decrease in blood pressure.

3. Increase in height.

4. Increase in frequency.

5. Summative effects.

6. Incomplete inhibition.

7. Complete inhibitory effect.

8. Irregular fluctuations, especially noticeable at the base of each cardiac pulsation.

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

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

Leonarde Keeler (1930) wanted to compare the peak of tension polygraph technique with the word association method. Seventy-five subjects took a one-chart peak of tension on which of ten cards they had chosen. If the chosen card, placed by chance, was first or last in the sequence, the test was repeated in a different sequence. The deception criteria were a rise in blood pressure followed by a release in tension after the chosen card, and the greatest suppression in the respiratory tracing. There were 71 correct decisions of 75 (95%) on the first trial. Post-test interviews attributed the failures to a lack of interest or concern which resulted in a lack of responses. Here we have pneumograph suppression and a rise and relief in the cardiograph pattern established as valid deception criteria. By comparison, the word association test of 30 students was correct in 19 (63%). For results of another comparison see the work of John E. Winter (1936) in a dormitory theft case.

Professor John E. Winter (1936) investigated thefts in the women's dormitories at West Virginia University with two methods: Jung's word association test with a chrono-scope for reaction time, and a Larson type polygraph test employing respiratory and cardiovascular measures, from separate devices. The breathing curve was rated as regular or irregular; light or deep. The blood pressure curve was rated as regular or irregular, and medium or strong. Winter gave three levels of significance to the results of each of the methods: 0 for no significance, "nothing to indicate guilt;" 1 for "some significance and points in direction of guilt;" and 2 for "distinct signs of guilt." There were 25 women suspects and each received two Larson type tests, with consistent responses except for the culprit. The first test of each subject was labeled practice. From the respiration recording there were 24 zeros, including the thief, who confessed. On her practice she scored a 2 on her cardiograph curve, the only one to do so. She was given a post-confession test where she again scored a 2 on the cardiograph curve. This may be the first case of numerical scoring. Word association cleared 19 innocent suspects, and had the thief among the five who scored a 1.

Winter's polygraph apparatus was reported as "an ordinary pneumograph, a Baumanometer, an improved form of the Erlanger capsule for high and low air pressure, and a MacKenzie polygraph for a continuous record of breathing and heart action." For a picture and description of the MacKenzie polygraph, see *Polygraph* (1992) 21(4) 349-350.

C.D. Lee wrote an article, "The Lie Detector," published in the September, 1937 issue of the *Fingerprint and Identification Magazine.* Lee illustrates the article with a picture of a chart from the examination of Jerone Selz who confessed to murder after the test. There was a double rise and fall in the cardiograph pattern and suppressed respiration following the question, "Did you kill Mrs. Rice?" The remainder of the article is about the instrument and testing.

Leon G. Turrou (1938) in his book Nazi Spies in America describes several polygraph examinations given to suspects and witnesses involved in a German espionage ring. Turrou describes how the instrument functions (cardiograph and pneumograph), then quotes Keeler on the procedure. Eight suspects or witnesses were tested. Because many questions were asked each examinee, a system of asterisks was devised to give some indication of results. In the report, one asterisk after a question indicated a mild emotional reaction, two a strong emotional reaction, and three asterisks, quite an emotional reaction, "such as would be found when the subject is telling a whopper." One examinee was asked nine relevant questions. There were no asterisks behind four of the questions, two asterisks behind one question, and three asterisks behind four questions; a split call from a multiple issue relevant-irrelevant test format. During the testing of a suspect, 18 relevants were asked, and in the report there were no asterisks behind five of the relevant questions, one asterisk behind four of the relevants, two asterisks behind five relevants, and three asterisks behind four relevants. This evaluation of the charts was unusual, at least unusual to appear in the report. In reality, the asterisks were a numerical system, zero to three, for each question.

William M. Marston published a book in 1938. Under the heading "Judging a Polygraph Record," Marston states that changes in the blood pressure are the chief and only dependable criterion of deception. This is shown by the shifting of the entire mass of pulse tracings toward the upper edge of the recording strip. Variations in the pulse are not significant. Regarding breathing, Marston said marked changes in respiration

tracings that accompany changes in the blood pressure justify a judgment of deception. He noted that Benussi's breathing ratios are probably extremely significant of lying, but it has never proved practical. Marston said a sudden hump in the breathing record may be meaningful, as may a "shoulder" in either the inspiration or expiration tracing. Also indicative is 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.

The Reverend Walter G. Summers, S.J., prepared a paper on his work before his death on September 24, 1938. Published in 1939, it describes a sophisticated test format and means of chart analysis. In a theft cases there would be three relevant questions. In sequence the questions asked about knowledge, guilt, and possession. Called "significant" questions, examples were, "Do you know who took the money?", "Did you take the money?", and "Have you the money on your person?" He said that within one record there were usually included three different but related significant questions, each of which was asked three times. Interspersed among the non-significant questions (irrelevants) are emotional standard questions (controls). An emotional standard question precedes each significant question. The format is three pairs of control-relevant question, with irrelevants put in as needed. Examples of irrelevants were "Are you wearing a black coat?" and "Did you eat breakfast this morning?" Examples of emotional standards, developed after extensive interviewing of the examinee were: "Where you ever arrested?" and "Do you own a revolver?"

The analytical system is modern. Summers said "...we contrast and compare the reactions to the significant questions with the reactions to the emotional standards. If the reactions to the 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." Professor Summers used a recording galvanometer, the Fordham Patho-meter, which he manufactured. A letter to the author from William E. Kirwan in 1952, indicated the New York State Troopers Scientific Laboratory was still using the Summer's technique, with excellent results (Kirwan 1952). Summers, who conducted laboratory and criminal cases, established the control question test concept, including the analytic procedure (Summers 1939).

Paul Trovillo (1942) wrote what is probably the first treatise on the topic of deception test criteria. The illustrations were taken from real cases. Although the electrodermal unit was not widely used, there is a good section of illustrations of GSR tracings. For the cardiograph he lists and illustrates:

1.  Common form of blood pressure rise (and return to baseline).

2.  Blood pressure increase. . . complicated by cyclical increase throughout the graph.

3.  Rapid rise and decline in blood pressure, accompanied by obliteration of pulse amplitude.

4.  Gradual increase in blood pressure.

5.  Constriction of pulse amplitude and gradual rise in blood pressure.

6.  Slight rise accompanied by rapid decline in blood pressure.

7.  Peak of tension.

8.  Rapid changes in heart rhythm.

9.  Another form of change in heart rhythm (includes general pulse irregularity).

10.  Complication of deception pattern – increase in blood pressure and return to baseline, variations in pulse frequency, and reduction of pulse amplitude.

11.  Reduction in pulse amplitude.

For the respiration tracing, he lists and illustrates:

1.  Suppression at point of deception.

2.  Respiratory block.

3.  Rise in baseline.

4.  Respiratory suppression preceding deception stimulus, followed by deeper respiration at point of deception.

5.  Regularity of respiration up to and through the deception stimulus, followed by irregular respiration.

6.  Respiratory irregularities up to point of deception, followed by regular respiration.

For the electrodermal he lists and illustrates:

1.  Comparatively large area of reaction at point of deception.

2.  Comparatively large magnitude of reaction at point of deception.

3.  Peak of tension test (experimental age test), reactions to each age up to and including the point of deception, then none.

4.  Peak of tension card test. The only large reaction.

5.  Peak of tension. Pattern at deception different from patterns at truthful answers.

6.  Gradual rise in the electrodermal pattern.

Trovillo then lists and illustrates what he calls ambiguities in the records. In the cardiograph tracing he shows the effect of body movements, a deep breath, general excitement, increase in blood pressure even at irrelevant questions, an absence of blood pressure and pulse rate changes during lying, inconsistency of reactions on questions involving guilt, startle response of innocent subjects, and a cardiac irregularity.

For ambiguous respiratory patterns he lists and illustrates: deception-like suppression found among some innocent examinees, effects of superfluous talking and physical movement, erratic breathing of an innocent person from great fear, a deep breath taken deliberately to obliterate suppression, normal shallow breathing following a deep breath, effect of sinus congestion, lack of response in known guilty subject, and respiratory tremor found in both relevant and irrelevant questions by an excited person.

For the ambiguous electrodermal patterns he lists and illustrates: over-activity of the reaction, effects of bodily movement, effect of deep breath at the very moment of response, unresponsiveness in guilty subject, inconsistent reactions in guilty subject, and guilt reactions in innocent persons.

In 1942, Fred E. Inbau published the first of his three books on *Lie Detection and Criminal Interrogation*. The techniques were relevant-irrelevant and peak of tension. In the section on deception criteria he notes that the criteria differ somewhat for the two techniques. For the cardiograph he mentions an increase in blood pressure and the illustration shows it returning to baseline after first going below the baseline. Other criteria include a sharp drop in blood pressure, and slowing of the pulse rate. For the respiration pattern he lists suppression, and heavy breathing about twenty or twenty-five seconds after the reply to a question. Inbau writes about the EDA and the lack of knowledge about it, and concludes that electrodermal tracings alone cannot be considered as adequate for deception diagnosis, but it may be occasionally helpful as an adjunct to the other recordings.

In 1943, C.D. Lee prepared an Instruction Manual for the Berkeley Polygraph. It is a complete text on conducting examinations and reading the charts. The methods are relevant-irrelevant and peak of tension. He notes the pattern of the innocent is one of regularity and uniformity with no marked difference between the effect produced by neutral questions and those related to the crime. The tension may remain constant, decrease, fluctuate slightly, but seldom increase. In the guilty, the tension is lacking in regularity and uniformity. Illustrations show a phasic rise and fall of the cardiograph pattern associated with deception. In the pneumograph, he shows repressed breathing, followed later by a sigh of relief. Most of the illustrations were of the cardiograph pattern, and the rise and fall of the cardiograph pattern is clearly the primary indication of deception.

Joseph W. Haney (1944) was a forensic psychologist and experienced polygraph examiner in the Chicago Crime Laboratory. He was interested in the catalogue of deception criteria by Paul V. Trovillo. Haney wondered if the respiration responses described by Trovillo might not be produced by a nondeceptive mental task as well as deception. Haney did that, producing charts with blocking (apnea), suppression, and baseline rises. Haney suggested that before using these as deception criteria, one should see if they occur also at irrelevant questions.

In 1948, Fred E. Inbau published the second edition of his book *Lie Detection and Criminal Interrogation.* In addition to the relevant-irrelevant and peak of tension tests there was the Reid control question test. The section on deception criteria has not changed in a significant way. For the cardiograph, Inbau mentions an increase in blood pressure, and the illustration shows it returning to baseline after first dropping below the baseline. Other criteria include a sharp drop in blood pressure, and slowing of the pulse rate. For the respiration system he lists suppression, and heavy breathing about twenty or twenty-five seconds after the reply to a question. In regard to the electrodermal channel, the author said electrodermal responses have been found to be of little practical value in diagnosing deception.

Baesen, Chung & Yang (1948-1949) tell us the chart criteria they used in a laboratory research project employing a two-channel Keeler polygraph. They used pulse rate changes, sudden and delayed drops in blood pressure, duration of rise and fall in blood pressure, and location of the dicrotic notch. Notice was taken of changes in respiration baseline, blocking and suppression of respiration either prior to, during, or immediately following the question.

In 1950, Colonel Ralph W. Pierce, president of Leonarde Keeler, Inc. was writing about the use of the peak of tension test. He wrote, "One man reacted to this test, his blood pressure rising until the question concerning the German Luger was asked, then falling off. He also showed marked irregularity in his breathing up to the question about Luger, followed by regularity to the end of the test. The galvanometer pen also rose sharply at the question concerning the Luger. This man also reacted similarly to the other tests referring to the disposition of the gun, its condition, etc." The examinee confessed to the crime. Colonel Pierce's description has tonic changes in the cardiograph and pneumograph tracings, but a phasic response in the electrodermal, channel showing a combination of deception criteria.

## Abandoned Methods for Detecting Deception

In the period before 1950 there were two techniques that were subject to considerable research as means for detecting deception. Their criteria for deception were not related to the methods in the polygraph technique.

Luria (1930, 1932) developed a lie detection method that involved tremors and motor movement. It received some research attention in the United States but was not used in criminal cases (Berrien, 1939; Morgan & Ojemann 1942).

From the turn of the century into the 1930s the word association test was considered a method for detecting guilt in criminal cases (Wertheimer & Klein 1904, Jung 1919). However, Larson (1922) found "the association words with time reaction do not give as satisfactory results as the cardio-respiratory changes." Larson added, "We can say this definitely in cases where the suspect has subsequently confessed where, although there were marked and striking changes in the tracings, the findings by association method were not significant." Keeler (1930) found the association method performed poorly when compared to polygraph test results. Winter (1936) in a real case of theft involving 25 students, found the association method had the thief among five in a narrowed pool of suspects, but his cardio-pneumo method identified the culprit, followed by a confession. Although word association with reaction time remains as a psychological tool, its use in solving crime has disappeared.

# References

Baesen, H. V., Chung, C. & Yang, C. (1948-49). A lie detector experiment. *Journal of Criminal Law, Criminology, and Police Science*, 39, 532-537.

Benussi, V. (1914). Die atmung asymptome der luge (The respiratory symptoms of lying). *Archiv fur die Gestamte Psychologie*, 11, 244-273. Translated and printed in *Polygraph*, 4(1), 52-76.

Berrien, F. K. (1939). Finger oscillations as indices of emotion. I Preliminary Validation, *Journal of Experimental Psychology*, 24, 485-498; and II Further validation and use in detecting deception, *Journal of Experimental Psychology*, 24, 609-620.

Haney, J. W. (1943-44). The analyses of respiratory criteria in deception tests, a possible source of misinterpretation. *Journal of Criminal Law and Criminology*, 34, 268-273.

Inbau, F. E. (1942). *Lie Detection and Criminal Interrogation*. Baltimore: Williams & Wilkins Company.

Inbau, F. E. (1948). *Lie Detection and Criminal Interrogation*. Second Ed., Rev. Baltimore: Williams & Wilkins Company.

*Instruction Manual for the Berkeley Psychograph* (1943).  San Rafael, California:  Lee & Sons.

Jung, C. G. (1919).  *Studies in Word Association.*  New York:  Moffet, Yard & Company.

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

Kirwan, W. E., (16 Oct 1952).  Letter from the Director or the New York State Troopers Scientific Laboratory to Norman Ansley.

Larson, J. A. (1923).  The cardio-pneumo-psychogram and its use in the study of the emotions, with practical applications.  *Journal of Experimental Psychology*, 5(5), 323-328.

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

Lee, C. D. (Sep 1937).  The lie detector.  *Fingerprint and Identification Magazine.*

Lombroso, C., M.D. (1911).  *Crime:  Its Causes and Remedies*.  Boston:  Little, Brown & Company.  Excerpt reprinted in Polygraph (1991) 20 (1) 52-53.

Luria, A. R. (1930).  Die Methode der Abbildenden Motorik in der Titbestandiagnostik.  *Zeitschrift fur Angewandte Psycholgie*, 35, 139-183.

Luria, A. R. (1932).  *The Nature of Human Conflicts:  On Emotion, Conflict and Will.*  New York:  Liveright.

Marston, W. M. (1938).  *The Lie Detector Test.*  New York:  Richard R. Smith.  Reprinted by the American Polygraph Association, 1989.

Moore, J. R. (1955).  Defoe's project for lie-detection.  *American Journal of Psychology*, 68, 672.

Morgan, M. I. & Ojemann, R. H. (1942).  A study of the luria method.  *Journal of Applied Psychology*, 26, 168-179.

Pierce, R. W. (1950).  The peak of tension test.  *ISDD Bulletin*, 3 (3) 1, 4, 5.

Summers, W. G. (1939).  Science can get the confession.  *Fordhan Law Review*, 8, 335-354.

Trovillo, P. V. (1942).  Deception test criteria:  How one can determine truth and falsehood from polygraphic records.  *Journal of Criminal Law and Criminology*, 33 (4) 338-358.

Turrou, L. G. (1938).  *Nazi Spies in America.*  New York:  Random House.

Wertheimer, M. & Klein J. (1904).  Psychologische Tatbestandsdiagnostik.  *Achiv fur Kriminal-Anthropologie und Kriminalistic*, 15, 78-113.

Wilson, C. M. (1950).  Should graphs be released or shown after test?  *ISDD Bulletin*, 3 (3), 5, 8.

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

**Attachment 1**
**Deception Criteria for Lie Detection Pioneers**

| | Defoe 1730 | Lombroso 1911 | Benussi 1914 | Larson 1923 | Keeler 1930 | Winter 1936 | Lee 1937 | Turrou 1938 | Marston 1938 | Summers 1939 | Trovillo 1942 | Inbau 1942 | Lee 1943 | Haney 1944 | Inbau 1948 | Baesen et al 1949 | Pierce 1950 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Respiratory** | | | | | | | | | | | | | | | | | |
| rate decrease | | | | | | | | | | | | | | | | | |
| rate increase | | | | | | | | | | | | | | | | | |
| I/E ratio change | | | X | | | | | | X | | | | | | | | |
| amplitude increase | | | | | | X | | | X | | X | X | X | | X | | |
| amplitude decrease-suppression | | | | X | X | X | X | | X | | X | X | X | X | X | | |
| progressive increase - decrease | | | | | | | | | | | | | | | | | |
| progressive increase and return | | | | | | | | | | | | | | | | | |
| progressive decrease and return | | | | | | | | | X | | | | | | | | |
| baseline change - temporary | | | | | | | | | | | | | | | X | | |
| baseline change - permanent | | | | | | | | | | | X | | | X | X | | |
| apnea - holding (inhalation) | | | | | | | | | | | | | | | | | |
| apnea - holding (exhalation) | | | | X | | | | | | | X | | | X | X | | |
| **Electrodermal** | | | | | | | | | | | | | | | | | |
| amplitude change | | | | | | | | | | X | X | | | | | | X |
| complex response | | | | | | | | | | | | | | | | | |
| response duration and return | | | | | | | | | | | X | | | | | | |
| **Cardiovascular** | | | | | | | | | | | | | | | | | |
| baseline increase and decrease | | | | X | X | | X | | | | X | X | X | | X | | X |
| baseline increase | | | | X | | X | | | X | | X | | | | | | |
| baseline decrease | | X | | X | | X | | | | | | X | | | | X | |
| amplitude increase | | | | | | | | | | | | | | | | | |
| amplitude decrease | | | | | | | | | | | X | | | | | | |
| pulse rate increase | X | | | X | | | | | | | X | | | | | X | |
| pulse rate decrease | | | | | | | | | | | X | | | | X | X | |
| PVCs | | | | | | | | | | | | | | | | | |
| Numerical analysis | | | | | | X | | | X | | | | | | | | |

# Numerical Evaluation of the Army Zone Comparison Test

## Gary D. Light

## Abstract

This study involved a comparison between the spot-totaling and cumulative numerical evaluation procedures with the zone comparison test (ZCT) as applied by the United States Army Criminal Investigation Command (USACIDC) polygraph program.  The examinations utilized for this research were conducted 1 January through 31 March 1991.  Subsequently, the USACIDC polygraph program replicated this research with data from the calendar year 1997, and these data were incorporated into this study.   The cumulative test data analysis procedure was applied to the ZCT examinations from both time periods. A total of 358 confirmed deceptive examinations were identified.  The cumulative evaluation procedure correctly identified 241 (67%) of the confirmed deceptive examinations while 107 examinations (30%) would have been classified as no opinion and ten examinations (3%) would have been classified as false negatives.  Based upon these findings, it would appear that the cost of the cumulative evaluation procedure in terms of utility and accuracy of the PDD process (30% no opinion) is too significant for an agency supporting a law enforcement mission to accept.

In 1961, the ZCT, as developed by Cleve Backster, was adopted by USACIDC for use in polygraph examinations of criminal suspects (Brisentine, 1991).  In 1962, with the permission and assistance of Cleve Backster, the ZCT was incorporated into the formal lesson plan of the Polygraph School, United States Army Military Police School (USAMPS), Fort Gordon, GA (Decker, 1991). The ZCT, with certain modifications, is still being taught at the Department of Defense Polygraph Institute (DoDPI) (Cole, 1991)].  This questioning format of the ZCT is referred to as the "Army" ZCT.  One modification of the original ZCT protocol incorporated by USACIDC and USAMPS was a change in the procedure for test data analysis (scoring). The numerical evaluation of the ZCT was taught in two formats at USAMPS (Sneed, 1991). Originally, the numerical evaluation of the test data was based solely on the overall sum of the spot totals, referred to as "cumulative scoring" (Brisentine, 1991). The second numerical evaluation procedure, as taught at USAMPS and currently at DoDPI, is referred to as "spot-totaling". USAMPS adopted the spot-totaling procedure after a review of numerous criminal specific examinations.  During the USAMPS review, it was found that the cumulative method of chart evaluation was incorrectly assigning deceptive examinees No Opinion (NO) classifications (Sneed, 1991).

In the ZCT, a spot is the pairing of a relevant and comparison questions in which the physiological responses are compared component by component against one another, and a whole number value between –3 and +3 is assigned to each spot for each component. The difference between spot-totaling and the cumulative evaluation procedure is the decision rule.  The spot-totaling numerical evaluation procedure results in a determination of non-deception indicated (NDI) if the sum of all spot-totals greater than +5, with a positive score occurring in each spot.  The classification of Deception Indicated (DI) is made if the overall total is less than –5, or if the evaluation of any relevant question (spot-total) is -3 or less, regardless of the grand sum of all spot-totals.  A classification of NO is rendered when an the grand sum for the

examination falls between +6 and –6, or an examination has a spot-total of 0 to minus -2, regardless of the cumulative total.

The cumulative numerical evaluation procedure results in a determination of NDI if the sum of all of the spot-totals is +6 or greater. The classification of DI is rendered if the sum of the spot-totals is -6 or less. A classification of NO is rendered when the sum of the spot-totals is between +6 and -6. No decision is based on the score of an individual relevant question.

The ZCT is used in approximately half of the polygraph examinations conducted by the USACIDC polygraph program. Between 1980 and 1990, USACIDC conducted approximately 15,839 field examinations of criminal suspects utilizing the ZCT. During this period 76% of these examinations were confirmed as deceptive. Further, between 1980 and 1990, USACIDC maintained a confession rate of 70% of those examinations in which the suspect was called DI using USACIDC spot-totaling scoring rules. Since 1966, USACIDC has utilized the spot-totaling method exclusively for chart evaluation (Brisentine, 1991).

The research study of Capps and Ansley (1991) indicated that polygraph examinations utilizing the spot-totaling method of numerical evaluation might result in a significant number of false positive examinations. In this study, Capps and Ansley used confirmed polygraph examinations that were numerically evaluated, and opinions that were based upon the cumulative totaling procedure. In this research, an opinion of NDI was rendered even when spot-totals were in the minus ranges. The Capps and Ansley research reported that when utilizing spot-totaling rules, the false positive error is not as readily identified. The research indicated that by "classifying in this way (cumulative totaling procedure), false positive errors that may have come into existence by use of the spot rule are identified" (Capps and Ansley, 1991).

In 1985, Richard Weaver, utilizing 15 criminal specific examinations retrieved from the Wisconsin State Crime Laboratory, applied the Backster, Utah and USAMPS numerical

evaluation procedures. The results of these evaluation procedures indicated that although there was no significant difference in the opinions rendered when similar classification procedures (Utah and USAMPS) were utilized, a significant difference occurred when the weaker comparison question was compared to a relevant question. These findings indicate that the method of evaluating the test data can affect ZCT decisions.

## Method

This research project was originally initiated in response to the research project of Capps and Ansley (1991), that indicated that the cumulative evaluation process might be a more appropriate method of evaluation for the ZCT by a government agency. Specifically, the USACIDC polygraph program evaluated the spot-total and the cumulative evaluation procedures to determine which would be the most appropriate for that agency. The original research and report were completed in 1991, based upon the review of all DI examinations conducted by the USACIDC polygraph program between 1 January and 31 March 1991, utilizing the ZCT question format as taught at USAMPS. The 1991 research was an internal review of existing PDD procedures with no intent of publishing the results. In 1998, it was requested the research be edited for submission to this journal for possible publication. Prior to the 1991 report being edited and submitted for publication, the USACIDC program was provided with a copy of the draft report. The USACIDC program requested that they be allowed to replicate the data collection procedures completed in 1991. It was agreed that USACIDC personnel would review all DI examinations involving the ZCT for the calendar year 1997. The examinations that had been determined to be DI utilizing the spot-totaling procedure would subsequently be evaluated utilizing the cumulative evaluation procedure. The results of that procedure would then be compared to the data collected in 1991.

Between 1 January and 31 March 1991, 482 examinations were conducted and 145 of these examinations were opined to be DI. All of the test data for these 145 examinations were evaluated by the spot-totaling procedure. Of these 145 DI

examinations, 108 were verified as DI by use of confession.

Between 1 January and 31 December 1997, 959 examinations were conducted and 556 examinations were opined to be DI. All of the test data for these 556 examinations were evaluated by the spot-totaling procedure. Of these 556 DI examinations, 250 were verified deceptive by posttest confession of the examinee.

The 358 examinations (1991 and 1997 totals) that were confirmed deceptive consisted of examinations in which the examinee confessed to being involved in the investigation for which the polygraph examination was conducted. The quality control section of the USACIDC polygraph program routinely reviews all examinations that are completed for accuracy, and to determine if the opinion rendered by the original examiner can be supported by the agency's investigation. A part of that quality control process is to determine if the examinee has made statements which support the opinion which is rendered based upon the physiological responses to the relevant questions. When statements are made by the examinee that are against the self-interest of the individual and are consistent with the deceptive physiological data, this examination is considered confirmed. As will be discussed later, when inconsistent results are found between the case facts and the physiological responses to the relevant questions, a review of this discrepancy is completed and this examination is categorized in the USACIDC polygraph program data base as a "contradicted examination".

All of the 358 DI examinations used in this research project underwent the USACIDC quality control review to verify that each examination was confirmed as deceptive. The criterion of confessions as ground truth is a subject for debate (Patrick and Iacono, 1991). This research project is not a validation study, but is concerned only with the impact the cumulative evaluation procedure would have when applying this test data evaluation procedure in lieu of the spot-total analysis procedure.

The Lafayette Factfinder analog polygraph was the primary instrument used in 1991. The 1997 examinations were conducted mainly with the Axciton (Axciton Systems, Houston, Texas) computerized polygraph system. All polygraph examinations administered underwent the USACIDC quality control process, which ensured each examination complied with the policies and standards required by USACIDC. With minor modifications, the USACIDC policies adhere to the procedures as taught by DoDPI for the conduct of the ZCT. The examiners who collected the examinations for this project were either certified examiners, or interns completing their internship for the USACIDC polygraph program. Each examiner had a bachelor's degree and at least five years of criminal investigative experience as a USACIDC special agent.

The following ZCT question format was utilized by USACIDC in the conduct of these examinations:

> Neutral
> Sacrifice Relevant
> Symptomatic
> Comparison
> Primary Relevant
> Comparison
> Primary Relevant
> Symptomatic
> Comparison
> Secondary Relevant

The review of the relevant literature for the original USACIDC research project was completed in 1991. The review of the literature was not updated for the 1997 data. The 1997 data was incorporated to this research project at the request of USACIDC polygraph program. It is believed that the replication of the procedures by different researchers and involving data collected six years apart would significantly add to this research.

## Results

### DI Confirmed Examinations in 1991

Of the 145 DI examinations, 108 examinations were confirmed by confession. When utilizing cumulative evaluation procedures, 74 (69%) of the examinations were determined to be DI. Of the remaining

examinations, 32 (30%) examinations would have been determined to be NO and two (1%) would have been classified as false negative. (See Table 1.)

**Table 1.**
**Polygraph decisions in 1991 as they would be affected by two types of decision rules: cumulative and spot totals. (n=108 confirmed deceptive examinations)**

Results

| Method | DI | NO | NDI |
|---|---|---|---|
| Cumulative | 74 (69%) | 32 (30%) | 2 (1%) |
| Spot-Totaling | 108 (100%) | 0 | 0 |

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

### DI Confirmed Examinations in 1997

Of the 556 DI examinations, 250 examinations were confirmed by confession. When utilizing cumulative evaluation procedures, 167 (67%) of the examinations were determined to be DI. Of the remaining examinations, 75 (30%) examinations would have been determined to be NO and eight (3%) would have been classified as false negative.

**Table 2.**
**Polygraph decisions in 1997 as they would be affected by two types of decision rules: cumulative and spot totals. (n=250 confirmed deceptive examinations)**

Results

| Method | DI | NO | NDI |
|---|---|---|---|
| Cumulative | 167 (67%) | 75 (30%) | 8 |
| Spot-Totaling | 250 (100%) | 0 | 0 |

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

### Combined Confirmed DI Examinations in 1991 and 1997

Of the 701 total DI examinations, 358 examinations were confirmed by confession. When utilizing cumulative evaluation procedures, 241 (67%) of the examinations were determined to be DI. Of the remaining examinations, 107 (28%) examinations would have been determined to be NO and ten (3%) would have been classified as false negative.

**Table 3.**
**Polygraph decisions in 1991 and 1997 combined as they would be affected by two types of decision rules:  cumulative and spot totals. (n=358 confirmed deceptive examinations)**

Results

| Method | DI | NO | NDI |
|---|---|---|---|
| Cumulative | 241 (67%) | 107 (30%) | 10 (3%) |
| Spot-Totaling | 358 (100%) | 0 | 0 |

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

# Discussion

This research indicates that the cumulative evaluation procedure has a significant negative impact upon the outcome of the ZCT.  Applying the cumulative evaluation procedure to the confirmed DI examinations increased the NO rate to 30%. This NO ratio is significantly higher than the 13% NO ratio attained by USACIDC field examiners during the 1990 calendar year.  The 13% NO ratio includes all USACIDC examinations, called truthful or deceptive, with all question formats for that time period. The USACIDC NO percentage, when NDI examinations are not considered, was less that 8% for the calendar year 1990.

A 30% rate of NO outcomes would have a significant negative impact upon an agency supporting a criminal investigative mission. Attachment A outlines seven case histories that illustrate the problem.  The case histories clearly illustrate that persons who have confessed to homicide, child molestation, forcible rape, larceny, indecent assaults, frauds, and other felony offenses would not have been correctly identified. Six of the seven examples provided were deceptive examinees who would have been classified as NO.  The number of these suspects who would have failed to confess subsequent to further polygraph testing is unknown.  However, the likelihood of resolving issues through testimonial evidence is reduced with each subsequent series of questions required to obtain a conclusive opinion. Case history 6 is a troubling example.  That guilty suspect would have been misclassified as NDI of a forcible rape without the spot-totaling decision rule.  The suspect may not have been interrogated at all, if the cumulative evaluation procedure were accepted as the test data evaluation procedure for USACIDC.

The rationale for the use of the cumulative evaluation procedure is the concern for the false positive, that is, the polygraph examination identified the individual as having committed the offense, when in fact the person was not involved in the incident.  The concern is that, "the classification used by those government agencies that employ the spot rule does not allow for this (false positive), since a subject with a minus three in any one overall spot-total is classified as practicing deception (and filed accordingly) regardless of whether or not he is truthful" (Capps and Ansley, 1991).  The USACIDC polygraph program, since 1976, conducts a review of all final reports of investigation in which a polygraph examination was completed (Brisentine, 1991). These reports of investigations are compared to the polygraph examination report that was conducted in support of that investigation. These two reports are compared to assure that the polygraph examination is consistent with the results of the field agent that conducted the original investigation.  This final review is to identify those examinations in which the case agent's final report of investigation is contradicted by the polygraph report.  Once a polygraph report is identified as being possibly contradicted by the final report of investigation, an attempt is made to identify the cause for this conflict.   The number of contradicted examinations is approximately 6

per year. Those examinations have traditionally been both false negative and positive. The occurrence of a significant number of false positive examinations in USACIDC is not in evidence.

The concerns of researchers that false positive examinations can occur in up to 50% of the examinations (Ben-Shakhar and Furedy, 1990) is not supported by the USACIDC evidence. There is little doubt that if false positive decisions occurred as suspected by some critics of PDD, such a significant error rate would be identified through the administrative processing of USACIDC PDD examinations. The evidence is in sharp contrast to these hypothetical projections.

The impact of the false positive for a specific issue examination in a law enforcement setting can be addressed by establishing priorities based upon costs and utilities. This concept was recognized by Ben-Shakhar, Lieblich, and Bar-Hillel (1982), when they noted that one of the few situations in which polygraph tests could have positive costs and utility is the police investigation. They recognized that the costs associated with a false positive in the criminal specific examination are minimal while the utility is significant. Ben-Shakhar and Furedy (1990), placed the costs associated with the false positive in the context of the USACIDC specific issue examination when stating the examination is, "a case where the outcome would only be a decision of the sort of whether to continue the interrogation of a given suspect or to release the person and concentrate on alternate leads".

## Summary

The use of a NO in test data evaluation is a necessary safety valve that precludes forcing a conclusive opinion when that opinion cannot be defended with the available physiological response patterns. However, if USACIDC field examiners rendered NO classifications in over 30% of the examinations conducted, it would be a disservice to the person undergoing the examination and to USACIDC field elements. The logic in utilizing the spot-totaling numerical evaluation procedure for criminal specific examinations is apparent for USACIDC. This study was consistent with the USAMPS review which found that the costs associated with the artificially created NO were too great when taking into consideration that the actual occurrence of the false positive is minimal. As clearly demonstrated in this study, the cumulative evaluation procedure failed to conclusively identify numerous deceptive persons, to include six sex offenders, 12 felons involved in crimes against property, four drug offenders, and five other felons involved in crimes against persons (these examples were extracted from the 1991 data). In a criminal specific issue examination, the concern for the false positive exists in the federal government, but there is a paucity of research which indicates that its actual occurrence is significant. The costs of the cumulative evaluation procedure on the utility and accuracy of the PDD process is high, and is considered too significant for an agency supporting a law enforcement mission to accept.

# References

Backster, C. ( 1961) Report on Polygraph Technique Trends. Speech given at the 1961 Annual ASI Report on Polygraph Technique Trends. New York, NY.

Ben-Shakhar, G. & Furedy, J. (1990). *Theories and Applications in the Detection of Deception*. New York: Springer-Verlag.

Ben-Shakhar, G., Lieblich, I., and Bar-Hillel, M. (1982). An evaluation of polygrapher's judgments: A review from a decision theoretic perspective. *Journal of Applied Psychology*, 67, 701-713.

Brisentine, R. (1991). Personal communication. Former Director United States Army Criminal Investigation Command, Dundalk, MD Polygraph Program, personal conversation.

Capps, M. H., & Ansley, N. (1992). Analysis of federal polygraph charts by spot and chart total. *Polygraph*, 21(2), 110-131.

Capps, M. (1990). Evaluation of Two Federal Zone Comparison Score Systems. Unpublished research.

Cole, R. (1991). Personal communication. Instructor Department of Defense Polygraph Institute, Fort McClellan, AL.

Decker, R. (1991). Personal communication. Former Director, United States Army Military Police School, Fort McClellan, AL.

Mason, P. L. (1998). Polygraph validity with urinalysis results as ground truth. *Polygraph*, 27(1), 45-48.

Patrick, C. & Iacono, W. (1991). Validity of the Control Question Polygraph Test: The Problem of Sampling Bias. *Journal of Applied Psychology*, 76, 229-238.

Sneed, E. (1991). Personal communication. Former Instructor, Department of Defense Polygraph Institute, Fort McClellan, AL.

Weaver, R. (1985). Effects of differing numerical chart evaluation systems on polygraph examination results. *Polygraph*, 14, 1, 90 - 104.

# Attachment A

The following are selected summations of the actual polygraph examinations extracted from the USACIDC files. These examinations were originally opined to be DI utilizing the spot-total analysis evaluation procedure. When applying the cumulative evaluation procedure, a NO or a false negative determination resulted for each of the following examinations.

**Case 1. Homicide**

During a failed drug transaction, a drug dealer produced a .45 caliber automatic pistol and shot one of the two persons attempting to make the drug purchase. The victim died as a result of a head wound. The other person attempting to purchase the drugs fled the area unharmed. The drug dealer, who did the shooting, was arrested for the murder. Additional information was developed that another person assisted and conspired with the drug dealer to kill the two persons during the drug transaction. The drug dealer who shot and killed the individual declined to talk about the incident. The suspected co-conspirator consented to undergo the examination. During the post test phase, the individual admitted to planning to "rip-off" the victim by using the .45 automatic. He further admitted to assisting the shooter in the actual robbery.

|  | Spot I Question #5 | Spot II Question #7 | Spot III Question #10 |
|---|---|---|---|
| Spot Scores | +2 | +4 | -4 |

Cumulative total = +2

## Case 2.  Child molesting

Two daughters alleged that their father had been sexually molesting them for several years. The children had made the same allegations three years earlier, but the allegations could not be substantiated and the children continued to reside with their father.  Based upon the new allegations made to a social worker by the children, the father underwent a polygraph examination. The father confessed after the examination to having committed the sexual acts reported by the two daughters.

|  | Spot I<br>Question #5 | Spot II<br>Question #7 | Spot III<br>Question #10 |
|---|---|---|---|
| Spot Scores | +2 | -3 | -3 |

Cumulative total = -4

## Case 3.  Theft

During the U.S. military operation "Just Cause," several barracks rooms on a U. S. Army installation were broken into, and over $5,000.00 worth of property was stolen.  The rooms belonged to soldiers deployed to Panama in support of operation Just Cause.  A soldier that was not deployed was a suspect, and he consented to a polygraph examination.  After failing the examination, the soldier admitted to stealing the master keys to the rooms and to assisting three other soldiers in stealing the property.

|  | Spot I<br>Question #5 | Spot II<br>Question #7 | Spot III<br>Question #10 |
|---|---|---|---|
| Spot Scores | +7 | -4 | +1 |

Cumulative total = +4

## Case 4.  Sexual harassment

A female alleged that, while she was at work, a male employee sexually assaulted her by kissing her without her permission and, subsequently, by touching her around the private parts of her body.  She alleged that all of these acts occurred against her will.  A polygraph examination was conducted of the male employee who confessed after testing to having fondled the victim as alleged.

|  | Spot I<br>Question #5 | Spot II<br>Question #7 | Spot III<br>Question #10 |
|---|---|---|---|
| Spot Scores | +1 | +2 | -3 |

Cumulative total = 0

**Case 5.  Fraud**

A person reported that jewelry and money in the amount of $350.00, was stolen from her room while she was on vacation.  The facts surrounding the theft, as related by the complainant, were not consistent with those related by witnesses.  Subsequent to the polygraph examination, the alleged victim confessed to having fabricated the theft of the property in order to make a false claim for reimbursement.

|  | Spot I<br>Question #5 | Spot II<br>Question #7 | Spot III<br>Question #10 |
|---|---|---|---|
| Spot Scores | +7 | 0 | -4 |

Cumulative total = +3

**Case 6.  Rape/Indecent Assault**

A female alleged that while at a party she became highly intoxicated and passed out.  She stated that when she woke up she found that her panties had been removed.  The female provided the name of a male soldier who was present in the room prior to her passing out.  A polygraph examination was administered of the male who subsequently admitted to engaging in sexual intercourse with the female against her will.

|  | Spot I<br>Question #5 | Spot II<br>Question #7 | Spot III<br>Question #10 |
|---|---|---|---|
| Spot Scores | +7 | +4 | -4 |

Cumulative total = +7

**Case 7.  Wrongful Use of a Controlled Substance (Urinalysis)**

During a command directed urinalysis test, a soldier rendered a urine sample which tested positive for the presence of marihuana.  The soldier stated he was given and smoked a cigar that could possibly have contained marihuana.  The soldier denied knowing the cigar contained marihuana at the time he smoked the cigar.  The soldier consented to a polygraph examination and later admitted he knowingly consumed marihuana.

|  | Spot I<br>Question #5 | Spot II<br>Question #7 | Spot III<br>Question #10 |
|---|---|---|---|
| Spot Scores | +2 | -4 | +3 |

Cumulative total = +1

# Numerical Scoring Systems in the Triad of Matte Polygraph Techniques

## James Allan Matte

The Matte family of polygraph techniques consist of the Matte Quadri-Track Zone Comparison Technique which is a single-issue test, the Matte Quinque-Track Zone Comparison Technique which is an exploratory multiple-issue test, and the Matte Suspicion-Knowledge-Guilt (S-K-G) test designed to identify the examinee who has major involvement, some direct involvement, or guilty knowledge regarding a specific issue.

The numerical quantification system in the analysis of the physiological data used in the aforementioned triad of Matte polygraph techniques employs the Backster chart interpretation rules (See pages 398-406, Matte 1996), with some minor changes described herein.

To attain an objective measure of the reactions or lack of reaction to each relevant and control question in each of the three tracings (pneumograph, electrodermal, cardiograph), the numerical scoring system designed by Cleve Backster (Backster 1963) is used in the triad of Matte polygraph techniques. This system provides the forensic psychophysiologist (FP) with a means of objectively evaluating each relevant question versus its neighboring control question, hereafter referred to as a spot (control vs. relevant), in each tracing according to chart interpretation rules with penalties for violation of those rules, by the assignment or scoring of each spot with a number from a seven-position scale.

Value
| | | |
|---|---|---|
| +3 | MT | Maximum Truthful Score |
| +2 | T | Truthful Score |
| +1 | t | Minimum Truthful Score |
| 0 | ? | |
| -1 | d | Minimum Deception Score |
| -2 | D | Deception Score |
| -3 | MD | Maximum Deception Score |

Numbers preceded by a minus sign fall into the deceptive area; numbers preceded by a plus sign fall into the truthful area.

The following are departures from the Backster Rules. Most of these departures were implemented prior to 1980 and all of them were in effect prior to the validity study of the Matte Quadri-Zone Comparison Technique (subsequently renamed Quadri-Track ZCT) published in 1989, with the exception that as a result of this validity study, the threshold or minimum score required to reach a conclusion for the truthful in the Matte Quadri-Track ZCT was reduced from +4 per chart to +3 per chart.

### Backster "Either-Or" Rule:

To arrive at an interim spot analysis tracing determination of (+2) or (-2) there must be a significant and timely tracing reaction in either the red zone (relevant) or green zone (control) being compared.

Backster ZCT: If the red zone indicates a lack of reaction, it should be compared with the neighboring green zone containing the larger

timely reaction. If the red zone indicates a timely and significant reaction it should be compared with the neighboring green zone containing no reaction or the least reaction.

Matte QTZCT: Red Zone is always compared with the green zone question preceding it. the red zone questions are switched in position with each chart conducted, thus are alternately compared against each green zone question. (Matte 1996)

Tracings Included: respiration, electrodermal, and cardiograph.

**Backster "Green Zone 'Yes' Answer Penalty Rule:**

If a "yes answer is given to a green zone question which is a reversal of the answer given during the pretest question review, that green zone cannot be used as a spot analysis 'presence-of-reaction" zone.

A green zone involving such an answer reversal can be used as a spot analysis "lack-of-reaction" zone where no reaction, or a reaction significantly smaller than the red zone reaction, is indicated.

Backster ZCT: Such use should be avoided if another adjacent lack-of reaction green zone, properly answered, is available.

Matte QTZCT: The forensic psychophysiologist cannot jump to another Track to make a comparison. See Figure 3 for Primary, Secondary, and Inside Tracks.

Tracings Included: respiration, electrodermal, and cardiograph.

**Backster "Green Zone Abuse" Rule:**

If the intensity of a green zone reaction appears to be at least four times as dramatic as a minor reaction in the red zone, it is not proper to feature the minor red zone reaction and compare it with the other neighboring green zone which may show a lesser reaction or no reaction.

Matte QTZCT: This rule does not apply to the Quadri-Track or Quinque-Track ZCT

inasmuch as the FP must compare each red zone question to the preceding green zone question, and cannot jump to another track to make a comparison. (Matte 1996)

Tracings Included: respiration, electrodermal, and cardiograph.

**Backster "Question Pacing" Upgrading Rule:**

To upgrade a ((+2) or (-2) interim spot analysis rating to a (+3) or (-3) final spot rating each of the two zones being inter-compared must embrace a minimum of twenty seconds and a maximum of thirty-five seconds.

Note: Question pacing is measured from the first word of one question to the first word of the question that follows.

Matte QTZCT: Requires a minimum of twenty-seconds between the answer to a test question to the commencement of the next question that follows it, not to exceed thirty seconds.

Tracings Included: respiration, electrodermal, and cardiograph.

Matte "Dual-Equal Strong Reaction" Rule: An Exception to the Backster's One-to-One Rule.

When the red and green zones being inter-compared both contain timely, specific and significant reactions of maximum and equal strength, a minus one (-1) score is assigned to that spot.

Tracings Included: respiration, electrodermal, and cardiograph.

Note: When there is a presence of mild reaction which would warrant only a minimum score of -/+1 in both the relevant question and its neighboring control question respectively of equal magnitude, such as in Figure 1, where there is no presence of parasympathetic activation (questions 46-33) a numerical value of zero must be assigned to this spot in the respiration tracing. However when there is a presence of strong reaction which would be manifested by distinct

activation of both sympathetic and parasympathetic systems in both zones being inter-compared of equal magnitude (Figure 2), a minimum deception score of -1 must be given to this spot. This rule applies only to the pneumograph and cardiograph tracings, not the electrodermal. The electrodermal tracing is excluded because it is more volatile and sensitive to extraneous stimuli.

The aforesaid rule is based on the premise that both zone questions appear to be equally threatening to the examinee, the degree of threat being proportionate to the degree of the responses, which indicate that while the examinee may be attempting deception to the relevant question, its neighboring control question may be too intense due to faulty structure, embraces an equally or more serious unknown crime, or a countermeasure attempt was made. A sophisticated guilty examinee may be able to cause a reaction on the control question but cannot control an oncoming reaction to the relevant question.

**Figure 1**
**Equally Mild Reactions**



1 -
46 (control)

1 1 -
33 (relevant)

1
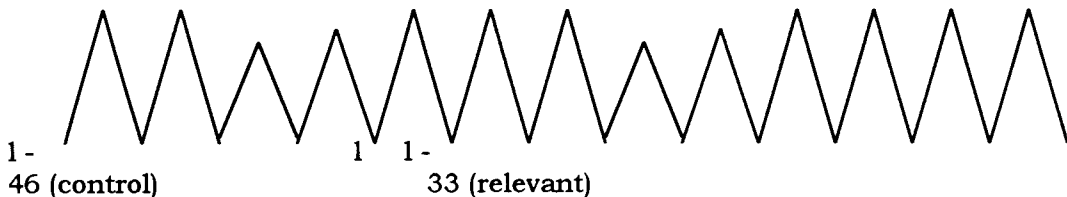
**Figure 2**
**Equally Strong Reactions**



1 1 -
46 (control)

1 1 -
33 (relevant)
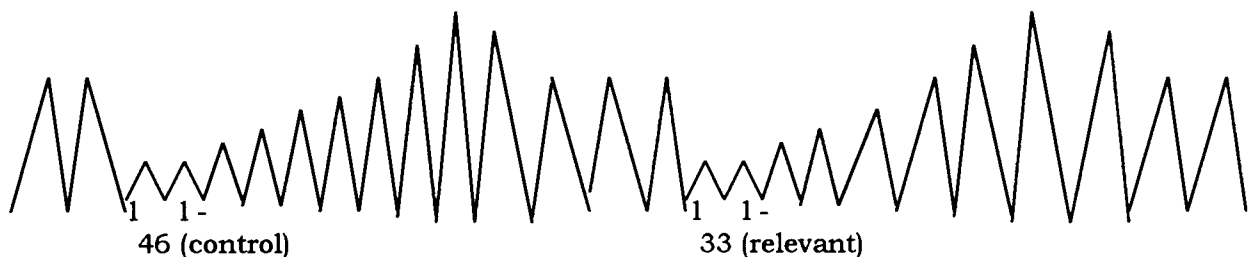
Appendix 1 depicts the Matte Quadri-Track Zone Comparison Test structure which shows that the vertical score tallied from spots 1, 2 and 3 are combined for a total score, inasmuch as all spots deal with the same single issue. Appendix 2 depicts the Tri-Spot Quantification System for the Quadri-Track ZCT. and Figure 3 shows the Conclusion Table from which a determination is made as to Truth, Indefinite (Inconclusive), or Deception from the total scores tallied from spots 1, 2, and 3. It should be noted that in the Matte Quadri-Track Zone Comparison Technique and the Matte Suspicion-Knowledge-Guilt (SKG) tests, a minimum of two polygraph charts (tests) must be conducted to reach a conclusion of truth, deception or inconclusive, and in the Matte Quinque-Track Zone Comparison Technique, a minimum of three polygraph charts must be

conducted to reach a conclusion. This requirement increases external reliability. The American Polygraph Association standards require that a minimum of two charts be conducted to reach a determination of truth or deception. (See Appendices 1 and 2 for the Matte Quadri-Track Zone Comparison Test Structure and Tri-Spot Quantification System).

**Figure 3**

| | CONCLUSION TABLE |
|---|---|
| RESULTS FOR 1 CHART | CIRCLE APPROPRIATE NUMBER BELOW |
| | +27 to +3    +2 to -4    -5 to -27 |
| | TRUTH         INDEFINITE   DECEPTION |
| RESULTS FOR 2 CHARTS | CIRCLE APPROPRIATE NUMBER BELOW |
| | +54 to +6    +5 to -9    -10 to -54 |
| | TRUTH         INDEFINITE   DECEPTION |
| RESULTS FOR 3 CHARTS | CIRCLE APPROPRIATE NUMBER BELOW |
| | +81 to +9    +8 to -14    -15 to -81 |
| | TRUTH         INDEFINITE   DECEPTION |
| RESULTS FOR 4 CHARTS | CIRCLE APPROPRIATE NUMBER BELOW |
| | +108 to +12    +11 to -19    -20 to -108 |
| | TRUTH         INDEFINITE   DECEPTION |

**Matte Quinque-Track Zone Comparison Technique:**

Appendix 3 shows the Matte Quinque-Track Zone Comparison Technique test structure. The Matte Quinque-Track ZCT is an Exploratory Multiple-Issue test, thus all four spots are vertically and independently scored but cannot be horizontally tallied because each spot deals with a different issue. A minimum of three charts (ideally four) must be conducted to attain a minimum number of comparisons for each spot. Appendix 4 is the Matte Quinque-Track ZCT Quantification system, including its Conclusion Table.

**Matte Suspicion-Knowledge-Guilt (SKG) Test:**

The Matte Suspicion-Knowledge-Guilt Test, hereafter referred to as the S-K-G test is designed to provide the forensic psychophysiologist with a single test capable of identifying the examinee who has major involvement, some direct involvement, or guilty knowledge, yet containing similar controls to that found in the Matte Quadri-Track Zone Comparison Technique. For detailed information about the SKG test, consult Matte (1996, chapter 17).

Appendix 5 shows the SKG Test structure and quantification system. It should be noted that test question number 31 is treated as a control question. Question number 31, which relates to unfounded suspicion on the part of the examinee, is not a relevant question but rather a control, inasmuch as it can be readily assumed that an examinee who shows no reaction to any of the relevant questions but shows a reaction to suspicion can be excluded as a participant or witness in the crime. Therefore, the forensic psychophysiologist can choose between control question 48, which is a control question which encompasses both periods normally covered by control questions 46 and 47 in the Quadri-Track ZCT, and control question 31 for the greatest physiological evidence of sympathetic and parasympathetic action, which the FP then uses as the control question to compare against relevant questions 42, 34, 33, and 32 individually.

There will be few occasions when question 31 (control) will exceed control question 48 in overall sympathetic/parasympathetic activity. However, its inclusion is necessary to offset the chance that an examinee's suspicion of someone may be so strong that, without the suspicion question

on the test, he or she may show reaction on the knowledge question because there is no other question in that general category to relieve the energy.

Figure 4 below shows the Comparison Score Table.

## Figure 4

### Comparison Score Table

| Relevants | 42 | 34 | 33 | 32 | 24 |
|---|---|---|---|---|---|
| Controls | 48or31 | 48or31 | 48or31 | 48or31 | 23 |
| Tally | (+/- ) | (+/- ) | (+/- ) | (+/- ) | (+/- ) |

Figure 5 below reflects the SKG Conclusion Table. It should be noted that a minimum of two polygraph charts, as in any test, must be conducted to insure external reliability, before a determination of truth or deception can be rendered.

## Figure 5

### S-K-G Conclusion Table

Results For 1 Chart: Circle Appropriate Number Below

| +2 Or More Truth | +1 To -2 Inconclusive | -3 Or More Deception |
|---|---|---|

Results For 2 Charts: Circle Appropriate Number Below

| +4 Or More Truth | +3 To -5 Inconclusive | -6 Or More Deception |
|---|---|---|

## References

Backster, C. (1963). *Backster Standardized Polygraph Notepack and Technique Guide.* New York: Backster School of Lie Detection.

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

Matte, J. A. (1996). *Forensic Psychophysiology Using the Polygraph: Scientific Truth Verification - Lie Detection.* Williamsville, New York: J.A.M. Publications.

Matte, J. A. (1989). A field validation study of the Quadri-Zone Comparison Technique. *Polygraph,* 18(4), 187-202.

Matte, J. A. and Reuss, R. M. (1989). Validation Study on the Quadri-Zone Comparison Technique. Research Abstract, LD 01452, Vol. 1502, 1989, University Microfilm International.

# Appendix 1

## MATTE QUADRI-TRACK ZONE COMPARISON TEST STRUCTURE



(Cannot jump track to make comparison)

| | OUTSIDE TRACK | | PRIMARY TRACK | | SECONDARY TRACK | | INSIDE TRACK | | OUTSIDE TRACK |
|---|---|---|---|---|---|---|---|---|---|
| PNEUMO TRACING | | | | | | | | | |
| ELECTRODERMAL (GSR/GSG) TRACING | Neutral, Irrelevant Question / Sacrifice/Preparatory Relevant Question | Symptomatic Question (Outside Issue) | Reviewed, Exclusive Control Question | Strong Relevant Question | Reviewed, Exclusive Control Question | Strong Relevant Question | Fear of Error Control Question | Hope of Error Relevant Question | Symptomatic Question (Outside Issue) |
| CARDIO TRACING | | | | | | | | | |
| QUESTION NUMBER | 14J    39 | 25 | 46 | 33 | 47 | 35 | 23 | 24 | 26 |
| COLOR CODE | Y    YR | B | G | R | G | R | Gw | Rw | B |
| TRI-ZONE COMPARISON | | ZONE | ZONE | ZONE | ZONE | ZONE | ZONE | ZONE | ZONE |

**SPOT ONE ANALYSIS SCORE** +-

**SPOT TWO ANALYSIS SCORE** +-

**SPOT THREE ANALYSIS SCORE** +-

COLOR LEGEND:
B    Symptomatic (Outside Issue)
G    Exclusive Control Question
R    Relevant Question (Strong)
w    Indicates Zone is influenced by Zones in Spots #1 and #2
Gw   Inside Issue Control Question (Variable strength)
Rw   Inside Issue Relevant Question (Variable strength)
YR   Sacrifice Relevant Question
Y    Neutral Question (Irrelevant)

ZONES
1. Black (Symptomatic)
2. Green (Exclusive Control)
3. Red (Strong Relevant)
Note: White (w) suffix to a Zone places that Zone in the *Inside Track* to recoup response scores lost as a result of an *Inside Issue*.
TRACK Identifies a pair of questions related for comparison/quantification (G & R Zone) or evaluation (B Zone).

THREE SPOTS SCORED AND TALLIED FOR
A GRAND TOTAL = TRUTH, DECEPTION, INCONCLUSIVE

SPOT Identifies a Track which is quantified.

© 1995 by James Allan Matte

# Appendix 2

Matte Quadri-Track Zone Comparison Test

Tri-Spot Quantification System

| SPEC-1 | SPOT ONE | | | | | | SPOT TWO | | | | | | SPOT THREE | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TRUTH | INDEF | DECEP | | | | TRUTH | INDEF | DECEP | | | | TRUTH | INDEF | DECEP | |
| PNE(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |
| GSR(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |
| CAR(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |
| **SPEC-2** | | | | | | | | | | | | | | | | |
| | TRUTH | INDEF | DECEP | | | | TRUTH | INDEF | DECEP | | | | TRUTH | INDEF | DECEP | |
| PNE(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 ·2 | +1 0 -1 | -2 -3 | = ( ) |
| GSR(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |
| CAR(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |
| **SPEC-3** | | | | | | | | | | | | | | | | |
| | TRUTH | INDEF | DECEP | | | | TRUTH | INDEF | DECEP | | | | TRUTH | INDEF | DECEP | |
| PNE(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |
| GSR(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |
| CAR(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |
| **SPEC-4** | | | | | | | | | | | | | | | | |
| | TRUTH | INDEF | DECEP | | | | TRUTH | INDEF | DECEP | | | | TRUTH | INDEF | DECEP | |
| PNE(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |
| GSR(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |
| CAR(33) | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (35) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) | (24) | | +3 +2 | +1 0 -1 | -2 -3 | = ( ) |

TARGET ( )       TOTAL:                ( )                         TOTAL:        ( )                         TOTAL: ( )

GRAND TOTAL: ( )

FOR ( ) CHARTS.

# Appendix 3

## MATTE QUINQUE-TRACK ZONE COMPARISON TEST STRUCTURE
### (EXPLORATORY)



| | | OUTSIDE TRACK | FIRST TRACK | | SECOND TRACK | | THIRD TRACK | | FOURTH TRACK | | OUTSIDE TRACK |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PNEUMO TRACING | | | | | | | | | | | |
| ELECTRODERMAL (GSR/GSG) TRACING | | Symptomatic Question (Outside Issue) | Non-Current, Exclusive Control Question | Relevant Question | Non-Current, Exclusive Control Question | Relevant Question | Non-Current, Exclusive Control Question | Relevant Question | Suspicion – Control Question | Knowledge – Relevant Question | Symptomatic Question (Outside Issue) |
| CARDIO TRACING | | | | | | | | | | | |
| QUESTION NUMBER | 14J | 39 | 25 | 46 | 45J | 47 | 45K | 48 | 45L | 031 | 032 | 26 |
| COLOR CODE | Y | YR | B | G | R | G | R | G | R | G | R | B |
| TRI-ZONE COMPARISON | | ZONE | ZONE | ZONE | ZONE | ZONE | ZONE | ZONE | ZONE | ZONE | ZONE | ZONE |

(Outside track labels: Neutral, Irrelevant Question / Sacrifice/Preparatory Relevant Question)

(Cannot jump track to make comparison)

**COLOR LEGEND:**
- B   Symptomatic (Outside Issue)
- G   Exclusive Control Question
- R   Relevant Question (Strong)
- YR  Sacrifice Relevant Question
- Y   Neutral Question (Irrelevant)
- O   Prefix - Option Use.

**ZONES**
1. Black (Symptomatic)
2. Green (Exclusive Control)
3. Red (Strong Relevant)

TRACK  Identifies a pair of questions related for comparison (G&R Zone) or evaluation (B Zone)

Note: Fourth Track includes Suspicion Question 031 Green Zone versus Knowledge Question 032 Red Zone.

SPOT ONE ANALYSIS SCORE  +-
SPOT TWO ANALYSIS SCORE  +-
SPOT THREE ANALYSIS SCORE  +-
SPOT FOUR ANALYSIS SCORE  +-

SPOT Identifies a Track which is quantified.

ABOVE FOUR SPOTS ARE VERTICALLY AND INDEPENDENTLY SCORED BUT CANNOT BE HORIZONTALLY SCORED FOR A TOTAL TALLY BECAUSE EACH SPOT DEALS WITH A DIFFERENT ISSUE. A MINIMUM OF THREE CHARTS (IDEALLY FOUR) MUST BE CONDUCTED TO ATTAIN A MINIMUM NUMBER OF COMPARISONS FOR EACH SPOT.

## Appendix 4

### MATTE QUINQUE-TRACK ZONE COMPARISON TECHNIQUE

| TARGET (    ) | | USED ON CHART NR. | |
|---|---|---|---|
| 14J | WERE YOU BORN IN THE U.S.? | 13L (Last Name) | 13F (First Name) |
| 39 | RE: WHETHER OR NOT (YOURSELF) | DO YOU INTEND TO ANSWER TRUTH-TRUTHFULLY EACH QUESTION ABOUT THAT? | |
| 25 | ARE YOU COMPLETELY CONVINCED THAT I WILL NOT ASK YOU AN UNREVIEWED QUESTION DURING THIS CHART? | | |
| 46 | BETWEEN THE AGES OF (    ) AND (    ) - DO YOU REMEMBER: | | |
| 45J | | | |
| 47 | DURING THE FIRST (    ) YEARS OF YOUR LIFE - DO YOU REMEMBER: | | |
| 45K | | | |
| 48 | DURING THE FIRST (    ) YEARS OF YOUR LIFE - DO YOU REMEMBER: | | |
| 45L | | | |
| 031 | | | |
| 032 | | | |
| 26 | IS THERE SOMETHING ELSE YOU ARE AFRAID I WILL ASK YOU A QUESTION ABOUT, EVEN THOUGH I TOLD YOU I WOULD NOT? | | |

### QUANTIFICATION SYSTEM
### QUINQUE-TRACK ZONE COMPARISON TECHNIQUE

| CHART1 | TI | INDEF | DI | | | TI | INDEF | DI | | | TI | INDEF | DI | | | TI | INDEF | DI | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PO 45J | +3+2 | +1 0-1 | -2-3 | ( | )45K | +3+2 | +1 0-1 | -2-3 | ( | )45L | +3+2 | +1 0-1 | -2-3 | ( | )32 | +3+2 | +1 0-1 | -2-3 | ( ) |
| GR 45J | +3+2 | +1 0-1 | -2-3 | ( | )45K | +3+2 | +1 0-1 | -2-3 | ( | )45L | +3+2 | +1 0-1 | -2-3 | ( | )32 | +3+2 | +1 0-1 | -2-3 | ( ) |
| CO 45J | +3+2 | +1 0-1 | -2-3 | ( | )45K | +3+2 | +1 0-1 | -2-3 | ( | )45L | +3+2 | +1 0-1 | -2-3 | ( | )32 | +3+2 | +1 0-1 | -2-3 | ( ) |

| CHART2 | TI | INDEF | DI | | | TI | INDEF | DI | | | TI | INDEF | DI | | | TI | INDEF | DI | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PO 45J | +3+2 | +1 0-1 | -2-3 | ( | )45K | +3+2 | +1 0-1 | -2-3 | ( | )45L | +3+2 | +1 0-1 | -2-3 | ( | )32 | +3+2 | +1 0-1 | -2-3 | ( ) |
| GR 45J | +3+2 | +1 0-1 | -2-3 | ( | )45K | +3+2 | +1 0-1 | -2-3 | ( | )45L | +3+2 | +1 0-1 | -2-3 | ( | )32 | +3+2 | +1 0-1 | -2-3 | ( ) |
| CO 45J | +3+2 | +1 0-1 | -2-3 | ( | )45K | +3+2 | +1 0-1 | -2-3 | ( | )45L | +3+2 | +1 0-1 | -2-3 | ( | )32 | +3+2 | +1 0-1 | -2-3 | ( ) |

| CHART3 | TI | INDEF | DI | | | TI | INDEF | DI | | | TI | INDEF | DI | | | TI | INDEF | DI | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PO 45J | +3+2 | +1 0-1 | -2-3 | ( | )45K | +3+2 | +1 0-1 | -2-3 | ( | )45L | +3+2 | +1 0-1 | -2-3 | ( | )32 | +3+2 | +1 0-1 | -2-3 | ( ) |
| GR 45J | +3+2 | +1 0-1 | -2-3 | ( | )45K | +3+2 | +1 0-1 | -2-3 | ( | )45L | +3+2 | +1 0-1 | -2-3 | ( | )32 | +3+2 | +1 0-1 | -2-3 | ( ) |
| CO 45J | +3+2 | +1 0-1 | -2-3 | ( | )45K | +3+2 | +1 0-1 | -2-3 | ( | )45L | +3+2 | +1 0-1 | -2-3 | ( | )32 | +3+2 | +1 0-1 | -2-3 | ( ) |

TARGET (    )      TOTAL (    )      TOTAL (    )      TOTAL (    )      TOTAL (    )

### CONCLUSION TABLE

| | NDI | INCL | DI | 45J | 45K | 45L | 032 |
|---|---|---|---|---|---|---|---|
| THREE CHARTS | +27 to +5 | +4 to -7 | -8 to -27 | ( ) | ( ) | ( ) | ( ) |
| FOUR CHARTS | +36 to +7 | +6 to -9 | -10 to -36 | ( ) | ( ) | ( ) | ( ) |
| FIVE CHARTS | +45 to +9 | +8 to -11 | -11 to -45 | ( ) | ( ) | ( ) | ( ) |

Each of the 4 SPOTS (45 series and 032) are scored, tallied and evaluated separately against the Conclusion Table according to the number of charts conducted.

STIMULATION TEST DATA - NUMBER SELECTED:_____CHART NR:_____

# Appendix 5

| S-K-G Test | 13L | | Used on Chart Nr. | |
|---|---|---|---|---|
| 25 | Are you completely convinced that I will not ask you an unreviewed question during this chart? | | | |
| 39 | Regarding the: | Do you intend to answer truthfully each question about that? | | |
| 42 | Same as above | Before that_____occurred - Did you definitely know it was about to happen? | | |
| 34 | Same as above | At the very time that_____ occurred - were you _____ (on the scene) | | |
| 48 | During the first (    ) years of your life - Do you remember: | | | |
| 33 | Same as above | Did you (yourself) | | |
| 32 | Same as above | Do you know for sure (who) | | |
| 31 | Same as above | Do you suspect anyone in particular of | | |
| 23 | Are you afraid an error will be made on this test regarding: (whether or not you were involved in this crime) | | | |
| 24 | Are you hopeful an error will be made on this test regarding: (whether or not you were involved in this crime) | | | |
| 26 | Is there something else you are afraid I will ask you a question about, even though I told you I would not? | | | |

## S-K-G QUANTIFICATION SYSTEM SCORE TABLE

SKG-1

| PNE 42 Score (    ) | 34 Score (    ) | 33 Score (    ) | 32 Score (    ) | 24 Score (    ) |
|---|---|---|---|---|
| GSR 42 Score (    ) | 34 Score (    ) | 33 Score (    ) | 32 Score (    ) | 24 Score (    ) |
| CAR 42 Score (    ) | 34 Score (    ) | 33 Score (    ) | 32 Score (    ) | 24 Score (    ) |
| SKG-2 | | | | |
| PNE 42 Score (    ) | 34 Score (    ) | 33 Score (    ) | 32 Score (    ) | 24 Score (    ) |
| GSR 42 Score (    ) | 34 Score (    ) | 33 Score (    ) | 32 Score (    ) | 24 Score (    ) |
| CAR 42 Score (    ) | 34 Score (    ) | 33 Score (    ) | 32 Score (    ) | 24 Score (    ) |

42 Total (    )  34 Total (    )  33 Total (    )  32 Total (    )  24 Total (    )

## S-K-G CONCLUSION TABLE

| | TRUTH | INCONCLUSIVE | DECEPTION |
|---|---|---|---|
| RESULTS FOR 1 CHART: | +2 or more | +1 to -2 | -3 or more |
| RESULTS FOR 2 CHARTS: | TRUTH | INCONCLUSIVE | DECEPTION |
| | +4 or more | +3 to -5 | -6 or more |

S-K-G CONCLUSIONS: TRUTHFUL to relevant question(s) _____

DECEPTIVE to relevant question(s)_____

TARGET (    )          INCONCLUSIVE to relevant question(s)_____

# The Academy for Scientific Investigative Training's Horizontal Scoring System and Examiner's Algorithm System for Chart Interpretation©

## Nathan J. Gordon

## Abstract

The Horizontal Scoring System was developed at the Academy for Scientific Investigative Training in 1981 as a means of eliminating subjectivity in scoring and the skewing of test results by the subjective selection of control questions for comparison to the relevant question.  It allows for objective numerical chart analysis, while eliminating the subjectivity of the assignment of numbers from the traditional seven-position scale, and the additional subjectivity of the selection of which control question to use for comparison to the relevant question.  The Examiner Algorithm System was developed at the Academy in 1997, and utilizes discrete  measurements which allow the examiner to accurately and consistently determine what constitutes the greatest reaction.

In 1963, Backster developed a numerical scoring system where values ranging from a +3, to a -3, were assigned to each relevant question's independent physiological parameter after it was compared to those same parameters of a control question selected by the examiner. The decision of whether this "control vs. relevant" comparison yielded no difference (0), a slight difference (±1), a clear difference (±2), or a huge vs. no difference (±3), and the corresponding score for the comparison is left to the subjectivity of the individual examiner.    There    are    also differences in opinion on how the examiner should select the control question for comparison to the relevant question (Weaver, 1980).    The Army compares the relevant question to the strongest control question, skewing their test toward truthfulness. Backster compares the relevant question to the weakest control question, skewing the test toward non-truthfulness.  The University of Utah compares the relevant question to the control question preceding it, not skewing the test outcome toward any direction.  Weaver also pointed out that the matter is further complicated in that there are clear differences among these three groups as to what constitutes a reaction.  The Army's position is that any change from the norm constitutes a

reaction, while Backster clearly discriminates between what he defines as reaction and relief, and Utah takes a position somewhere in between.

With so much subjectivity and difference in opinions among major groups in our profession,  it is not difficult to imagine that examiners of different schools of thought could have differences of opinion analyzing the same polygraph charts, even though they are all using "numerical analysis."  These conditions led to our search to eliminate the subjectivity of what number to assign from the traditional seven-position scale, and which control question to select for comparison to the relevant question and to the development of the Horizontal Scoring System (Gordon & Cochetti, 1987).

In the Horizontal Scoring System the subjectivity of control question selection is totally eliminated by the examiner comparing all of the control and relevant questions in each  individual  parameter,  creating  a hierarchy of the greatest to least reaction. For example, in a standard Backster "You Phase" format there is a maximum of three control questions (numbers 46, 47, and 48) and three relevant questions (numbers 33, 35 and 37),

---

making a total of six reactions being compared in each individual parameter. The examiner employing Horizontal Scoring identifies the greatest reaction in thoracic breathing, and assigns it a six (6), the next greatest a five (5), the next greatest a four (4), the next greatest a three (3), the next greatest a two (2), and the smallest reaction a one (1).

The same process is then repeated in the abdominal breathing patterns, the electrodermal patterns and the cardiograph patterns. After all the hierarchies have been established, each question's pneumograph scores (thoracic and abdominal) are averaged, and then added to the electrodermal score and the cardiograph score, resulting in a total question score for each of the control and relevant questions. This total question score is then assigned a plus (+), in each of the control questions, and a minus (-), in each of the relevant questions. Since the Backster "You Phase" is a single-issue technique, the scores can be combined for a total chart score.

In a general question technique, such as Reid, Arther or the MGQT, there are four relevant questions and two control questions of varying weights. The same ranking of six to one can be performed, since one is still comparing six questions in each parameter. If two pneumographs are being used, we would again average them for a single pneumograph score, and again add them to the electrodermal and cardiograph score for each question for a total question score. We would again assign a plus (+) to the control question scores and a minus (-) to the relevant question scores. However, we cannot combine the scores since the relevant questions of inquiry do not represent a single issue. We must now compare each relevant question's total score with the total score of the control question with which it would traditionally be compared. The difference between these scores will represent that relevant question's final score. For example, if relevant question #3 had a total score of -15, we would compare it to control question #6, which had a total score of a +9, and derive the final score for relevant question #3 as a -6: (-15) - (+9) = -6. We would follow the same process to evaluate relevant question #5, and then use the same process to compare relevant questions #8 and 9, to control question #10.

In situations where there are equal reactions between questions in the parameter being scored, we average the scores of the positions they are vying for. For example, we have identified the greatest reaction in a given parameter, and assign it a 6. There are now two equal reactions competing for ranks of 5 and 4. We average those two numbers (5+4 = 9; 9/2 = 4.5) and assign each a 4.5.

The cutoffs we are currently using are ±1.5 per relevant question, per chart. For three charts, with three "single issue" relevant questions we use a ±13, and for three charts with two "single issue" relevant questions a ± 9. For a single question (i.e., relevant question #3 in MGQT) we use a ±3 for two charts, and a ±4.5 for three charts. These numbers need to be reevaluated to determine if adjusting them would result in even more accurate results.

Previous research has shown the Horizontal Scoring System to be highly valid and reliable (Horvath, 1985, and, Driscoll & Honts, 1987). Both studies concluded that Horizontal Scoring was as accurate as traditional scoring, and the latter study stated that Horizontal Scoring was much easier to teach and apply. While Horizontal Scoring succeeded in removing the subjectivity of what number from a seven-position scale to apply, and which control question to select for comparison to the relevant question, it still left the question of what constitutes the "greatest" reaction to the subjectivity of the individual examiner.

In 1997, we devised a mathematical method (algorithm) for measuring changes for each physiological parameter, which in itself reflects the degree and importance of the psychophysiological reactions occurring . The algorithm works independently of any understanding of the psychological or physiological basis of why, or what is actually happening within the parameter. It is designed to simply apply objective mathematical equations to measure what is occurring, and remove examiner subjectivity. Since examiners are applying the same formula to determine the degree of reaction, reliability between examiners' interpretations of the same charts is dramatically increased. Examiner experience in chart interpretation is thereby negated.

## Pneumograph

I initially theorized a method for interpreting the pneumograph that was a dual system monitoring changes in pneumograph suppression (PS) and pneumograph duration (PD). I strongly believe these two reactions reflect the major changes occurring in this parameter:

(PS + PD) = pneumograph reaction.

To confirm my theory, students at the Academy for Scientific Investigative Training were instructed to measure in millimeters the height of each of the breathing cycles in a specific question. The first four cycles following answering distortion were measured and then totaled. Each control and relevant reaction was then assigned a value for the greatest (defined by the smallest total number, which represented the greatest overall suppression) to the least reaction (Horizontal Scoring System). As previously stated, the highest value assigned was determined by the total of the number of relevant and control questions utilized in the polygraph technique being employed for the exam in question. Each pneumograph was similarly scored based on duration, and given scores from greatest to least. To do this, chart time was measured, in millimeters, from the end of exhalation in the answering distortion cycle to the beginning of inhalation in the fifth respiration cycle.

In 1997, Emanuel Cohen, one of the students attending the Academy for Scientific Investigative Training from Israel, listened to my lecture on interpreting the pneumograph based on my formula of PS + PD. He then suggested a simpler mathematical equation for establishing pneumograph reactions:

1. Measure in millimeters the heights of the first four cycles after answering distortion in the tracing being evaluated.

2. Measure the duration of these four cycles from the end of the exhalation cycle containing answering duration, until the beginning of the fifth inhalation cycle.

3. Divide the duration by the total of the four cycle heights in the reaction being evaluated, which in essence gives the amount of suppression and duration in the tracing.

4. The larger the number, the greater the reaction.

In Figure 1, the thoracic respiration has been scored using the algorithm and then assigned numbers from 6 to 1 in accordance with the Horizontal Scoring System. The greatest reaction was to question R12, and it received a 6. Questions R9 and R6 were tied as the next greatest reaction, and since they were vying for horizontal positions (ranks) 5 and 4, they were each given a 4.5, which represents the average of those two positions (5+4 = 9; 9/2 = 4.5). The three control questions, C5, C8 and C11, were also tied for horizontal positions 3, 2 and 1. Each received the average of those scores (3+2+1+ = 6; 6/3 = 2), a 2.

In Figure 2 the abdominal respiration tracing is scored, and the two pneumograph scores for each question will be averaged, so that the pneumograph score only represents one-third of the questions total score.

## Electrodermal

Measurement of the electrodermal recording is performed by multiplying the height by the base of the tracing. The height of the tracing is established by a straight line drawn from the highest peak to the base. The duration is established by measuring the distance of a straight line drawn from the beginning of the reaction, straight out until the point it intersects the downward movement of the tracing in automatic mode. Multiplying these two measurements together reflects tracing amplitude plus duration, and the larger the number, the greater the reaction. In Figure 3, the electrodermal responses are scored.

## Cardiograph

The cardiograph tracing measurement is established by drawing a twenty-second straight line out from the bottom of the cardiograph tracing at the beginning of the

question. A measurement, in millimeters, is made to determine the height of any changes that occur in the baseline of the cardiograph above that line. In essence, we are measuring increases in blood volume. In Figure 4, the cardiograph is scored and tied Horizontal position scores are averaged (R6 & R12, C8 & R9).

In Figure 5, the results are entered onto a Horizontal Scoring System sheet, and since it was a single-issue examination the scores can be combined, for a total chart score of -22.5. The same procedure would then be followed for the other charts in the examination and a total examination score derived.

In conclusion, the Horizontal Scoring System has been successfully used in the field since 1981, and the Examiner Algorithm System since 1997. The Academy's Horizontal Scoring System clearly removes examiner subjectivity in the assignment of numerical values to reactions and the skewing of test results by arbitrary control question selection. This system is also easier to analyze and teach, allowing for broader application. The Examiner Algorithm System simplifies and standardizes the process of determining the degree of a reaction. The reliability and validity of the algorithm in a blind evaluation study will be published later this year.

# References

Driscoll, L.N., & Honts, C.R. (1988). An evaluation of the reliability and validity of rank order and standard numerical scoring of polygraph charts. *Polygraph*, 16(4).

Gordon, N.J., & Cochetti, P.M. (1987). The horizontal scoring system. *Polygraph*, 16(2), 116-125.

Horvath, F. (1985). Personal communication.

Weaver, R. S. (1980). The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. *Polygraph*, 9, 94-108.

**Figure 1**
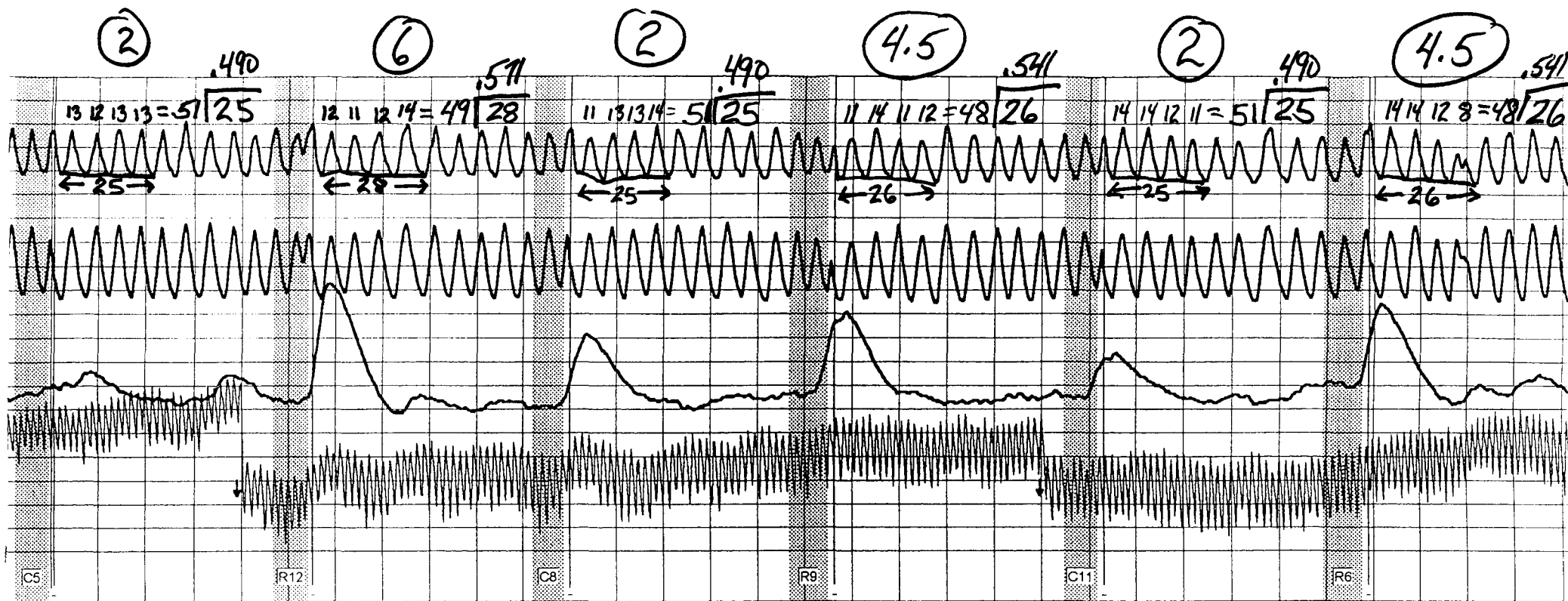
Exam 1

| Gain Settings: | P2 | P1 | GS | CA |
|---|---|---|---|---|
| Recorded: Start | 8.5 | 4.2 | 2.0 | 1.9 |
| Recorded: End | 8.5 | 4.2 | 2.0 | 1.9 |
| Printed: Start | 8.5 | 4.2 | 2.0 | 1.9 |
| Printed: End | 8.5 | 4.2 | 2.0 | 1.9 |

Subject:
Examiner: njg
Date: 1/13/99 View Size: Normal
Time Start: 2:29:44 PM   End: 2:34:09 PM Duration: 04:25
Cuff Pressure Start: 68 End: 70
98-05-01R (Zone)

Subject:
Examiner: njg
Date: 1/13/99 View Size
Time Start: 2:29:44 PM
Cuff Pressure Start: 68
98-05-01R (Zone)

# Figure 2

Subject:
Examiner: njg
Date: 1/13/99 View Size: Normal
Time Start: 2:29:44 PM   End: 2:34:09 PM Duration: 04:25
Cuff Pressure Start: 68  End: 70
98-05-01R  (Zone)

| Gain Settings: | P2 | P1 | GS | CA |
|---|---|---|---|---|
| Recorded: Start | 8.5 | 4.2 | 2.0 | 1.9 |
| Recorded: End | 8.5 | 4.2 | 2.0 | 1.9 |
| Printed: Start | 8.5 | 4.2 | 2.0 | 1.9 |
| Printed: End | 8.5 | 4.2 | 2.0 | 1.9 |

Subject:
Examiner: njg
Date: 1/13/99 View Size
Time Start: 2:29:44 PM
Cuff Pressure Start: 68
98-05-01R  (Zone)

**Figure 3**

Pneumo Averages:
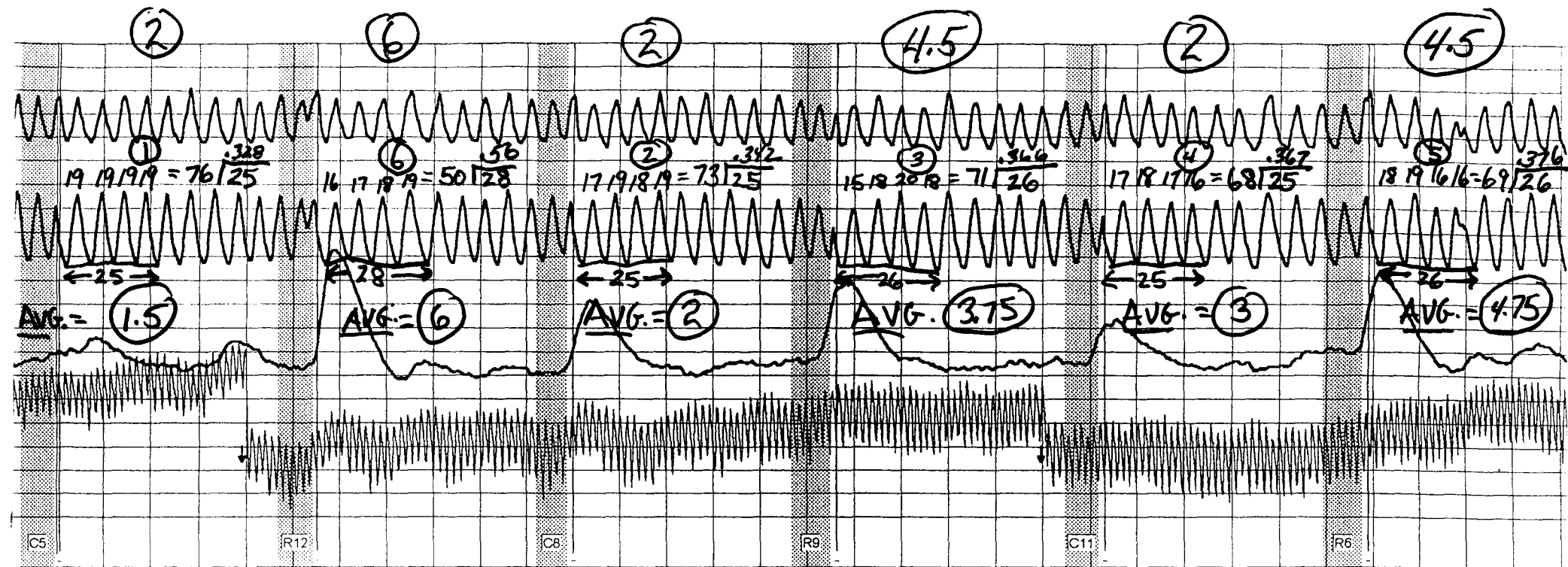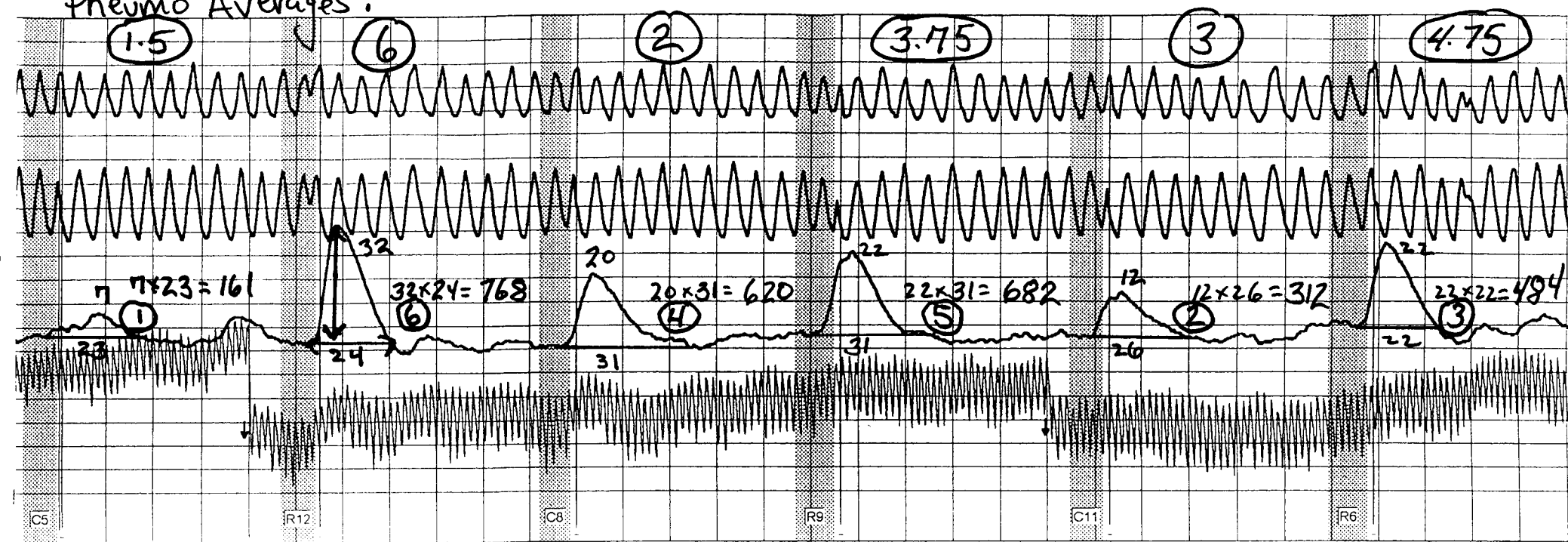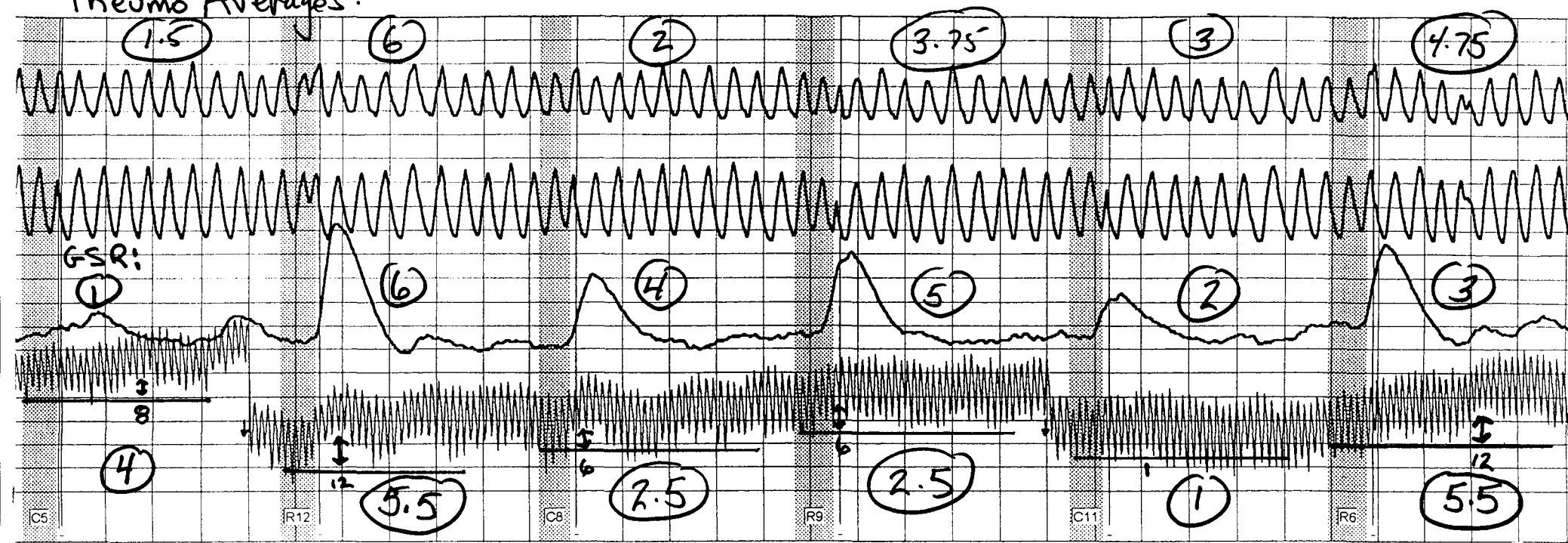
(1.5)   (6)   (2)   (3.75)   (3)   (4.75)

7×23 = 161    32×24 = 768    20×31 = 620    22×31 = 682    12×26 = 312    22×22 = 484

32   20   22   12   22
23   24   31   31   26   22

C5   R12   C8   R9   C11   R6

| | CA | | page 1 of 3 | 99-05-01 | Exam 1 | Gain Settings: | P2 | P1 | GS | CA | | page 2 of 3 | 99-05-01 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1.9 | | | Subject | | Recorded: Start | 8.5 | 4.2 | 2.0 | 1.9 | | | Subject: |
| | 1.9 | | | Examiner: njg | | Recorded: End | 8.5 | 4.2 | 2.0 | 1.9 | | | Examiner: njg |
| | 1.9 | | | Date: 1/13/99 View Size: Normal | | Printed: Start | 8.5 | 4.2 | 2.0 | 1.9 | | | Date: 1/13/99 View Size |
| | 1.9 | | | Time Start: 2:29.44 PM   End: 2:34:09 PM Duration: 04:25 | | Printed: End | 8.5 | 4.2 | 2.0 | 1.9 | | | Time Start: 2:29.44 PM |
| | | | | Cuff Pressure Start: 68 End: 70 | | | | | | | | | Cuff Pressure Start: 68 |
| | | | | 98-05-01R (Zone) | | | | | | | | | 98-05-01R (Zone) |

**Figure 4**

Figure 5

# Academy for Scientific Investigative Training's
## Horizontal Scoring System©

**CHART 1**

| | Q# 5 | Q# 12 | Q# 8 | Q# 9 | Q# 11 | Q# 6 |
|---|---|---|---|---|---|---|
| P1 | 2 | 6 | 2 | 4.5 | 2 | 4.5 |
| P2 | 1 | 6 | 2 | 3 | 4 | 5 |
| Avg. Pneumo | 1.5 | 6 | 2 | 3.75 | 3 | 4.75 |
| E | 1 | 6 | 4 | 5 | 2 | 3 |
| C | 4 | 5.5 | 2.5 | 2.5 | 1 | 5.5 |
| Total | +6.5 | -17.5 | +8.5 | -11.25 | +6 | -13.25 |

**CHART 2**

| | Q# | Q# | Q# | Q# | Q# | Q# |
|---|---|---|---|---|---|---|
| P1 | | | | | | |
| P2 | | | | | | |
| Avg. Pneumo | | | | | | |
| E | | | | | | |
| C | | | | | | |
| Total | | | | | | |

**CHART 3**

| | Q# | Q# | Q# | Q# | Q# | Q# |
|---|---|---|---|---|---|---|
| P1 | | | | | | |
| P2 | | | | | | |
| Avg. Pneumo | | | | | | |
| E | | | | | | |
| C | | | | | | |
| Total | | | | | | |

**OVERALL DETERMINATION**

| Q# | CHART 1 | CHART 2 | CHART 3 | TOTAL SCORE |
|---|---|---|---|---|
| 5 | +.5 | | | |
| 12 | -17.5 | | | |
| 8 | +8.5 | | | |
| 9 | -11.25 | | | |
| 11 | +6 | | | |
| 6 | -13.25 | | | -22.5 |

Determination: _____

# The Control Question Technique:  A Search for
# Improved Decision Rules

## Eitan Elaad

## Abstract

Analysis of polygraph examinations is done with two main methods.  In the semi-objective numerical scoring method, the examiner assigns a numerical score between -3 and +3 for every comparison between the responses to relevant and adjacent control questions.  Then, the numerical scores are summed up to yield a final score.  Another, less systematic method of analyzing the charts, use global evaluations.  Here, the examiner collects reactions which are liable to be of help in reaching a decision.  Then, he or she selects the blatant responses and reach a decision.  The present study was designed to examine these two interpretation methods and to look for better decision rules.  Two samples of verified polygraph examination records, conducted with the common Control Question Technique, were selected for the purposes of the present study.  Results suggested to replace the raw scores of the numerical scoring method with the mean per comparison point scores and combine both interpretation methods.  It was further suggested to use an inconclusive region with boundaries skewed in the direction of the negative scores.

The most preferred polygraph method in field practice is the Control Question Technique (CQT; Raskin, 1989; Reid & Inbau, 1977; Podlesny, 1993).  Briefly, the CQT contains an extensive pre-test interview in which the examinee is given the opportunity to talk about the offense and to present his or her version of the crime.  During this interaction between the examiner and the examinee, the questions are formulated.  Then, the examinee is attached to the polygraph and is asked a series of questions while continuously measuring the various physiological reactions.  The questions are of the following three types: (a) Relevant questions - directly crime-relevant questions of the "did you do it?" type (e.g., "Did you break into Mr. Smith's store last Friday night?"). (b) Control questions - focusing on general, non-specific misconducts, of a nature as similar as possible to the issue under investigation (e.g., "Have you ever taken something valuable without permission?"). (c) Irrelevant questions - focusing on completely neutral issues, (e.g., are you sitting on a chair?).  These are intended to absorb the initial orienting response evoked by any opening question, and

to enable rest periods between the more loaded questions.  Typically, the whole question series is repeated three or four times.

The inference rule underlying the CQT is based on a comparison of the responses evoked by the relevant and control questions.  Deceptive individuals are expected to show more pronounced responses to the relevant questions, whereas truthful individuals are expected to show the opposite pattern of responsivity (i.e., more pronounced responses to the control questions).

Polygraph records conducted with the control question test are usually interpreted according to one of two main methods.  The first is the semi-objective numerical scoring (NS) technique (Barland & Raskin, 1975).  This procedure is a routine at the Polygraph Unit of the Israeli Police.  According to the numerical scoring procedure relevant data is gathered by systematically addressing every comparison point (a single comparison of response magnitude between a relevant and a control question for each physiological measure by one examiner), and numbers (-3,-2,-1, 0, 1, 2,

3) are assigned to each comparison. The absolute value of the assigned number reflects the magnitude of the difference between the responses evoked by the two questions within the pair (e.g., -3 or +3 reflect a very large difference, -1 or +1 reflect a small difference and 0 reflects no difference), and the sign of the assigned number reflects the direction of the difference, such that positive numbers are associated with a pattern of larger physiological reactivity to the control question, and negative numbers reflect the opposite pattern. These numbers are then summed up across question pairs, across physiological measures and across polygraph charts to yield a total score. Thus, if for example a polygraph examination is based on three charts and three physiological measures and if two pairs of relevant-control questions are identified for each chart, then the total score ranges between -54 and +54.

While the first stage of the NS which notes relevant data and gathers it, has the advantage that all possible information is referred to in a reliable manner (when conducted by knowledgeable and well trained polygraph examiners), the second stage, which integrates the information, suffers from oversimplicity. Thus, all three physiological measures are given the same weight, no attention is paid to dynamic changes in the subject's responses during the test session, and the consistency of the responses over time is not considered.

Another method of CQT record analysis is the global record evaluation (GRE) in which the evaluator studies the record and notes any significant changes in response patterns between the relevant and control questions. This type of data analysis presents the opportunity to treat the record as a whole rather than a combination of all its components. However, the GRE is based mainly on the subjective impression of the evaluator, and since evaluators differ in the weight they place on various physiological measures and in their emphasis on the dynamic changes during the test, no definite decision rules are applied and the evaluator makes the data selection while collecting it. Consequently, the final result is based on some few salient cues which pass the selection process and this exhibits relatively poor

interevaluator reliability (Podlesny & Raskin 1977).

The purpose of this article is to introduce some decision rules which exhibit improvement of the CQT detection rate, while using the NS and the GRE techniques, on data collected from two independent, and completely different samples of verified polygraph records.

# Method

Two samples of real life criminal polygraph records, which were verified by confession and/or conviction of the guilty party, were utilized for the purposes of the present analysis. The two samples were randomly drawn from the Israel Scientific Interrogation Unit's pool of verified polygraph tests. One sample consisted of 69 records of severe crimes (SC) (e.g. murder, attempted murder) and the other sample selected 60 records of minor crimes (MC) (e.g. theft, burglary, fraud). The MC sample was divided evenly between actually guilty and actually innocent suspects' records (30 in each) while the SC sample consisted of 51 innocent and 18 guilty suspects' records. All the records were of polygraph examinations conducted by trained field examiners according to control question techniques (Reid & Inbau 1977, Backster 1969). Three physiological measures were present on each test record. The first, recorded thoracic and abdominal respiration. The second recorded skin resistance responses (SRR), and the third recorded the cardio-vascular activity (e.g. blood pressure, blood volume and heart rate).

**Polygraph Record Analysis**

The records were independently analyzed by eight (MC) or ten (SC) experienced polygraph examiners. This was done twice, in two counterbalanced sessions separated in time. In one, the evaluator was asked to make a blind interpretation of the record according to the semi-objective numerical scoring technique. In the other, the same polygraph examiners were asked to analyze the records according to the global record evaluation method. Here, the evaluator was asked to make an overall decision regarding each relevant question in the record. The overall

decision was expressed on a 5-point scale which defines at its extreme points strong confidence in the truthful or deceptive result, at the 2nd and 4th points low confidence in the results and on the 3rd point an inconclusive decision. While evaluating the records, the examiners were blind to the guilt or innocence of the examinee, to the frequency of deceptive and truthful records in the sample, to the criminal case the record referred to, and to the outcomes of his previous evaluation of the same record (on the second analysis of the record). Every record was evaluated by three polygraph examiners assigned randomly with the exemption that no examiner was to score a record of an examination originally conducted by himself. The composition of the three scorers changed from one record to another according to a Latin Square design.

# Results

## Reliability in record scoring

To compute the interscorer agreement rate for the numerical scoring technique, the final scores obtained were condensed into 5 agreement regions according to the 5 main results of field tests used for polygraph examinations in the Israeli police. For the present study purposes, the two extreme points gathered all final scores less than -7 (clear deception results) or larger than +7 (clear truthful results). In the 2nd region, all final scores between -3 and -7 inclusive, were gathered (reserved deception results). Similarly, in the 4th region all final scores between +3 and +7, inclusive, were gathered (reserved truthful results). The 3rd region gathered all final scores between -2 and +2 inclusive (inconclusive region).

The difference (d) between every pair of examiners who scored the same record was determined. The lowest value of such difference was 0 (two final scores in the same region) whereas the largest value that could have been computed for such a difference was 4 (two scores in extreme opposite regions). The difference score was subtracted from the largest value of 4 in order to compute the agreement score of that pair. The largest possible difference (d=4) was then multiplied by the total number of pairs, defining the

highest possible difference score. The ratio between the sum of all agreement scores and the highest possible difference score, determined the interscorer agreement rate. It turned out that the agreement rates computed for the severe crime and the minor crime records were, 85.05% and 87.5%, respectively. Both agreement rates are fairly high and suggest that the numerical scoring was done in a reliable manner.

## Analyzing the charts using a zero cutoff

Selection of a zero cutoff point yield two decision regions. One gathered all mean total scores (across scorers) which are less than 0 and indicate deception (DI). The other corresponded to scores which were equal or greater than 0 and were defined as no deception indicated (NDI). The distribution of DI and NDI decisions according to a zero cutoff for each crime sample, is presented in Table 1.

Table 1 reveals that innocent examinees yielded more truthful than deceptive outcomes whereas guilty suspects were found more deceptive than truthful. However, when the severe crime sample is considered, the false positive error rate (records of actually innocent suspects which were classified as DI) was considerably larger that the false negative rate (43.1% and 5.6%, respectively). For the minor crime sample, an identical false positive and false negative error rate (23.3%) was found.

The selected zero cutoff point is by no means the best classification rule for guilty and innocent examinees. In real life polygraphy an inconclusive region is included to reduce the error rate.

## Including an Inconclusive Region

The inclusion of an inconclusive region was suggested by Barland and Raskin (1975), and is used routinely in field polygraphy. Table 2 presents the results for both the MC and SC samples with the inclusion of the inconclusive region. The inconclusive regions were selected by searching for the optimal discrimination between guilty and innocent suspects, thus, its boundaries are asymmetrical around zero and are skewed in the direction of the negative scores for both MC and SC samples.

**Table 1**
**Frequencies of truthful and deceptive decisions obtained from numerical scoring of the charts using a zero cutoff.**

|  | Actually Guilty | | Actually Innocent | | Across Suspects | |
|---|---|---|---|---|---|---|
| Decisions | DI | NDI | DI | NDI | DI | NDI |
| Severe Crimes | 17 | 1 | 22 | 29 | 39 | 30 |
| Minor Crimes | 23 | 7 | 7 | 23 | 30 | 30 |
| Across Crimes | 40 | 8 | 29 | 52 | 69 | 60 |

Note.  DI - Deception indicated; NDI - No deception indicated

**Table 2**
**Distribution of raw NS scores in three decision regions including skewed inconclusive regions selected for each sample.**

|  | Actually Guilty | | | Actually Innocent | | | Across Suspects | | |
|---|---|---|---|---|---|---|---|---|---|
| Decisions | DI | Inc | NDI | DI | Inc | NDI | DI | Inc | NDI |
| Severe Crimes | 14 | 3 | 1 | 5 | 17 | 29 | 19 | 20 | 30 |
| Minor Crimes | 20 | 5 | 5 | 4 | 4 | 22 | 24 | 9 | 27 |
| Across Crimes | 34 | 8 | 6 | 9 | 21 | 51 | 43 | 29 | 57 |

Note.   DI - Deception indicated; NDI - No deception indicated;
INC - Inconclusive.

The selected inconclusive regions for the Severe Crime sample  -4 / -1
and for the Minor Crime sample   -3 / 0.

Table 2 reveals the importance of the inconclusive region for reducing the false positive error rate.  This is manifested in both SC and MC samples.  When the SC sample is considered, the inclusion of the inconclusive region eliminates 17 out of 22 (77%) false positive errors without any loss of true negative decisions.  Regarding the MC sample, 3 out of 7 (43%) of the false positive errors were shifted into the inconclusive region, while only 1 out of 23 (4%) true negative decisions was lost.  The results for the false negative decisions are not as impressive.  For the SC sample the inclusion of the inconclusive region did not effect the single false negative decision, however it reduced the true positive decisions from 17 to 14 (an 18% improvement).

For the MC sample, 2 of 7 (29%) false negative decisions were shifted into the inconclusive region and the same happened to 3 out of 23 (13%) true positive decisions.

**Mean per Comparison Point**

The NS technique is a procedure to note and gather data in a reliable manner but is not necessarily the optimal method to integrate this data.  Table 3 presents the frequencies and percentages of the comparison point scores, with positive, negative and zero signs, gathered from two collections of data for guilty and innocent examinees separately.

**Table 3**
**Comparison point score signs for guilty and innocent.**

| Criterion | Actually Guilty | | | Actually Innocent | | |
|---|---|---|---|---|---|---|
| Score Sign | - | 0 | + | - | 0 | + |
| Severe Crimes | Total = 1674 | | | Total = 4749 | | |
| Number | 693 | 704 | 277 | 1146 | 2165 | 1438 |
| Percent | 41.4 | 42.1 | 16.6 | 24.1 | 45.6 | 30.3 |
| Minor Crimes | Total = 2928 | | | Total = 2784 | | |
| Number | 1081 | 1232 | 615 | 620 | 1159 | 1005 |
| Percent | 36.9 | 42.1 | 21.0 | 22.3 | 41.6 | 36.1 |
| Mean Percent Across Crimes | 39.2 | 42.1 | 18.8 | 23.2 | 43.6 | 33.2 |

Table 3 reveals that zero scores are consistently more than 40% of all scores assigned to the polygraph records. For the guilty suspects, across crime samples, scores with negative signs exceed the scores with positive signs by more than 20% while for the innocent suspects the positive scores are only 10% more than the negative scores. A closer look at Table 3 reveals that the difference originates from the severe crime sample where the guilty suspects produce about 25% more negative scores than positive scores and the innocent group produce positive scores in excess of negative ones by only 6%. The very high proportion of negative scores in the innocent group implies that a relatively minor deviation from the average may have serious consequences.

The minor crimes sample seems to be less vulnerable. It presents similar but opposite figures for guilty and innocent suspects with about a 15% difference between them.

Table 4 sums up the figures of Table 3 into percentages of agreement and disagreement with the ground truth and separates them into the three physiological measures.

**Table 4**
**Percent agreement between the ground truth and comparison point score signs for the two crime samples and the three physiological measures.**

| Measures | Severe Crimes | | | Minor Crimes | | |
|---|---|---|---|---|---|---|
| | Agreement % | Zero Scores % | Disagreement % | Agreement % | Zero Scores % | Disagreement % |
| Respiration | 38.35 | 35.73 | 25.92 | 44.54 | 32.83 | 22.64 |
| SRR | 26.34 | 54.65 | 19.01 | 32.72 | 49.16 | 18.12 |
| Cardiovascular Activity | 34.84 | 43.62 | 21.53 | 32.30 | 43.59 | 24.11 |
| Mean | 33.18 | 44.67 | 22.15 | 36.52 | 41.86 | 21.62 |

The results of the two samples, over physiological measures, are very similar: About 36% of the responses in the MC sample and 33% in the SC sample are scored in accordance with the actual state of the suspect toward the investigated crime. About 22% of the responses in the two samples are in the opposite direction. This means that the decision is based, on the average, on 13% of the responses in the test. When the different physiological measures are regarded individually, Table 4 shows that for severe crimes a difference of 13% is preserved for the respiration and cardiovascular measures but for the SRR the difference decreases to only 7%. The difference for the SRR in the MC sample is 14%, but for respiration it increases to 22% and for the cardiovascular it decreases to only 8%. The zero scores percents for the two samples are very similar, with a relatively large proportion of these scores for the SRR (about half of the scores) and a relatively low percent for respiration.

In light of the results presented in Tables 3 and 4, the widespread use of raw scores sums to determine polygraph outcomes should be reconsidered. Many polygraph records exhibit a small difference between the number of positive and negative scores and the use of raw scores sums may lead to a decision based on a record which is not clear enough. The problem is not acute for short records since the raw scores sum will probably be assigned to the inconclusive region. Records with many comparison points and a small difference between the percentages of positive and negative scores, increase the probability that the raw score sum will fall in a decision region.

To demonstrate the long records problem, the records of both MC and SC samples were divided into two groups according to the length of the record. The cutoff point was 27 comparisons. Consequently, the SC sample was divided into 35 short and 34 long records and the MC sample was divided into 26 short and 34 long records. The distribution of raw scores results for actually guilty and actually innocent suspects according to the length of the records are presented in Table 5.

**Table 5**
**Distribution of raw NS scores according to the ground truth and the length of the records.**

Length of Records

| Record's Length | | Actually Guilty | | Actually Innocent | | Across Subjects | |
|---|---|---|---|---|---|---|---|
| | | Long | Short | Long | Short | Long | Short |
| Severe Crimes | DI | 9 | 5 | 5 | 0 | 14 | 5 |
| | Inc | 0 | 3 | 5 | 12 | 5 | 15 |
| | NDI | 1 | 0 | 14 | 15 | 15 | 15 |
| Minor Crimes | DI | 9 | 11 | 4 | 0 | 13 | 11 |
| | Inc | 4 | 1 | 1 | 3 | 5 | 4 |
| | NDI | 5 | 0 | 11 | 11 | 16 | 11 |

Note.   The selected inconclusive regions for the Severe Crime sample
-4 / -1 and for the Minor Crimes  -3 / 0.

Table 5 presents a strong connection between long records and false decision probability.   All 6 false positive and false negative decisions in the SC sample, and all 9 errors of both types in the MC sample are found in long records.  It seems that when a clear result is manifested on the record after three repetitions of the question battery the

polygraph examiner tend to end the test and report the result. When the decision is not clear the examiner tends to continue with the test hoping that the forthcoming repetitions would clarify the result. Unfortunately, this approach may lead to false decisions.

To overcome the problem, the raw score sum was replaced with the mean score per comparison point (MCP). The MCP is considered as a single score assigned by one scorer to one comparison of responses to adjacent relevant and control questions in a single physiological measure. The theoretical maximum MCP is 3.00 and will occur only when all three examiners decide that the responses to all control questions are greater than those to all adjacent relevant questions and the difference is judged to has the maximum value (+3). This should occur in all three measures for all the questions in the record. In general, a positive mean score

indicates a relatively stronger response to the control questions. Conversely, -3.00 is the theoretical minimum score possible, and negative mean score indicates that the responses to the relevant questions in a given record are stronger than those to the adjacent controls. In order to compare the validity of the MCP with that of the NS raw scores, the MCP score for each record was computed.

The optimal inconclusive region for the MC sample has boundaries of -.2 and +.1. Thus, all MCP scores less than -.2 are considered as deception indicators (DI), while all MCP scores greater than +.1 are considered as no deception indicators (NDI). All other MCP scores are considered inconclusive. Similarly, the optimal inconclusive region for the SC group was determined. The region has boundaries of -.2 and 0. Table 6 presents the distribution of MCP scores in each of the three decision regions for both SC and MC groups.

**Table 6**
**Distribution of MCP scores according to the skewed inconclusive regions for the two samples.**

|  |  | Actually Guilty | Actually Innocent | Across Subjects |
|---|---|---|---|---|
| Severe Crimes | DI | 12 | 1 | 13 |
|  | Inc | 5 | 21 | 26 |
|  | NDI | 1 | 29 | 30 |
| Minor Crimes | DI | 17 | 2 | 19 |
|  | Inc | 9 | 9 | 18 |
|  | NDI | 4 | 19 | 23 |

Note.   The selected inconclusive regions for the Severe Crime sample
-.2 / 0 and for the Minor Crimes  -.2 / +.1.

Table 6 reveals that the distribution of MCP scores in the three decision regions eliminate many of the false positive decisions found for the raw score sums. This is true for both MC and SC groups. When the SC group is considered, 4 out of 5 (80%) false positive decisions were shifted into the inconclusive region while none of the 29 true negative decisions were affected. For the MC group, 2 out of 4 (50%) false positive decisions were included in the inconclusive region at the price

of 3 out of 22 (14%) true negative decisions. In the actually guilty suspect group, the only false negative decision in the SC sample was unaffected by the change from NS raw scores to MCP scores, but 2 out of 14 (14%) true positive decisions were shifted into the inconclusive region. For the MC group a similar percent of false negative (1 out of 5, (20%)) and true positive (3 out of 20 (15%)) decisions were changed to inconclusives.

**The Intersection Rule**

The effect of a hybrid method, based on the NS technique and the GRE method of record interpretation, upon correct and incorrect decision rates, was investigated. For this purpose, the evaluations of the three evaluators on the 5-point scale were summed up yielding a 13 point continuum in which the first score (3) reflects an unanimous decision of the three evaluators that the suspect is undoubtedly deceptive, while the last (15) reflects a consensus about a strong truthful decision. The 13-point continuum was divided into three decision regions which best discriminate between guilty and innocent suspects. The first consisted of the five extreme DI evaluations (sums of 3 to 7, inclusive). The second gathered all evaluation sums between 8 and 10 inclusive, (inconclusive region), and the third consisted of all the sums greater than or equal to 11 (NDI region).

A new decision rule was now defined, according to which a result would be considered truthful only if it is included in both NS and GRE NDI regions. The same rule was applied to the deceptive results. All other decisions which did not meet the demands of the intersection rule were regarded as incon-clusives. Table 7 presents the distribution of MCP scores in the three decision regions after applying the intersection decision rule.

**Table 7**
**Distribution of MCP scores in the three decision regions for the two samples after applying the intersection decision rule.**

|  |  | Actually Guilty | Actually Innocent | Across Suspects |
|---|---|---|---|---|
| Severe Crimes | DI | 11 | 0 | 11 |
|  | Inc | 6 | 30 | 36 |
|  | NDI | 1 | 21 | 22 |
| Minor Crimes | DI | 17 | 1 | 18 |
|  | Inc | 11 | 10 | 21 |
|  | NDI | 2 | 19 | 21 |

Table 7 reveals that for the MC group the intersection rule eliminated 3 out of 6 (50%) errors (1 false positive and 2 false negative errors). Furthermore, the rule has no effect on the correct detection rate of both true positive and true negative decisions. When the SC group is considered, the decision rule shifted the only false positive error to the inconclusive region but did the same to 8 out of 29 (28%) true negative decisions. For the guilty group of subjects, the intersection rule reduced the true positive decisions from 12 to 11 (-8%). While not being impressive, even the results for the SC group seem to justify the use of the intersection rule.

For both samples the decision rule eliminated half of the errors. The price in reduction of correct decisions was relatively low (0% in the MC group and 22% in the SC group).

## Conclusions and Suggestions

The robustness of the present results lean on the fact that the reported decision rules apply to two independent and completely different samples of polygraph records. On this ground it is recommended to use both semi-objective numerical scoring and global record evaluation techniques. For each, an inconclusive region should be included because, as the present analyses show, such a region considerably reduces both false positive and false negative errors, at a price of relatively lower reduction of the correct decisions. The inconclusive region boundaries seem to be slightly skewed in the direction of

the negative scores, and this tendency is somewhat stronger for the SC group when compared to the MC group. It is further recommended to replace the NS raw scores sums with MCP scores which standardize results of records with varied length and eliminate many of the errors which result from this variation. Finally, it is suggested to apply an intersection decision rule to both MCP and GRE results. This may lead to a further substantial reduction of both false positive and false negative errors from a rate of about 10% to a rate of about 5%. At the same time the inconclusive region is expected to grow from about 30% to about 40%.

It must be noted, however, that a large proportion of the polygraph records in the present study were verified by confessions. This may introduce a sampling bias since the confession is not independent of the polygraph outcomes and the probability that a suspect will confess may depend on the polygraph results. Hence, deceptive outcomes may encourage interrogation efforts to induce a confession, while a truthful outcome may convince the interrogator to stop interrogating. This may result in an underestimation of the false-negative rate. On the other hand, one cannot exclude the possibility of false confessions. In this case the false negative rate may be overestimated. Hence, the verified sample is not representative of the population of polygraph examinations (Elaad and Schachar, 1985).

# References

Backster, C. (1969). *Technique fundamentals of the tri-zone polygraph test.* New York: Backster Research Foundation. Inc.

Barland, G.H. & Raskin, D.C. (1975). An evaluation of field techniques in detection of deception. *Psychophysiology*, 12, 321-330.

Elaad, E. & Schahar, E. (1985). Polygraph field validity. *Polygraph*, 14, 217-223.

Podlesny, J.A., & Raskin, D.C. (1977). Physiological measures and the detection of deception. *Psychological Bulletin*, 84, 782-799.

Podlesny, J.A. (1993). Is the guilty knowledge polygraph technique applicable in criminal investigations? A review of FBI case records. *Crime Laboratory Digest*, 20, 57-61.

Raskin, D.C. (1989). Polygraph techniques for the detection of deception. In D.C. Raskin (Ed.), *Psychological Methods in Criminal Investigation and Evidence*. New York, Springer Publishing Company.

Reid, J.E., & Inbau, F.E. (1977). *Truth and deception: The polygraph ("lie detection") technique.* 2nd ed. Baltimore: Williams & Wilkins.

# Rank Order Analysis

## Kathleen Miritello

The name Rank Order Analysis literally explains this procedure, as it is a "ranking" of the questions on a chart from greatest to least responsiveness. In rank order chart analysis each physiological parameter (respiration, electrodermal, and cardiovascular) is evaluated separately. The benefits of rank order analysis include:

1. A systematic approach to chart evaluation which has been proven in research studies to be more accurate than a visual or global review.

2. Each physiological parameter is evaluated separately and given equal emphasis in the decision process. An examiner cannot ignore or overlook any of the parameters based on individual preferences or biases.

3. Rank order analysis provides the examiner with a cumulative picture of the consistency and intensity of reactions on the polygraph charts. It serves to organize the data contained in the charts, and is especially helpful when reactions are inconsistent, chart quality is marginal, or numerous charts have been run.

4. Use of this procedure enhances the examiner's confidence in the final conclusion by providing a good foundation for the decision. It can serve as a quality control check of the charts in the field when an examiner is without benefit of an on-the-spot QC for his or her charts.

5. A systematic and objective chart analysis prevents the examiner from ignoring or rationalizing the pattern of reactions on the charts. Rank order ratios force the examiner to recognize which questions were consistently most reactive, despite any possible behavioral clues which may appear to be indicating truthfulness.

6. The recorded rank order analysis results allow for a more effective quality control review by providing a definitive record of the examiner's chart analysis which can be compared with the findings of the QC supervisor. Any source of possible disagreement is readily identified.

7. The reliability of chart analysis, an important measure which is related to validity, can be clearly demonstrated through use of rank order analysis.

8. The chances of an improper inconclusive call are reduced with rank order analysis. In some cases which might otherwise be considered inconclusive, the use of this technique can define potential areas of deception more clearly in the charts, steering the examiner's interrogation in the proper direction in order to gain information.

Rank order chart analysis is performed by the following steps.

**Step 1**. The number of questions to be evaluated per chart determines the values to be assigned in ranking the questions. Count the number of questions to be evaluated on each chart and subtract 1. The resulting number should be awarded to the question you feel has the largest reaction in respiration. For example, if there are 6 questions to be evaluated, 6 - 1 = 5. The largest respiration reaction will be given a 5. The second largest reaction should be given a 4, and so on down to 0 for the question with the least or no

reaction. This procedure is repeated for the electrodermal and cardiovascular patterns, again ranking the questions in order of response down to 0. (See example below.)

**Chart #2: Questions A, B, C, D, E, and F.**

| Question | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Rank | 2 | 5 | 3 | 4 | 1 | 0 |

**Step 2**. If you believe reactions to questions are of equal strength, add the value of the ranks they would be assigned and divide by the number of questions involved. In the example below, question B stands out clearly as being the most reactive. It is assigned a rank value of 5. The next most reactive questions, D and C, appear to be of equal strength. The values that would be assigned next are 4 and 3. These are added together, and then divided by the number of tied questions: 4 + 3 = 7; -7/2 = 3.5. Questions D and C are then both assigned values of 3.5. You continue to assign the remaining values where the ranks left off since values 4 and 3 have already been assigned, 2 is given to the next most reaction question, A, and so on down to 0.

| Question | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Rank | 2 | 5 | 3.5 | 3.5 | 1 | 0 |

**Step 3**. If you feel an entire parameter cannot be evaluated (e.g., PVC contamination, totally erratic respiration, completely flat electrodermal tracing), give all questions the average rank for that parameter. In the following example, acceptable tracings exist in the respiration and electrodermal channels, but the ranks of the cardiovascular channel are averaged due to excessive PVC contamination.

| Question | A | B | C | D | E | R |
|---|---|---|---|---|---|---|
| Respiration | 2 | 5 | 3.5 | 3.5 | 1 | 0 |
| Electrodermal | 3 | 4 | 2 | 5 | 0 | 1 |
| Cardiovascular | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 | 2.5 |

**Step 4**. Next, the totals are added for each question by summing the values assigned to each parameter.

| Question | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Rank | 2 | 5 | 3.5 | 3.5 | 1 | 0 |
| Electrodermal | 3 | 4 | 2 | 5 | 0 | 1 |
| Cardiovascular | 2.5 | 5 | 4 | 2.5 | 1 | 0 |
| Total | 7.5 | 14 | 9.5 | 11 | 2 | 1 |

**Step 5**. This process is completed for each separate chart. The examiner has the option of evaluating only the relevant and comparison questions, or also evaluating irrelevant questions. At a minimum, all relevant questions and at least one comparison question should be evaluated.

**Step 6**. The final step is to calculate ratios for each question. Ratios provide some indication of the degree of separation between the questions by placing them against a 0 to 1.0 scale. Also, in tests where the relevant questions are not asked an equal number of times, ratios make adjustments for the differences in the resulting totals to allow for a fair comparison between questions. Ratios are calculated by dividing the total value for each question by the maximum possible total for that chart. This maximum value is determined by summing the values accorded to the highest rank in each parameter. In the examples we have shown, the highest value for each parameter was a 5, since we were evaluating 6 questions. A maximum of 5 for the respiration channel, plus 5 for the electrodermal channel, plus 5 for the cardiovascular channel, equals a maximum possible total of 15. Using the totals from the example shown above, the ratio for A would be 7.5 divided by 15, or .50. The remainder of the ratios have been calculated and are shown below.

| Question | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Respiration | 2 | 5 | 3.5 | 3.5 | 3.5 | 0 |
| Electrodermal | 3 | 4 | 2 | 5 | 0 | 1 |
| Cardiovascular | 2.5 | 5 | 4 | 2.5 | 1 | 0 |
| Totals | 7.5 | 14 | 9.5 | 11 | 2 | 1 |
| Ratios | .50 | .93 | .63 | .73 | .13 | .07 |

To calculate rank order ratios for a question after several charts, the totals from each chart are added together and divided by the maximum possible value for each of the charts. For example, question G received the following total values:

Chart 4: 11    (Maximum possible value = 15)
Chart 5: 9.5   (Maximum possible value = 15)
Chart 6: 7.5   (Maximum possible value = 12)
        (only five questions were ranked)

The total for question G would be 28 and the ratio is determined by dividing that figure by 42 (the sum of the maximum possible values) = .67.

However, in another example, question H is asked twice and results in the following figures:

Chart 4: 11    (Maximum possible value = 15)
Chart 5: 14    (Maximum possible value = 15)

Even though the final total, 25, is less than the total for question G, the ratio indicates that question G was actually more reactive; 25/30 = .83.

It is extremely important to keep in mind that rank order analysis does not provide any automatic cut-off scores which can be used to reach firm conclusions of truth or deception. The purpose of the rank order ratios is to provide the examiner with more accurate information for applying the principles of chart analysis already in use.

# Scoring in a Computer Age

## James Wygant

The extensive use of computer-aided chart evaluation brings new focus to manual scoring. The obvious questions are: 1) when is manual scoring necessary; and 2) what is the best means of resolving a discrepancy between a manual score and a computer result?

Examiners are not apt to find answers to those questions in any research, except to the extent that the value of manual scoring is well established by numerous studies over the past 25 years. Despite widespread acceptance of the concept of scoring, actual practice falls into four general categories. An examiner may manually score all charts from every examination, even the most obvious. He may score only the charts that are not obvious. He may score only the exams for which he can not otherwise reach a "global impression" of truth or lie. And finally, the examiner may score nothing, relying entirely on "global impressions" or on a computer if he or she has one. We could argue that these four options happen to have been listed in order of best to worst practice. Fortunately for the profession, informal surveys of examiners confirm that most manually score everything, even when using a computer.

If nearly everybody agrees that scoring is the best way to evaluate charts, there are varying opinions about what range of numbers to use when matching issue questions (relevants) against comparison questions (controls). Cleve Backster, who is generally credited with devising numerical scoring, set the range at +3 to -3. Zero meant no distinguishable difference between physiological responses on an issue question and a comparison question. A score of 1 meant a slight difference. Two meant a noticeable difference. Three was reserved for whoppers and for what Backster called "upgrading." A plus score meant greater response on the comparison question; a minus meant the issue question produced the greater response.

Backster applied scoring to what he called the "you phase" test, a reference to the "did you do it" thrust of the issue questions, basically a single-issue test by another name. Perhaps an unintended and unanticipated consequence of the increase in use of numerical scoring has been a gradual and universal change in test formats, even among those who don't know a "you phase" from a "U-boat." When testing specific allegations, most examiners now use a comparison question format that includes 2-3 issue questions and 2-3 comparison questions. The issue questions often constitute a single issue or at least issues that are so closely related that a lie to any one, or truth on any one, forces the same for the others. There are fewer and fewer instances in specific allegation testing of mixed issues and relevant-irrelevant formats, neither of which is as well adapted to scoring as a single issue.

The scoring range that Backster advocated, +3 to -3, is commonly identified as the seven-position scale, because it encompasses seven possible values, including zero. The range might just as well have been +99 to -99, although the addition and subtraction necessary to calculate totals would have been more cumbersome, and cutoffs applied to total scores would need to be higher. Still, any range of numbers will works with the appropriate cutoffs. There is nothing sacred about +3 and -3, except that those numbers offer a range of values that are relatively easy to use.

In actual practice, examiners who use the seven-position scale rarely assign values of +3 or -3. That is reflected in research by Capps and Ansley (1992a). In 440 cases, with 11,682 separate comparisons of responses, examiners assigned threes to the extent shown in Table 1. The numbers represent the percentage of all values that were either +3 or -3.

---

The author is a polygraph examiner in private practice. Reprint requests should be sent to him at The Mayer Building, 1130 SW Morrison St, Ste 220, Portland, OR 97205.

**Table 1.  Percentages of +3 or -3 scores among 440 cases.  (from Capps & Ansley, 1992[a]).**

| Chart | Pneumograph | Electrodermal | Cardiograph |
|-------|-------------|---------------|-------------|
| 1 | 1.3 | 18.6 | 2.8 |
| 2 | 1.2 | 19.6 | 2.2 |
| 3 | 3.1 | 23.9 | 4.3 |

Two things are obvious from the tabled data.  Threes are rarely given to pneumograph or cardiograph reactions; and they are used with the electrodermal function to such a disproportionate extent that the examiner gives substantially more weight to that one tracing.  The cause of this phenomenon rests within the nature of the tracings.  The electrodermal tracing is much simpler to score than the pneumograph or cardiograph tracings.  It consists of a smooth flowing line that rises and falls in measurable increments.  Differences that warrant a three in scoring are easier to observe and can be justified with a ruler.  The cardiograph and pneumograph tracings are more complex, relying to a greater extent on subjective judgement, and consequently invite argument about the magnitude of a score.  Polygraph examiners are like most other people, preferring a safe, conservative course rather than one that can lead to a dispute.  That translates to fewer threes given to respiration and cardiograph tracings.

**Other Scoring Methods**

Many examiners use a five-position scale, +2 to -2, and thereby avoid weighting the electrodermal channel.  In the past few years many have also begun using a three-position scale, +1 to -1.  The rationale for the three-position scale is that it is allegedly the most objective.   It relies simply on the presence or absence of a scorable response.  Those who use it will tell you that they distrust examiner judgment in deciding when the difference between issue and comparison questions is great enough to permit a +2 or -2.  On the other hand, those who use twos argue that not all differences in responses are of the same magnitude.  They assert that since some differences are clearly greater than others, provision must be made for that in scoring.

A separate Capps and Ansley study concluded that the inconclusive rate with a three-position scale was more than double that obtained with a seven-position scale, although accuracy remained equally high with both methods for those cases in which a decision was made.  The researchers used 100 verified single-issue examinations.  The difference in inconclusive rates was greatest for the 52 verified deceptive tests.  The three-position scale produced an inconclusive rate of 23%, while the seven-position rate was only 4%.

Apart from concern about inconclusive results, the various methods rarely contribute to any difference of opinion in test results.  Two examiners using different scoring ranges should theoretically arrive at the same result of truth or deception, if able to reach any conclusion at all.  If the conclusions differ, the fault does not lie with the numbers, but with the different judgments the two examiners have made regarding what they think they are seeing on the same set of charts.

Research might someday help resolve the issue of scoring ranges, but in the meantime we tolerate examiners using a variety of number systems.  Other non-numerical systems, for instance involving check marks, have also been proposed over the years.  All note-keeping systems have the same goal, to create a simple method for the examiner to track what he or she has observed through an entire examination.

The amount of data an examiner must consider is considerable.  A chart usually includes at least four channels: two respiration records, an electrodermal tracing, and a cardiograph tracing.   A typical chart might also consist of three issue questions and three comparison questions.  The examiner will usually assign three scores per issue question; one for respiration, one for electrodermal, and one for cardiograph.  To arrive at those scores, the examiner must make comparisons

between an issue question and at least one comparison question. For an examiner who looks at more than one comparison question, the numbers given here as examples are only going to get bigger. Even though the examiner assigns only one score for respiration, both the upper and lower tracings must be reviewed. The total number of judgments an examiner must make for one question on one chart is therefore at least eight (looking at each tracing for the issue question and one comparison question). For three issue questions on one chart, that is 24 judgments. For three charts, that is 72 judgments, a lot to remember accurately and to tabulate without some kind of system.

We will assume that an examiner is convinced that he needs to keep a written notation of what he is seeing during his chart evaluation. He must then decide what he will do with that data, after he's gone to the trouble of collecting it.

## Results Reporting

Most examiners refrain from referring to their scores when they report test results. If a defense attorney or deputy district attorney gets "trained" by one examiner regarding his cutoffs and the strength of his results, he is apt to assume that all examiners score the same. Although polygraph testing has become highly standardized, the fact is that some examiners score conservatively, others are more generous in assigning values, some use a range of +1 to -1, while others range from +3 to -3. One examiner's score on a set of charts may be +12, while another's is +6. That is not

a problem for either examiner, since they both conclude truthfulness. It may be a problem for an attorney who thinks that scores are as concrete as the inches on a ruler and that something must be wrong if there is a six-point difference.

The simplest method of avoiding that kind of confusion is to avoid reporting numbers, either those derived from manual scoring or from computer evaluation. The numbers are not the result. The result is what the examiner concludes from the numbers and anything else his training and experience cause him to believe is relevant and can be justified.

That also means that cutoffs are not absolute. Most examiners use +6 and -6 as minimum scores necessary for a finding of truthfulness or lying in a single-issue examination that consists of three issue questions and three charts. In fact, cutoffs as low as +3 and -3 have been shown to work nearly as well, while reducing the number of inconclusive results.

In an article about the selection of comparison questions in scoring, Michael Capps and Norman Ansley (1992b) coincidentally reported interesting data regarding cutoffs. They had eleven examiners score forty examinations, for a total of 440 conclusions. The tests were single-issue procedures consisting of two issue questions. Best results were obtained from comparing each issue question to the stronger adjacent control. The results were tabulated with various cutoffs. (See Table 2)

**Table 2. Correct, incorrect, and inconclusive outcomes for eleven examiners scoring 40 examinations. (From Capps & Ansley, 1992b).**

| Cutoff | Number Correct | Number Incorrect | Number Inconclusive | Percent Correct |
|--------|----------------|------------------|---------------------|-----------------|
| +/-6 | 361 | 10 | 69 | 97.30% |
| +/-5 | 368 | 13 | 59 | 96.60% |
| +/-4 | 382 | 13 | 45 | 96.70% |
| +/-3 | 386 | 17 | 37 | 95.80% |
| +/-2 | 395 | 24 | 22 | 94.30% |
| +/-1 | 402 | 28 | 10 | 93.50% |

It doesn't take any advanced math to see that as the cutoffs were lowered two things happened: the number of inconclusive results declined and the number of mistakes increased, though not as dramatically. Errors went up from 10 to 17 when the cutoff was changed from +/-6 to +/-3. At the same time, inconclusives were cut nearly in half, from 69 to 37, representing a drop in the inconclusive rate from 15.7% to 8.4% and a corresponding rise in the error rate from 2.7% to 4.4%.

The point of reviewing these data is not to advocate lower cutoffs, but only to illustrate that there are two primary determining factors in setting cutoffs. The first is accuracy. The second is inconclusive rate. Any examiner who is suffering an inconclusive rate greater than 20% may want to consider either scoring more aggressively or lowering cutoffs, or both. On the other hand, any examiner who has an exceptionally low inconclusive rate is probably making too many mistakes.

## Human vs. Computer

That brings us to a question raised in the first paragraph: what should an examiner do about a difference of opinion between his or her score and a computer result? That is a question every examiner who uses computer-aided scoring must expect to address at some point. A difference is certain to occur. It should not happen often, and it should rarely be a circumstance of opposite opinions, one concluding truth while the other concludes lying. This is an area too new for research results to be available. Anecdotal information suggests that differences are occurring for most examiners less than 10% of the time; and most of the differences are only one degree (the examiner's manual score indicating a definite result while the computer says it's inconclusive, or the reverse of that.

Any examiner who frequently disagrees with his or her computer may want to ask some human colleagues to review a few examinations. If everybody is getting results that differ from the examiner who gave the test, that examiner needs to figure out why.

Occasional differences can derive from several sources. In using computer-aided evaluation, it is imperative that the examiner

not mix issues. The computer software was written with the assumption that the examiner is submitting a single-issue test, or at least one in which the issue questions address something in which lying to one presumes lying to all. Asking a computer to evaluate a test in which one question asks about stealing money from the victim, one asks about raping the victim, and one asks about the suspect's alibi is an invitation to disaster. The people who wrote the software assumed the examiner would be using the best test available, which research indicates is the single issue. The software programmers have repeatedly cautioned that a computer conclusion of truth or lie in a mixed issue test may not be reliable.

Another consideration that is within the examiner's control is the care he or she exercises in manually removing distortion from the computer's consideration. Although the computer algorithms are good at identifying many forms of distortion, they can miss the most obvious. It is the examiner's responsibility to feed in good data if he or she expects to get back reliable results.

There is one other common source of differences that, unlike the others, is totally beyond the examiner's control. It is the method by which the charts are evaluated. Examiners are generally limited to comparing an issue question to a control on either side of it. To jump much further away from the issue question makes the comparison more difficult and is usually not attempted unless there are no better options. A computer does not suffer that same handicap. It can make comparisons and do statistical analyses that encompass all of the charts, not just limited portions of any particular chart. That does not mean that its results are necessarily more accurate or reliable, but they are certainly more comprehensive.

Faced with a difference, what should a conscientious examiner do? Since the examiner must assume responsibility for the result, he or she should not automatically either accept or reject a computer result that differs. Further testing or review by another examiner is often helpful and probably should be the first consideration, before issuing a report in a case where examiner and computer disagree. There may occasionally be cases in

which the examiner is comfortable reporting that his or her result was definite truth or lie, while the computer was inconclusive, or vice versa.

Whether an examiner even mentions a difference with a computer depends to a great extent on how he or she routinely writes reports. If reports typically make no reference to a computer result, the examiner can continue to give an opinion without mentioning any difference with the computer. It may lead to some surprises, disputes, and disappointments later on, but at least there can be no demonstration that the examiner tried to conceal information he or she would have otherwise ordinarily reported. The examiner in those circumstances can always argue that he considered the computer result advisory, along with his own score, his observations, and whatever else he believes justifiably contributes to his opinion.

When an examiner does routinely mention the computer result in his reports, omitting it when there is a difference of opinion can lead to serious problems. To reveal a difference belatedly suggests an attempt to hide something and can damage the examiner's credibility.

The styles for describing a non-agreeing computer opinion vary considerably. A report might say: "Based on a review of my own numerical evaluation and other factors that included computer evaluation of these charts, I conclude that these charts indicate deception/truthfulness with regard to the questions about the issue." There is no mention of what the computer concluded, only that it was taken into consideration. This is probably the least desirable mode of reporting. A more candid version would say: "I conclude from my numerical evaluation of these charts that they indicate lying/truthfulness with regard to the questions about the issue. Computer evaluation of these same charts produced an inconclusive result." The reverse situation, manual inconclusive and definite computer result, might be worded: "It is my opinion, based on my evaluation of these charts, that they are inconclusive. Computer evaluation indicated lying/truthfulness to the questions about the issue, which I do not find supported by the charts."

The common factor in all of these samples is that the examiner (the big "I" in the examples above) assumes full responsibility for the result he or she is delivering. There is no attempt to characterize the conclusion as something beyond the examiner's control, the automatic determination of either a computer or some mystical scoring system. After all, if an examiner wants to give the impression that this game is played solely by the numbers, then he is advancing the notion that he is about as skilled as the Roto-Rooter man.

Whether any examiner reports any computer result is ultimately a matter of personal preference. How he or she responds to any differences of opinion, whether with a computer or another examiner, is always a reflection of the examiner's professional stature and self-confidence.

# References

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

Capps, M.H., & Ansley, N. (1992b). Strong control versus weak control. *Polygraph*. 21(4), 341-348.

# Short report

# Proposed Method for Scoring Electrodermal Responses

## Donald Krapohl

The material in this report is drawn from a larger study which is in progress. The data from 300 polygraph cases were randomly selected from the Department of Defense Polygraph Institute confirmed case database. Half of the charts were from truthful examinees and the other half were from deceptive examinees. All cases were single-issue examinations conducted in the field on computer polygraphs by federal or law enforcement polygraph examiners. Ground truth was established by confession of the subject, confession of someone who exculpated the subject, or other irrefutable forensic evidence. Accuracy of the original examiner decision was not a criterion for inclusion of the cases in the database. The test format for all cases was the DoDPI Zone Comparison Test (DoDPI, 1992).

The electrodermal response (EDR) data were standardized as the ratio of the relevant question EDR amplitude divided by the amplitude of the greater adjacent comparison question, within each chart. A total of 2700 EDR ratios were calculated from the 300 cases (3 charts consisting of 3 relevant questions each per case). The distribution of ratios was sequentially divided into seven groups such that all groups had the same number of ratios. Each division corresponded to one of the values used in the seven-position scale: +3, +2, +1, 0, -1, -2, -3. Ratios greater than 1.00 indicated greater EDR amplitudes to the relevant question than to the comparison question, while ratios smaller than 1.00 indicated the EDR to the relevant question was less than the EDR to the comparison question. Table 1 lists the cutting scores based on the present data.

The scores obtained using this method were compared to those of the 2:1, 3:1, 4:1 requirements of the conventional (DoDPI, 1999) scoring rules with these 300 cases. The proposed method produced a mean score per case more in the correct direction for deceptive $(t(149)=11.34, p<.01)$ and nondeceptive $(t(149)=15.25, p<.01)$ than did the conventional scoring rules (see Table 2). Moreover, the proposed method produced a higher proportion of EDR total scores in the correct direction than did the conventional method (86.3% versus 74.0%, z=3.77, p<.01). In short, the proposed method outperformed the conventional method for scoring the EDR with these 300 field criminal cases.

## Table 1
**Cutting scores for the seven-position scale and the corresponding ratio of relevant question EDR amplitudes and comparison question EDR amplitudes. (n=2700 ratios)**

| 7-Position Value | EDR Ratios (R/C) |
|:---:|:---:|
| +3 | <.44 |
| +2 | .44 - .67 |
| +1 | .68 - .92 |
| 0 | .93 - 1.20 |
| -1 | 1.21 - 1.60 |
| -2 | 1.61 - 2.44 |
| -3 | >2.44 |

**Table 2**
**Mean values for the conventional and proposed method of scoring the electrodermal channel by case (n=150 deceptive and 150 nondeceptive cases)**

|  | Conventional | Proposed |
|---|---|---|
| Deceptive cases | -4.37 | -8.33 |
| Nondeceptive cases | 3.53 | 8.43 |

**Table 3**
**Conventional and proposed ratios for scoring EDRs**

|  | Greater response to relevant question | | | Greater response to comparison question | | |
|---|---|---|---|---|---|---|
| Comparison | R:C | R:C | R:C | R:C | R:C | R:C |
| Score | -3 | -2 | -1 | +1 | +2 | +3 |
| Conventional | 4 : 1 | 3 : 1 | 2 : 1 | 1 : 2 | 1 : 3 | 1 : 4 |
| Proposed | 2.45 : 1 | 1.61 : 1 | 1.21 : 1 | 1 : 1.09* | 1 : 1.49* | 1 : 2.27* |

\* Note: These values are reciprocals of the ratios found in Table 1. They have been inverted in Table 3 to facilitate the comparison with the more familiar conventional ratios.

In the first of two cross validations of the proposed scoring rules, a holdout sample consisting of 30 confirmed deceptive and 30 confirmed nondeceptive cases were tested. Applying the proposed scoring method to these new data, the mean of the total EDR scores for deceptive cases was more negative than those scores rendered by the conventional scoring rules (-10.5 versus -5.6) and this difference was significant (t(29)=7.37, p<.01). Similarly, the mean of the total EDR scores for nondeceptive cases was more positive than those of the conventional scoring system (7.1 versus 2.7) and this difference was also statistically significant (t(29)=5.62, p<.01). The proposed method produced 54 of 60 scores in the correct direction, while the conventional method produced only 49, though the difference in the proportions was not significant (z=1.32, p>.01). A second cross validation with 100 new cases will be completed and reported later this year.

It is interesting to note that the proposed scoring rules are not symmetrical. The assignment of a negative value to the relative sizes of the EDR amplitudes (R/C) requires a stronger response to the relevant question than does the assignment of a positive value if the comparison is in the other direction (C/R). In other words, the thresholds for assigning +1,+2, and +3 are lower than the thresholds for assigning a -1, -2 and -3. Table 3 compares the proposed and conventional ratios. There are many possible explanations for this response pattern, but a working hypothesis is that, on average, nondeceptive examinees do not respond as strongly to the comparison questions as the deceptive examinees respond to relevant questions. In support of this hypothesis, Figure 1 shows the relative sizes of EDRs to comparison and relevant questions for the 300 deceptive and nondeceptive subjects used to develop this scoring method. If nondeceptive subjects do not react electrodermally to comparison questions as strongly as deceptive subjects do to relevant questions, optimal scoring rules would be asymmetrical, such as was found here. The present findings, of course, require replication. If confirmed, all scoring methodologies will have to accommodate the asymmetry in response patterns for truthful and deceptive examinees.

**Figure 1**



Relative EDR amplitudes for relevant and comparison questions for deceptive and nondeceptive examinees (n=1350 samples per bar)
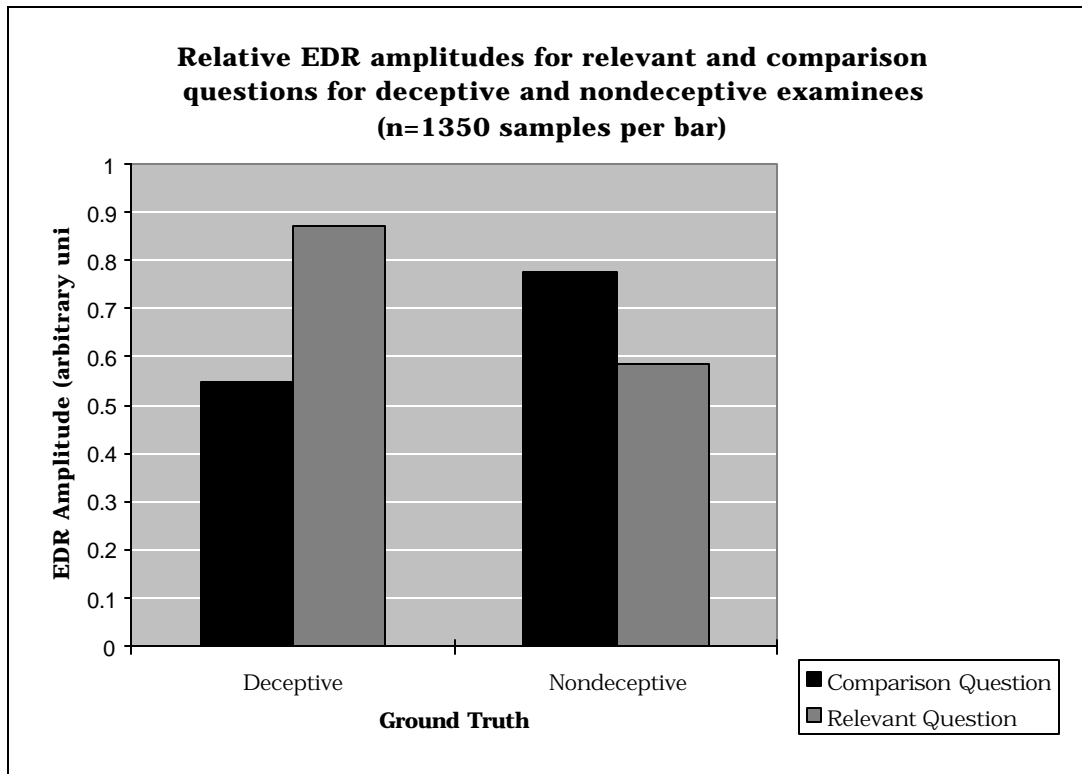
## Summary

Despite the near universality of the conventional EDR ratio scoring rules in polygraph training and practice, the present writer could find no support in the literature for setting the EDR ratios at 2:1, 3:1 and 4:1. The notion that psychophysiological responding is so mathematically convenient is probably unduly optimistic. Though the conventional EDR ratios have been part of validated scoring procedures over the last 30 years, the present findings suggest that those EDR scoring rules have not been optimized, and that decision accuracy would benefit by the development of empirically based scoring rules. The method outlined in this report is one such attempt. A forthcoming full report will include the second cross validation data, and similar analyses of the cardiograph and pneumograph tracings.

## References

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

Department of Defense Polygraph Institute (1999). *Test data analysis (manual scoring process), seven-position numerical analysis scale.* Ft. McClellan, Alabama.

# Chart Interpretation – A Bibliography

## Norman Ansley

A selective bibliography on the topic of chart interpretation. It includes recent developments in the use of computers. Most of these papers are available in the files of the DoDPI, the files of the APA References Service, and the files of the author.

Key words: Algorithms, Bibliography, Chart Interpretation, Computer Charts, Hand Scoring

Abrams, S. (1989). *The Complete Polygraph Handbook.* Lexington, MA: Lexington Books.

Abrams, S. (1977). *A Polygraph Handbook for Attorneys.* Lexington, MA: Lexington Books.

Abrams, S. (1973). Problems in Interpreting Polygraph Examinations in Malingerers. *Polygraph,* 3 (4), 377-380.

Abrams, S. (1975). The Validity of the Polygraph With Children. *Journal of Police Science and Administration,* 3 (3), 310-311.

Abrams, S. (1974). The Validity of the Polygraph With Schizophrenics. *Polygraph,* 3 (3), 328-337.

Abrams, S. & Weinstein, E. (1974). The Validity of the Polygraph With Retardates. *Journal of Police Science and Administration,* 2 (1), 11-14.

Adachi, K. (1989). Discriminant Functions Based on Models for Randomized Block Design for Detection of Deception. *Reports of the National Institute of Police Science,* 42 (3), 24-31.

Adachi, K. & Suzuki, A. (1990). Correlation Structure Among Response Measures in Detection of Deception. *Reports of the National Institute of Police Science,* 43 (3), 13-16.

Adams, T. (1982). A Neurophysiological review and a Proposed Rationale for Interpreting Electrodermal Polygraph Records. *Polygraph,* 11 (4), 285-303.

Amsel, T. T. (1997). Fear of Consequences and Motivation as Influencing Factors on Psychophysiological Detection of Deception. Polygraph, 26 (4), 255-267.

Angus, J. & Castelaz, P. (1993). Artificial Neural Network Analysis of Polygraph Signals (Final Report) DoDPI 93-R-0010.

Ansley, N. Capillary Responses as a Polygraph Channel. In N. Ansley (Ed), *Legal Admissibility of the Polygraph.* Springfield, IL: Charles C. Thomas, 1975, pp. 257-265.

Ansley, N. (1959). A Comparison of Arm-Cuff and Wrist-Cuff Blood Pressure Patterns in Polygraph Charts. *Journal of Criminal Law, Criminology, and Police Science,* 50 (2) 192-194.

Ansley, N. (1990). *Scientific Research on Detecting Deception and Verifying Truth.* Washington: United States Department of Defense.

Ansley, N. (in press). The Validity of the Modified General Question Test (MGQT).

Aragaki, F. (1979). Intensity of the Suspicion Expressed in Questions and the Physiological Responses. Paper presented at the National Convention of Police Psychology at the National Research Institute of Police Science.

Arther, R. O. (1955). Blood Pressure Rises on Relevant Questions in Lie Detection – Sometimes an Indication of Innocence Not Guilt. *Journal of Criminal Law, Criminology, and Police Science,* 46, 112-115.

Arther, R. O. (1971). Breathing Analysis, Part I. *Journal of Polygraph Studies,* 5 (4) 1-4, Part II, *Journal of Polygraph Science,* 5 (5), 1-4.

Arther, R. O. (1977). Miscellaneous Chart Analysis Principles, Part I, *Journal of Polygraph Science,* 11 (5), 1-4, Part II, *Journal of Polygraph Science,* 12 (1), 1-4.

Arther, R. O. (1970). Spot Responders. *Journal of Polygraph Studies,* 4 (6), 1-2.

Arther, R. O. (1976). Known-Lie Test Chart Analysis. *Journal of Polygraph Science,* 11 (2), 1-4.

Arther, R. O. (1996). The Two Types of Spot Responders. *Journal of Polygraph Science,* 30 (4), 1-4.

Backster, C. (1974). Anticlimax Dampening Concept. *Polygraph,* 3 (1), 48-50.

Backster, C. (1963). The Backster Chart Reliability Rating Method. *Law & Order,* 1 (1), 63-64.

Backster, C. (1996). The Backster Zone Comparison Technique, An Update from the Source. Paper delivered at the American Polygraph Association Seminar, New Orleans.

Backster, C. (1964). Cardio Baseline Changes Sometimes Misleading. *Military Police Journal,* p. 24.

Backster, C. (1963). New Standards for Chart Interpretation – Do the Charts Speak for Themselves? *Law and Order,* 11 (6), 67-71.

Backster, C. (1964). Polygraph Spot Analysis versus Non-localized Analysis. *Law & Order,* 12 (2), 31-32.

Backster, C. (Sep 1963). Polygraph Technique Trouble Shooting. *Law & Order,* pp. 54-55.

Baesen, H. V., Chung, C. & Yang, C. (1948-49), A Lie Detector Experiment. *Journal of Criminal Law, Criminology, and Police Science,* 39, 532-537.

Barland, G. H. (1972). Reliability of Polygraph Chart Evaluations. *Polygraph,* 1 (4), 192-206.

Barland, G. H. (May 1981). A Validation and Reliability Study of Counterintelligence Screening Test. Security Support Battalion, 902nd Military Intelligence Group, Fort George G. Meade, MD.

Benussi, V. (1914). Die Atmung Asymptome der Luge (The Respiratory Symptoms of Lying). *Archiv fur die Gestamte Psychologie,* 11, 244-273. Translated and printed in *Polygraph,* 4 (1), 52-76.

Biran, I. (1975). Chart Interpretation: An Interdisciplinary Paradigm. *Polygraph,* 4 (1), 7-17.

Blackwell, N. J. (1994). An Evaluation of the Effectiveness of the Automated Scoring System (PASS) in Detecting Deception in a Mock Crime Analog Study. DoDPI 94-R-003.

Burtt, H.E. (1921). Further Techniques for Inspiration Expiration Ratios. *Journal of Experimental Psychology*, 4, 106-111.

Burtt, H. E. (1921). The Inspiration Expiration Ratio During Truth and Falsehood. *Journal of Experimental Psychology*, 4, 1-23.

Capps, M. H. & Ansley, N. (1992). Analysis of Federal Polygraph Charts by Spot and Chart Total. *Polygraph*, 21 (2), 110-131.

Capps, M. H. & Ansley, N. (1992). Analysis of Private Industry Polygraph Charts by Spot and Chart Total. *Polygraph*, 21 (2), 132-142.

Capps, M. H. & Ansley, N. (1992). Anomalies: The Contributions of the Cardio, Pneumo, and Electrodermal Measures Toward a Valid Conclusion. *Polygraph*, 21 (4), 321-340.

Capps, M. H. & Ansley, N. (1992). Comparison of Two Scoring Scales. *Polygraph*, 21 (1), 39-43.

Capps, M. H. & Ansley, N. (1992). Numerical Scoring of Polygraph Charts: What Examiners Really Do. *Polygraph*, 21 (4), 264-320.

Capps, M. H. & Ansley, N. (1992). Strong Control versus Weak Control. *Polygraph*, 21 (4), 341-348.

Capps, M. H. , Knill, B. L. & Evans, R. K. (1993). Effectiveness of Symptomatic Questions. *Polygraph*, 22 (4), 285-298.

Capps, M. H., Knill, B. L., Evans, R. K. & Johnson, G. J. (1995). Cognitive Arousal as a Means of Evaluation. *Polygraph*, 24 (1), 1-5.

Castelaz, P. F. & Angus, J. E. (1994). Deception Classification via Nonconventional Processing Techniques. *SPR Abstracts*, S12.

Clingan, L. (1980). To Score or Not to Score. *The Police Polygraphists Journal*, 23-25.

Decker, R. E. (Fall 1981). Chart Analysis. *C.A.P.E. Newsletter*, 1 (4), 5-24.

Decker, R. E., Stein, A. E. & Ansley, N. (1972). A Cardio Activity Monitor, *Polygraph*, 1 (3) 108-124.

Duston, B. M. (Dec 1990). A Comparison of Discriminant Analysis Techniques for Classifying Polygraph Charts. Naval Oceans Systems Center, San Diego.

Duston, B. M. (Nov 1992). Statistical Techniques for Classifying Polygraph Data. Naval Command, Control, and Ocean Surveillance Center, San Diego.

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, L. A. Jr. (1984). Analysis of Agreement in Polygraph Charts. *Polygraph*, 13 (2), 193-197.

Elaad, E. (1985). Decision Rules in Polygraph Examinations. *In Anti-Terrorism; Forensic Science; Psychology in Police Investigations; Proceedings of IDENTA '85.* Jerusalem: Heiliger & Co., 167-178.

Elaad, E. (1990). Detection of Guilty Knowledge in Real-Life Criminal Investigations. *Journal of Applied Psychology,* 75 (5), 521-529.

Elaad, E., Ginton, A. & Jungman, N. (1992). Detection Measures in Real-Life Criminal Guilty Knowledge Tests. *Journal of Applied Psychology,* 77 (5), 757-767.

Elaad, E. & Kleiner, M. (1990). Polygraph Chart Interpreter Experience on Psychophysiological Detection of Deception. *Journal of Police Science and Administration,* 17 (2), 115-123.

Franz, M. L. (1989). Technical Report: Relative Contributions of Physiological Recordings to Detect Deception. Contract No. MDA904-88-M-6612, Argenbright Polygraph, Inc.

Furgerson, R. M. (Oct 1989). Evaluating Investigative Polygraph Results. *FBI Law Enforcement Bulletin,* 58 (10), 6-11, reprinted in *Polygraph,* 18 (4), 208-215.

Gately, R. D. (1983). Specific Testing With the Atari 800 Computer. Unpublished manuscript.

Geddes, L. A. & Newberg, D. C. Cuff Pressure Oscillations in the Measurement of Relative Blood Pressure. *Polygraph,* 6 (2), 113-122.

Glassford, E. & Nyboer, J. (1975). Emotional Stress and Oscillometric Variations of the Pulse Curve. *Polygraph,* 4 (2), 108-117.

Golden, R. I. & Hunter, F. L. (Aug 1970). The Measurement of Upper and Lower Respiration in Lie Detection. Paper presented at the American Polygraph Association Seminar, Los Angeles.

Gordon, N. J. & Cochetti, P. M. (1987). The Horizontal Scoring System. *Polygraph,* 16 (2), 116-125.

Haney, J. W. (1943-44). The Analyses of Respiratory Criteria in Deception Tests, A Possible Source of Misinterpretation. *Journal of Criminal Law and Criminology,* 34, 268-273.

Haney, K. W. (1972). Question Spacing and the Length of Reactions. *Polygraph,* 1 (4), 234-237.

Harrelson, L. H. (1965). Chart Interpretation and Question Formulation. *In Keeler Polygraph Institute Second Annual Seminar.* Chicago: Keeler Polygraph Institute.

Harrleson, L. H. (1964). *The Keeler Technique,* 2nd ed. Chicago: Keeler Polygraph Institute.

Hathaway, S. R. & Hanscom, C. B. (1958). "The Statistical Evaluation of Polygraph Records." In V. A. Leonard (Ed.) *Academy Lectures on Lie Detection,* v.2. Springfield, IL: Charles C. Thomas, 112-121.

Hilliard, D. L. (1979). A Cross Analysis Between Relevant Questions and a Generalized Intent to Answer Truthfully Question. *Polygraph,* 8 (1), 73-77.

Holmes, W. D. (1958). "The Degree of Objectivity in Chart Interpretation." In V. A. Leonard (Ed.) *Academy Lectures on Lie Detection,* v.2. Springfield, IL: Charles C. Thomas, 83-85.

Honts, C. R. & Driscoll, L. N. (1988). A Field Validity Study of the Rank Order Scoring System (ROSS) in Multiple Issue Control Question Tests. *Polygraph,* 17 (1), 1-15.

Horvath, F. (1994). The Value and Effectiveness of the Sacrifice Relevant Question: An Empirical Assessment. *Polygraph*, 23 (4), 261-279.

Howland, D. P. (1978). An Application of the Positive Control Concept of Polygraph Examination. *Polygraph*, 7 (4), 286-294.

Howland, D. (1981). Positive Control Question Technique; Pre-Test Interview and Chart Interpretation. *Polygraph*, 10 (1), 37-41.

Inbau, F. E. (1942). *Lie Detection and Criminal Interrogation*. Baltimore: Williams & Wilkins Co.

Inbau, F. E. (1948). *Lie Detection and Criminal Interrogation*. Second Ed., Rev. Baltimore: Williams & Wilkins Co.

Inbau, F. E. & Reid, J. E. (1953). *Lie Detection and Criminal Interrogation*. Third Ed., Rev. Baltimore: Williams & Wilkins Co.

*Instruction Manual for the Berkeley Psychograph* (1943). San Rafael, CA: Lee & Sons.

Jayne, B. (1981). Purposeful Non-Cooperation: A Diagnostic Opinion of Deception. *Polygraph*, 10 (3), 156-174.

Joseph, C. N. (1957). "Compensatory Responses and Irregularities in Polygraph Chart Interpretation." In V. A. Leonard (Ed.) *Academy Lectures in Lie Detection*, v.1. Springfield, IL: Charles C. Thomas, 93-99.

Keeler, L. (1930). A Method for Detecting Deception. *American Journal of Police Science*, 1 (1), 38-51.

Keifer, R. (Jul 1993). Chart Interpretation. Paper presented at the American Polygraph Association Seminar, Newport Beach, CA.

Kircher, J. C. (1989). Response Quantification and Decision Making in the Detection of Deception: Current Status and Future Directions. *SPR Abstracts*, 26 (4A), 2-3.

Kircher, J. C. & Raskin, D. C. (1988). Human versus Computerized Evaluations of Polygraph Data in a Laboratory Setting. *Journal of Applied Psychology*, 73 (2), 291-302.

Kircher, J. C., Raskin D. C., Honts, C. R. & Horowitz, S. W. (1995). Lens Model Analysis of Decision Making by Field Polygraph Examiners. *SPR Abstracts*, S45.

Koll, M. (1979). Analysis of Zone Charts by Various Pairings of Control and Relevant Questions. *Polygraph*, 8 (2), 154-160.

Knapp, R. B. & Layeghi, S. (1994). Classification of Deception Using Fuzzy Pattern Recognition. *SPR Abstracts*, S12.

Kubis, J. F. (1973). Analysis of Polygraph Data, Dependent and Independent Situations. Part I. *Polygraph*, 2 (1), 42-58. Part II. *Polygraph*, 2 (2), 89-107.

Landis, C. & Gullette, R. (1925). Studies of Emotional Reactions. Systolic Blood Pressure and Inspiration-Expiration Ratios. *Journal of Comparative Psychology*, 5 (3), 221-253.

Landis, C. & Wiley, L. E. (1926). Changes in Blood Pressure and Respiration During Deception. *Journal of Comparative Psychology*, 6 (1), 1-19.

Larson, J. A. (1923). The Cardio-Pneumo-Psychogram in Deception. *Journal of Experimental Psychology*, 6 (6), 420-454.

Lee, C. D. (1953). *The Instrumental Detection of Deception – The Lie Test.* Springfield, IL: Charles C. Thomas.

MacNitt, R. D. (1942). Measures of Deception. *Journal of Criminal Law and Criminology*, 35 (3).

Magiera, A. C. (1975). "Patterns of Purposeful Distortions." In N. Ansley (Ed.) *Legal Admissibility of the Polygraph.* Springfield, IL: Charles C. Thomas, 199-209.

Marston, W. M. (1938). *The Lie Detector Test.* New York: Richard R. Smith. Reprinted by the American Polygraph Association, 1989, pp. 102-105.

Matte, J. A. (1996). *Forensic Psychophysiology Using the Polygraph, Scientific Truth Verification – Lie Detection.* Williamsville, NY: J.A.M. Publications.

Matte, J. A. (1980). *The Art and Science of the Polygraph Technique.* Springfield, IL: Charles C. Thomas.

Matte, J. A. (1981). Polygraph Quadri-Zone Reaction Guide. *Polygraph*, 10 (3), 186-193.

Minor, P. (Aug 1985). Modified General Question Technique (MGQT) Summary. Paper delivered at the American Polygraph Association Seminar, Reno, Nevada.

Minor, P. K. (Aug 1985). Modified Relevant-Irrelevant (MRI) Technique. Paper delivered at the American Polygraph Association Seminar, Reno, Nevada.

Minor, P. K. (1989). "The Relevant-Irrelevant Technique." In Abrams, S. *The Complete Polygraph Handbook.* Lexington, MA: Lexington Books, 143-152.

Minor, P. K. (Aug 1994). The Relevant-Irrelevant Technique. Paper delivered at the American Polygraph Association seminar at Newport Beach.

Minor, P. K. (Aug 1985). The Zone Comparison Test (ZCT). Paper delivered at the American Polygraph Association Seminar, Reno, Nevada.

Miritello, K. (Nov 1990). Rank Order Analysis. Unpublished manuscript.

Mohr, C. T. (1982). Problems Encountered in Testifying to Test Charts of Other Examiners. *Empire State Polygraph Society Newletter*, 6 (4), 8-11.

Moroney, William F. & Zenhausen, R. J. (1972). Detection of Deception as a Function of Galvanic Skin Response Recording Methodology. *Journal of Psychology*, 80 (2), 255-262.

Murray, K. E. (1989). Movement Recording Chairs: A Necessity? *Polygraph*, 18 (1), 15-23.

Nakayama, M. & Yamamura, T. (1990). Changes of Respiration Pattern to the Critical Question on Guilty Knowledge Technique. *Polygraph*, 19 (3), 188-198.

Nieman, D. A. (1982). Physiological Response to Rank Order of Problems as Measured by the Polygraph. Doctoral Dissertation, United States International University, UMI 43/03B, p. 856, DEC82-18282 vol.3, p.735.

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), 347-357.

Olsen, D. E., Harris, J. C., Capps, M. H. & Ansley, N. (1997). Computerized Polygraph Scoring System. *Journal of Forensic Sciences*, 42 (1), 61-71.

Olsen, D. E., Harris, J. C., & Chiu, Wendy W. (1994). The Development of a Physiological Detection of Deception Scoring Algorithm. *SPR Abstracts*, 31, S11.

Pierce, R. W. (1950). The Peak of Tension Test. *ISDD Bulletin*, 3 (3), 1, 4, 5.

Raskin, D. C., Barland, G. H. & Podlesny, J. A. (1977). Validity and Reliability of Detection of Deception. *Polygraph*, 6 (1), 1-39.

Reburn, J. W. & Mayo, J. F. (Apr 1975). The Effect of Noise on Polygraph Tracings. *Maryland Polygraph Review*, 8-9.

Reid, J. E. (1992). Controlled Breathing as an Indication of Deception. *Polygraph*, 11 (1), 12-13.

Reid, J. E. (1992). Interpretation of Truth and Deception in Polygraph Test Records. *Polygraph*, 11 (1), 46-61.

Reid, J. E. & Inbau, F. E. (1966). *Truth and Deception: The Polygraph ("Lie Detector") Technique.* Baltimore: Williams & Wilkins.

Reid, J. E. & Inbau, F. E. (1977). *Truth and Deception: The Polygraph ("Lie Detector") Technique.* Second Ed., Rev. Baltimore: Williams & Wilkins.

Robbins, N. E. & Penley, W. J. (1975). Review of Polygraph Charts of Non-Deceptive Subjects. *Polygraph*, 4 (3), 199-206.

Silverberg, B. A. (1980). A Note on the Diagnostic Value of the Dicrotic Pulse. *Polygraph*, 9 (2), 114-119.

Slowik, S. M. (1971). Comparison of Thoracic and Abdominal Respiration Deceptive Responses. Paper presented to the American Polygraph Association Seminar, Atlanta, Georgia.

Slowik, S. M., Buckley, J. P., Kroeker, L. & Ash, P. (1973). Abdominal and Thoracic Pneumograph Recordings. *Polygraph*, 2 (1), 12-27.

Summers, W. G. (1936). Guilt Distinguished from Complicity. *Psychology Bulletin*, 33, 787.

Summers, W. G. (1939). Science Can Get the Confession. *Fordham Law Review*, 8, 335-354.

Suter, James E. (1979). A Field Study of the Usefulness of the Cardio Activity Monitor. *Polygraph*, 8 (2), 112-119.

Suzuki, A. & Hikita, Y. (1981). An Analysis of Responses on Polygraph; A Dimunition of Responses. *Polygraph*, 10 (1), 1-7.

Suzuki, A., Ohnishi, K. Matsuno, K., & Arasuna, M. (1979). Amplitude Rank Score Analysis of GSR in the Detection of Deception: Detection Rates Under Various Examination Conditions. *Polygraph*, 8 (3), 242-252.

Suzuki, A., Watanabe, S., Ohnishi, K., Matsuno, K., & Arasuna, M. (1979). The Objective Analysis of GSR in the Detection of Deception. *Polygraph*, 8 (1), 53-63.

Thackray, R. I. & Orne, M. T. (1968). A Comparison of Physiological Indices in Detection of Deception. *Psychophysiology*, 4 (3), 329-339.

Timm, H. W. & Hedges, W. K. (1993). Factors Theoretically Affecting the Incidence of Deceptive Responses During Pre-employment Screening Procedures. *Polygraph*, 22 (3), 229-244.

Trovillo, P. V. (1942). Deception Test Criteria: How One Can Determine Truth and Falsehood from Polygraphic Records. *Journal of Criminal Law and Criminology*, 33 (4), 338-358.

Trovillo, P. V. (1942). Some Indices of Deception for the Interpretation of Polygraph (Lie Detector) Tests. *Psychological Bulletin*, 39, 599-600.

Turrou, L. G. (1938). *Nazi Spies in America*. New York: Random House.

United States Army Military Police School (USAMPS) (n.d.). Chart Interpretation – Chart Analysis, Classroom Issue, IN 330, IO 118.

United States Department of Defense Polygraph Institute (DoDPI) (1998). Instructions for Manually Scoring Charts Collected During a Psychophysiological Detection of Deception (PDD) Examination.

United States Department of Defense Polygraph Institute (6 June 1990). Peak of Tension (POT) Test Construction Summary Sheet.

van Herk, M. (1991). Numerical Evaluation: Seven Point Scale ± 6 and Possible Alternatives; A Discussion. *Polygraph*, 20 (2), 70-79.

Wainer, H. & Gruvaeous, G.T. (1974). Using Variance as a Discriminator in Lie Detection. *Journal of Applied Psychology*, 59 (1), 110-112.

Watanabe, S. & Suzuki, A. (1977). The Heart Rate as an Index of Lie Detecting Test. *Polygraph*, 6 (2), 103-112.

Watanabe, S. & Kajitani, M. (1995). Abstract: The Heart Rate Response as an Index of the Lie Detector Test. *Polygraph*, 24 (1), 56.

Weaver, R. S. (1985). Effects of Differing Numerical Chart Evaluation Systems on Polygraph Examination Results. *Polygraph*, 14 (1), 34-42.

Weaver, R. S. (1980). The Numerical Evaluation of Polygraph Charts: Evolution and Comparison of Three Major Systems. *Polygraph*, 9 (2), 94-108.

Wicklander, D. E. & Hunter, F. L. (1975). The Influence of Auxiliary Sources of Information in Polygraph Diagnosis. *Journal of Police Science and Administration*, 3, 405-409.

Winter, J. E. (1936). A Comparison of the Cardio-Pneumo-Psychograph and Association Methods in the Detection of Lying in Cases of Theft Among College Students. *Journal of Applied Psychology*, 20, 243-248.

Wygant, J. R. (1984). Rationale for Scoring. *Polygraph*, 13 (3), 263-266.

Yankee, W. J. (July 1968). A Report on the Computerization of Polygraphic Records. Paper presented to the Fifth Annual Seminar of the Keeler Polygraph Institute Alumni Association, Chicago.

Yankee, W. J. & Laughner, D. L. (1973). An Investigation of Mutual Relationships Between Sighs and Cardio Tracings in Chart Interpretation. *Polygraph*, 2 (1), 28-35.