

Utah Approach to Comparison Question Polygraph Testing

Mark Handler and Raymond Nelson

Abstract

In 1970, David Raskin, a psychologist and researcher at the University of Utah, began a study of the probable lie comparison question polygraph technique. Raskin and his colleagues systematically refined the elements of polygraphy by determining what aspects of the technique could be scientifically proven to increase validity and reliability (Raskin & Honts 2002). Their efforts culminated in the creation of what is known today as the Utah approach to the Comparison Question Test (CQT), an empirically consistent and unified approach to polygraphy. The Utah-CQT, was traditionally employed as a single issue Zone Comparison Test (ZCT). It is amenable to other uses including multi-facet testing of a single crime issue, as a Modified General Question Technique (MGQT) format, or as a multiple-issue (mixed-issue) General Question Technique (GQT). The Utah-CQT and the corresponding Utah Numerical Scoring System (Bell, Raskin, Honts & Kircher, 1999; Handler, 2006) resulted from over 30 years of scientific research and scientific peer-review. The resulting technique provides some of the highest rates of criterion accuracy and interrater reliability of any polygraph examination protocol (Senter, Dollins & Krapohl, 2004; Krapohl, 2006) when applied in an event-specific testing situation. The authors discuss the Utah-CQT using the Probable Lie Test (PLT) as well as the lesser known Directed Lie Test (DLT) and review some of the possible benefits offered by each method.

Test Structure and Administration

The Utah-CQT begins as other testing procedures do, with the pre-test interview, accomplished in a non-accusatory manner. The examiner should obtain the necessary test release that includes a brief statement of allegations or issues to be resolved, and if applicable, a statutory rights waiver and then collects general biographical and medical information from the test subject. Rapport-building discussion gives the examiner a chance to evaluate the test subject's suitability for the examination. Interaction with the test subject also gives the examiner the chance to do a rough assessment of the

test subject's verbal and mental abilities that will later be used to help word the examination questions. In the PLT version, the examiner uses this period of conversation to develop material for comparison questions to be used during the testing phase of the examination, although the nature of the issues to be resolved usually dictates the general content of the comparison questions. The examiner does not, however, lecture the test subject regarding past transgressions during this comparison question material review. This portion of the interview is conducted with open-ended questions and the careful use of suggestions as opposed to an interrogation of past deeds.

The version of this paper originally published in *Polygraph* was rewritten with greater detail for the journal *European Polygraph*, and the authors recommended the more detailed article for republication in *Polygraph* for this special edition. It appears here with the kind permission of the authors and the Editor of *European Polygraph*. The citation is: Handler, M. & Nelson, R., (2008). Utah approach to comparison question polygraph testing. *European Polygraph*, 2(2), 83-110.

The authors thank David Raskin, Charles Honts, Don Krapohl, John Kircher and Frank Horvath for their thoughtful reviews and comments to an earlier drafts of and revisions to this paper.

The examiner points out any monitoring or recording devices in the examination room and explains the purpose for having the exam monitored and/or recorded. In the Utah-CQT approach all examinations should be recorded in their entirety. In an age in which video and audio recording technology is easily available and fully integrated into all modern field polygraph systems, there is no reason to forgo the advantages of a complete video and audio recording of all polygraph examinations. It is only through complete recordings that meaningful quality assurance is possible. Frankness regarding monitoring devices helps assure the test subject the test will be conducted in a professional manner and may assist in convincing the test subject that the examiner is being open and truthful. Brief explanation of any quality assurance program also assists in establishing a professional and trustworthy atmosphere.

The examiner advises the test subject of the general nature of the allegations and the specific issues to be resolved by the examination. The test subject is then given the opportunity to provide a “free narrative” to discuss his or her knowledge of and/or role in the incident. The goal of the free narrative discussion is to obtain information from the test subject without confrontation or undue stress.

In general the examiner should allow the test subject to tell his or her story without interruption. The examiner informs the test subject of the case facts in a low-key approach and should advise the test subject that these are allegations and ensure the test subject understands the difference between allegations and facts known to be true.

The examiner should note inconsistencies or other matters to which he or she may wish to return once the test subject finishes the narrative. The examiner does not argue with the test subject nor does the examiner challenge the test subject’s version of the case facts. The examiner encourages the test subject to be candid in order to formulate the test questions in a succinct and clear manner.

In polygraph screening or monitoring programs (i.e., LEPET, security, PCSOT), the

Utah-CQT may be used as a mixed-issue (multiple-issue) examination, similar to the AFMGQT, in the absence of a known allegation or known incident. In these programs discussion of the known allegation or known incident will be replaced with a structured interview protocol, which addresses content areas pertinent to the risk or compliance issues under investigation. It should be noted that these applications of polygraph testing have not been investigated as thoroughly as other uses, and scientific investigation and verification of such uses are more limited.

This low key, non-accusatory approach presents the examiner as a neutral seeker of the truth and helps to allay fears of pre-conceived guilt. If there are inconsistencies or other matters that require follow-up or clarification before the examination, they are discussed at this time in a non-confrontational fashion.

After the narrative and the discussion of any other issues, the components are placed on the test subject. During this process, the functions of various polygraph component sensors are discussed, and a general explanation of the psychophysiology that underlies the polygraph test is provided. This may be done through a general discussion of the anecdotes that illustrate psychophysiological responding and various possible causes of arousal (Handler & Honts, 2007). The goal of this portion of the interview is to ensure in the test subject an understanding that lying will inevitably be associated with physiological response.

Once the components are placed on the test subject, the examiner conducts an acquaintance test. The acquaintance test is generally a known solution peak of tension test that is used to demonstrate the efficacy of the polygraph examination. Other approaches to the acquaintance test are not prohibited and would not invalidate an examination. In the known-solution acquaintance test, the test subject is told to select a number such that there will be some additional or padding questions before and after the selected number. This can be accomplished by directing the test subject to select a number between 3 and 6 and write that number on a piece of paper. The paper may then be

displayed in front of the test subject and the test subject is instructed to deny picking any number between 1 and 7 while the polygraph records his or her physiological reactions. The acquaintance test allows the examiner to ensure the production of adequate quality recordings and to take corrective actions to remedy any lack thereof.

The examiner can use the acquaintance test during the question review to demonstrate to the test subject that he or she is a suitable candidate for polygraph, and provide assurances that successful completion of the examination can be obtained by answering all of the test questions truthfully (in the PLC version of the examination) or that clear indication was found when the test subject was not answering truthfully (in the DLC version).

Following the acquaintance test the test questions are reviewed with the test subject for clarity. Some agency or local testing protocols may specify the test questions be fully reviewed prior to attaching any components to the test subject. There is no theoretical rationale to suggest this difference would invalidate an examination result. Attaching the sensors earlier may allow them to stabilize, especially the electrodes for electrodermal recording.

The examiner begins with the sacrifice-relevant question followed by the relevant questions. The sacrifice-relevant question is used to introduce the relevant issue under investigation during the testing and is not scored. In investigative polygraph testing, relevant question targets are dictated by the circumstances of the investigation and are commonly formulated around the most salient or intense aspects of the allegation. In screening programs, relevant questions should describe the test subject's involvement in possible behavioral concerns to risk managers or adjudicators and should be designed to add incremental validity to their particular program.

Polygraph screening targets would ideally be selected to investigate content areas pertinent to actuarial or empirically derived protocols for risk assessment and risk management. The fundamental requirement for relevant question target selection is that

the behavioral issue of concern provides information useful to the referring authority.

Effectively formulated relevant questions will directly assess the test subject's behavioral involvement in the issue of concern. Relevant questions should not introduce confusion through the use of language or concepts pertaining to psychological motivation or intent, as these are thought to introduce dimensions of excuse or rationalization on the part of test subjects or skillful liars. Conversely, truthful test subjects may produce spurious reactions because of the ambiguity and lack of concreteness of such questions. Direct questions with a simple grammatical structure are the best approach. Relevant questions should be free of idiomatic and legal jargon that is unfamiliar to the test subject, and should not include issues of psychological assessment or inference. Relevant questions are simple questions that can easily be answered "Yes" or "No." Reluctance, on the part of the test subject to provide a simple answer to a simple question may be an indicator of a non-testable issue or a test subject who is unable to disambiguate the issue. Discussion and resolution of this should be non-accusatory, but persistent enough to achieve a simple testable answer to a question that is behaviorally descriptive of the test subject's possible involvement in an issue of concern. The prevailing practice preference for relevant questions is they are usually answered "No," though certain exceptions have been suggested such as alleged victims of severe sexual assaults (Hardy & Murphy, 1996). The current authors found nothing to support that using "Yes" answered relevant questions would invalidate a test.

Next the examiner introduces the comparison questions. PLC questions are presented to the test subject as being necessary for further evaluating the test subject's character and the issue under investigation. PLC questions are based on transgressions whose subject matter is generally or conceptually related to the allegations of the examination and which virtually all persons may have committed, but which are likely to be denied in the context of the examination. PLC questions are broad in scope and usually based on actions

categorically similar to that of the issue under investigation. That is, relevant questions on theft would normally be associated with comparison questions about theft or general honesty. Relevant questions about violent acts are typically associated with comparison questions about causing harm. Standard comparison question construction, as taught in polygraph schools accredited by the American Polygraph Association and American Association of Police Polygraphists, is recommended for ensuring saliency. There is no reason, however, to prohibit the use of standard “lie” comparison questions in nearly any testing context.

Comparison questions in the Utah-CQT are traditionally “exclusive” in that they are separated from the relevant issue by time, place or category. Comparison questions not separated from the relevant issue are sometimes referred to as *non-exclusionary* type. Three studies (Horvath, 1988; Amsel, 1999; Palmatier, 1991) failed to establish any clear and consistent advantage of exclusionary comparison questions over non-exclusionary questions (Krapohl, Stern & Ryan, 2003). Podlesny & Raskin, (1978) showed some superiority for exclusionary questions, in that Skin Conductance Response (SCR) half-recovery time, SCR recovery half time width and Skin Potential Response (SPR) amplitude were significantly more effective with exclusive comparison questions.

Podlesny & Raskin (1978) also reported both types of comparison questions produced significant identification of innocent test subjects, but only exclusive comparison questions produced significant identification of guilty test subjects using numerical scores. Collectively these reports suggest that exclusionary comparison questions may hold no advantage over non-exclusionary comparison questions when data are evaluated using reaction criteria typically employed in field testing (for descriptions of those reaction criteria, see: Bell et al., 1999; Handler, 2006; Raskin & Honts 2002; Kircher et al., 2005). The use of exclusionary comparison questions may avoid possible criticism that the PLC questions are also relevant and may cause a false negative result. The current authors found nothing to suggest a test would be invalid should an

examiner choose to employ non-exclusionary type comparison questions.

As in other CQT techniques, the test subject is strongly, but indirectly, discouraged from making admissions to PLC questions. If the test subject makes an admission to a PLC question, the examiner notes that admission with some dismay, *“Really, you did something that would make me think you are a thief,”* and either minimizes the admission, “No, I am only concerned about serious things,” or modifies the comparison question. An example of the latter is: *“Other than what you told me about, before this year did you ever lie to anyone who trusted you?”*

Note the italicized modifier preceding the comparison question. The ultimate goal is to discourage admissions to PLC questions to ensure that the test subject perceives them as ambiguous and broad in nature. It is also important the examiner imply to the test subject that lying to any of the relevant or PLC questions will result in a failure of the polygraph test and the conclusion of deception to the relevant issue under investigation.

The examiner then introduces and reviews the neutral questions which provide time to return to a baseline when there is distortion or a physiological reaction to a specific question. Kircher, Kristjansson, Gardner, & Webb (2005) suggest inter-question intervals following a strong cardiovascular response should be increased to a minimum of 35 seconds to allow recovery, or a neutral question inserted. In general, the preferred approach is to wait to allow a return to, or at least toward, baseline levels. The neutral questions should be non-emotional in nature and are generally answered “Yes” to ensure the test subject is paying attention to the test questions. There is nothing to suggest, however, that an exam in which any neutral question is answered “No” would be invalid. The examiner may review additional neutral questions in case they are needed during testing to re-establish a baseline tracing.

The examiner next reviews the introductory question that is similarly worded to one of the “symptomatic” questions used in other CQT formats. The introductory question attempts to assure the test subject that no

un-reviewed questions will be asked during the examination and may allow an orienting response at the beginning of an examination. Research by Honts, Amato & Gordon, (2004) has failed to demonstrate the symptomatic question functions as described and may actually produce poorer accuracy, especially for innocent test subjects.

The consistent trend illustrated by these investigators and others suggests that the invention and addition of new types of questions should not be encouraged in an age of modern scientific polygraph testing unless research shows the efficacy of a new approach (Hilliard, 1979).

While it is wise for field examiners to adhere to the general principles and procedures taught in basic training, there is equal or greater wisdom in adapting field practices to conform to modern approaches with proven validity. We do not believe that minor departures from the above question sequences would cause a test to be invalid, and the varying formulations of the Utah-CQT since its emergence suggest that the scientists who developed the Utah-CQT method did not seek validity through simplistic adherence to a "paint-by-numbers approach", but sought demonstrable validity through the construction of CQT methods according to sound testing principles.

Three-Question Format

The Utah-CQT has two versions, a three-question version and a four-question version (Raskin & Honts, 2002).

The three-question version was the first designed and was primarily used for single-issue testing but can also be used for multiple-facet testing of a single known allegation. The three-question version of the Utah-CQT allows a great degree of flexibility in relevant question format.

The following describes an example of question numbering and type of question used in the three-question version of the Utah-CQT. For a *single-issue* examination, there will be three relevant questions, each slightly reworded.

Example of a Utah PLT 3-question wording

For an event-specific, *single-issue* test surrounding a bank robbery occurring last Thursday, one might ask the following questions:

Introductory 1: Do you understand I will only ask you the questions we discussed?

Sacrifice Relevant 2: Regarding whether or not you robbed that bank do you intend to answer all of these questions truthfully?

Neutral 1: Are the lights turned on inside of this room right now?

Comparison 1: (*Before turning X*), Did you ever do anything that was dishonest or illegal?

Relevant 1: Did you rob that bank located at ___ in Austin?

Neutral 2: Are you now physically located within the State of Texas?

Comparison 2: (*Between the ages of X and Y*), Did you ever take anything that did not belong to you?

Relevant 2: Did you rob that bank located at ___ in Austin last Thursday?

Neutral 3: Do you sometimes listen to music while riding in a car?

Comparison 3: Did you ever take anything from a place where you worked, (*before age X*)?

Relevant 3: Did you rob that bank at ___ on ___?

The examples above are shown with the exclusionary clause of the comparison question in brackets.

For a *multiple-facet* examination, the examiner has a choice of asking two reworded relevant questions with the same meaning and another relevant question that is directly related to the issue under investigation. This third relevant question can be an evidence-connecting, guilty knowledge or secondary involvement question.

A third alternative is to ask three separate relevant questions relating to the same specific issue under investigation. Readers are reminded that research has shown that accuracy rates are higher for tests in which the test subject is either completely truthful or deceptive to all of the test questions as opposed to just some of them (Honts, Kircher, & Raskin, 1988; Raskin, Kircher, Honts, and Horowitz, 1988; Barland, Honts and Barger, 1989).

The current authors would strongly recommend examiners, if possible, attempt to limit the examination to one in which the test subject is truthful or deceptive to all of the relevant questions.

If one were to construct a *multiple-facet* polygraph examination surrounding a single crime event involving a bank robbery, examples of alternative relevant questions may be:

Introductory 1: Do you understand I will only ask you the questions we discussed?

Sacrifice Relevant 2: Regarding whether or not you robbed that bank do you intend to answer all of the questions truthfully?

Neutral 1: Are the lights turned on inside of this room right now?

Comparison 1: Did you ever steal anything from someone who trusted you?

Relevant 1: Did you rob that bank at ___ on ___?

Neutral 2: Are you now physically located within the State of Texas?

Comparison 2: Did you ever steal anything from a friend or family member?

Relevant 2: Did you plan or arrange with anyone to rob that bank at ___?

Neutral 3: Do you sometimes listen to music while riding in a car?

Comparison 3: Did you ever steal anything from a place you worked?

Relevant 3: Did you participate in any way in the robbery of that bank?

Note that this example is provided with *non-exclusionary* comparison questions.

Four-Question Format

The four-question format is similar in design to a version of the Air Force Modified General Question Technique (DoDPI 2006) using pairs of relevant questions that are bracketed by comparison questions. This allows the examiner greater flexibility covering more than one aspect of the relevant issues and in scoring by using the surrounding comparison questions. The relevant questions can range from one to four distinct behavioral aspects or facets of a single crime or allegation. The question construction rules are the same as those described above for the multiple-facet version of the three-question version.

The following describes an example of question numbering and type of question used in the four-question version.

I1 Introductory
SR2 Sacrifice Relevant
N1 Neutral
C1 Comparison
R1 Relevant
R2 Relevant
C2 Comparison
N3 Neutral (optional) This neutral question may be inserted at the option of the examiner to allow some decrease of tension and recovery to baseline. If inserted, the examiner will skip over this neutral question during scoring.
R3 Relevant
R4 Relevant
C3 Comparison
N2 Neutral

In Test Operation

The test subject is instructed to sit still and answer each question truthfully. However, the approach is to avoid doing this in a heavy-handed manner. For example the following admonition would be typical for this approach:

"I need you to sit still during the asking of the questions. Movement will create distortion and artifacts in the recordings that will require me to repeat the questions and that will make the test longer."

The examiner rotates the neutral, comparison, and relevant (if desired) questions during the next and subsequent presentations. The examiner may prefer leaving the relevant questions always in the same position, and rotating only the comparison and neutral questions, making it easier to score the charts by having a fixed order of relevant questions. Moving the questions helps to prevent pattern recognition and anticipation of a specific order of questions during the examination.

The following are examples of serial positioning in the question strings showing one example of question rotation.

Three Question Version

First Chart

I1,SR2,N1,C1,R1,N2,C2,R2,N3,C3,R3

Second Chart

I1,SR2,N2,C3,R2,N3,C1,R3,N1,C2,R1

Third Chart

I1,SR2,N3,C2,R3,N1,C3,R1,N2,C1,R2

Four Question Version

First Chart

I1,SR2,N1,C1,R1,R2,C2,N3
(N3 is optional),R3,R4,C3,N2

Second Chart

I1,SR2,N2,C2,R1,R2,C3,N3
(optional),R3,R4,C1,N1

Third Chart

I1,SR2,N1,C3,R1,R2,C1,N3
(optional),R3,R4,C2,N2

As can be seen above, each relevant question has an opportunity to be compared to each comparison question across the three chart series. As discussed above, if the results are inconclusive after three charts, two additional charts are run. The examiner may simply use the first and second serial positioning question strings for the fourth and fifth chart.

After the third chart, the charts are numerically scored. However, the test subject is only told the examiner always stops at this point to carefully check the quality of the recordings before collecting more charts. If the scores meet the threshold of the decision criteria, the data collection phase is complete. If the test result is inconclusive following the first three charts, two additional charts are conducted following the same rotational patterns described above. Following the fifth chart, all scores are totaled to make a determination of veracity. The Defense Academy for Credibility Assessment (DACA), the Federal Training facility, (2006) permits the examiner to conduct just a fourth chart, and if necessary a fifth test chart. We are aware of no theoretical rationale or evidence to suggest this procedural difference would invalidate a test result.

The questions are presented to the test subject at least three times across three charts, with a brief discussion between charts to clarify and resolve any perceived problems raised by the test subject and to reinforce a focus on both the relevant and comparison questions (Raskin & Honts, 2002). After each presentation of the test questions, the examiner should ask the test subject if he or she has any concerns with the test questions. Honts (1999) reviewed data from 19 studies that involved 1092 polygraph tests and found between chart discussion (even when limited to only the comparison questions) reduced the risk of error. In 9 of the 11 studies which included examinations in which the questions were reviewed between charts both the relevant and comparison questions were reviewed. In 2 of those studies, only the comparison questions were reviewed. Honts showed that between-chart stimulation and question review reduced the false negative rate (54%), had a modest reduction of false positive rates (2.9%) and a substantial decrease in inconclusive outcomes for truthful test subjects (42%).

The following is typical of the type of exchange that might take place between charts.

Note how the examiner places equal emphasis on each group of questions during the stimulation and review.

Examiner: OK Roy, did you have any problems with any of those questions on the test?

Roy: No.

Examiner: Anything come to mind when I asked you those questions?

Roy: No.

Examiner: How about those questions about the drug transaction? Is it clear what I am asking you? Do you understand them?

Roy: Yep.

Examiner: How about those questions about lying? Any problem with any of those?

Roy: Nope.

There has been controversy surrounding the review of question between test charts. Abrams (1999) and Matte (2000) argued that review of comparison questions between tests is incorrect and Offe & Offe (2007) found no contribution to improved or degraded decision outcomes as a result of between test review of the test questions. In consideration of these findings, we feel it prudent to recommend a review of the questions between each chart, but find no reason to support an argument that the inclusion or exclusion of this review would cause a test result to become invalid or erroneous. Honts (1999) did not speculate as to the psychological cause of these findings. He correctly stated "The essence of science is empiricism. That is, scientific knowledge is built on data, not speculation nor authority." Honts chose to accept the data for what it stated on its' own merit.

Should a test subject make additional admissions to comparison questions or need to modify a relevant question, the examiner should do so and re-label the question. For example, if during a Utah PLT the test subject makes an admission to question C1 "Before this year did you ever steal anything from a business," the examiner can modify that question to "Other than what you told me about, before this year did you ever steal

anything from a business" and label that question C1a. The examiner should then review all test questions with the test subject. The examiner conducts the next two charts and again starts by instructing the test subject to sit still and answer all of the questions truthfully.

Test Data Analysis and Decision Criteria

The Utah Scoring System (Bell, Raskin, Honts & Kircher, 1999) is a simplified version of the numerical scoring techniques introduced by Backster in 1963 and modified by the US Army around 1970 (Weaver 1980; Swinford 1999).

The Utah scoring system is a simple and elegant scoring system designed to improve accuracy, reduce inconclusive results, and improve interrater reliability. It has fewer rules to follow and fewer criteria to score than the other scoring systems currently in use. The Utah Numerical Evaluation Scoring System was designed, refined and tested by Raskin and his colleagues.

The Utah scoring System is based on physiological response data that has been proven to be a valid and reliable indicator of sympathetic arousal. The inter-scorer correlations of results produced using the Utah Scoring system are typically around 0.90 (Bell et al., 1999). The accuracy of the Utah Scoring system from several analog studies was 90%, as reported by Bell et al., when averaged for programmed innocent and guilty test subjects. The results of field studies using the Utah scoring system are consistent with analog study results (Bell et al. 1999)

Numerical evaluation of the test data is accomplished by comparing the relative strengths of responses to comparison and relevant questions. The Utah system uses a 7-position numerical scoring approach. The relative strengths of physiological reactions for each sensor are compared and a score is assigned. The possible scores range from -3 to +3. The reaction of each relevant question is compared to the reaction to the preceding comparison question in the 3-question CQT format or the stronger of the two surrounding comparison questions in the 4-question MGQT format. If the relative strength of the

relevant question is greater than that of the comparison question, a negative value is assigned.

Conversely if the comparison question strength exceeds the relevant question strength, a positive score is assigned. If there is no observable difference, a zero is assigned. In some components there are minimum relative ratios that must be achieved in order to assign a score.

For the three-question version shown above, the relevant question is normally compared to the preceding comparison question for evaluation. If the preceding comparison question is distorted by an artifact, the examiner may use the closest artifact-free comparison question for evaluation.

For the four-question version shown above, the examiner compares the relevant question to the two bracketing comparison questions, component by component. For example, in the first chart of the four-question version shown above, R1 is compared to C1 and C2. The examiner will find the strongest reaction channel separately of each channel for C1 and C2 and use that to compare to the corresponding channel of R1. Using the reaction of the stronger bracketed comparison question has been shown to produce valid field results (Honts 1996; Raskin et al. 1988).

Physiological tracings that are affected by artifacts are excluded for evaluation purposes. If the test subject answered "Yes" to a comparison question during the test, the comparison question response may be used in scoring as long as the reviewed answer had been "No." (see Honts, Raskin, & Kircher, 1992). The examiner may insert a neutral question routinely after the second comparison question or any other time needed to reestablish tracing stability. During test data analysis, the examiner will skip over that neutral question.

The Utah Scoring System uses a total of seven primary scoring criteria in the respiration, cardiograph, electrodermal, and peripheral vasomotor activity channels. Values of -3, -2, -1, 0, +1, +2, and +3 are assigned by channel to each relevant question. As mentioned above, if the relevant

question is the larger of the two, the score will be a negative number. If the comparison question is the stronger of the two, the score will be a positive number, and no difference yields a score of zero.

Only one score of +/-3 can be assigned per chart, in the cardio and electrodermal channel, and only if the baseline for the channel is stable and the reaction is the largest in that channel on the chart. The relevant question totals are calculated after three charts and, if inconclusive, after five charts.

For the respiration channel, there are four empirically confirmed features that are considered diagnostic (ASTM 2005). Three of those features are captured by the phenomenon known as Respiration Line Length "RLL" (Timm, 1982). RLL is simply the measurement of the length of the respiration line for a fixed period of time. The total line length for the designated period of time between the relevant and comparison question or questions is compared. The greater the suppression the shorter the line length, thus the stronger the response. Those three features are suppression of respiration amplitude (Figure 2), reduction in the respiration rate (which includes changes in the inhalation/exhalation ratio if they result in respiration rate decreases, Figure 3) and apnea occurring near the exhalation cycle (Figure 4). The fourth respiration criterion is a temporary rise in the baseline of the tracing.

A respiration tracing is considered to be diagnostic if there are at least three successive cycles of an RLL feature or temporary baseline arousal. The exception to this is apnea, where there may not be any discernible cycles of respiration. While the thoracic and abdominal respirations are recorded separately, a single value is assigned. That value is based on the noted combined difference between the relevant and comparison questions.

The developers of the Utah-CQT have taught and practiced conservatism when evaluating the respiration channel. Bell and his colleagues used a sample of 50 polygraph examinations to conduct a survey that provided 450 numerical scores. Bell et al. (1999) tallied those scores to determine the

distribution of scores, and reported that respiration scores of 0 were assigned about 75% of the time. Scores of +/-1 were assigned about 20% of the time, and +/-2 or 3 less than 5% of the time.

For the electrodermal channel, scores are based primarily on a comparison of the peak amplitude (Figure 6), a criterion that has been empirically shown to be diagnostic. Amplitude is measured from the pre-stimulus baseline to the highest peak achieved within the scoring window (Bell et al., 1999). The ratio of the relevant and comparison question is calculated.

A score of +/-1 is assigned if the relative strength is twice as large, a score of +/-2 is assigned if the relative strength is three times as large and a score of +/-3 is assigned if the relative strength is four times as large. If the electrodermal tracing is labile, a score of 3 should not be assigned.

Duration of response and complexity can be considered as secondary reaction criteria. Reactions that have clearly longer duration or complexity may increase a 0 to a +/-1 or a +/-1 to a +/-2 (Figures 7 & 8). If the amplitude ratios are at least 1.5:1 with complexity over no complexity or increased duration of reaction time, this allows an increase of a score of 0 to +/-1. Similarly, a ratio of at least 2.5:1 to increase a score of +/-1 to +/-2 following the same rules regarding increased complexity or duration. Bell et al. noted in the electrodermal channel scores of 0 were assigned about 50% of the time, scores of +/-1 about 25%, +/-2 about 20 % and +/-3 less than 10% of the time.

Krapohl and Handler (2006) demonstrated that additional diagnostic information can be gained from interpreting smaller differences in response to relevant and comparison questions using federal ZCT examinations. While these were not of the Utah-CQT format, the current authors know of no reason to believe improved ratio values for scoring would not be amenable in a Utah-CQT format. Improved ratios are being investigated at the time of this writing.

For the relative blood pressure channel, relative strengths of reactions are assessed based on upward movement from baseline (baseline arousal) as shown in Figure 9. A minimum ratio of 1.5:1 is required for a score of +/-1; a ratio of 2:1 for a score of +/-2; and 3:1 for a score of +/-3. Bell et al. (1999) reported duration of the response may be considered when evaluating the relative strength of the reaction and a reaction with greater duration may increase a score from 0 to 1 or from 1 to 2 (see Figure 10).

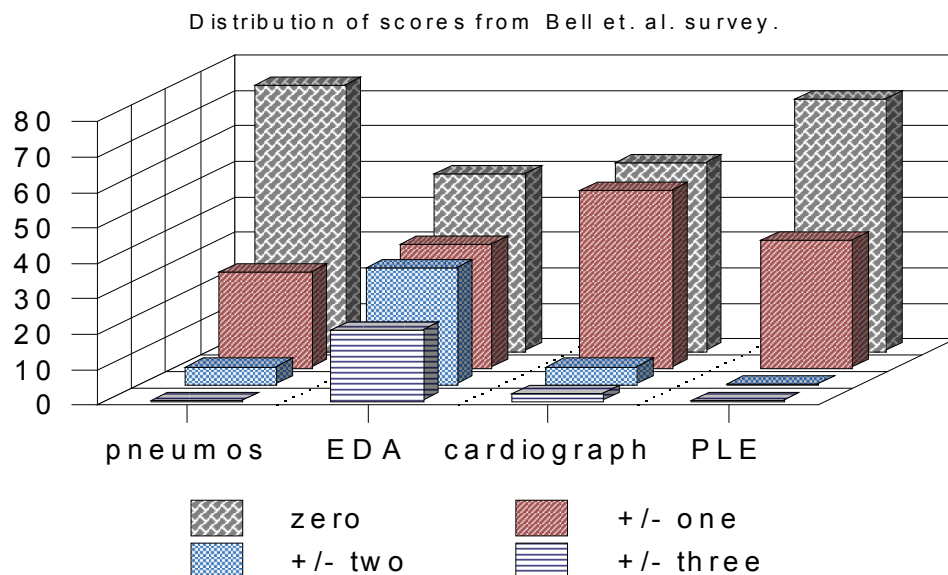
They did not, however, discuss the procedure for applying this rule and the current authors suggest limiting scoring of this channel to baseline arousal as suggested by Kircher et al (2005), Harris, Horner and McQuarrie (2000) and ASTM (2002). Bell et al. noted in relative blood pressure scores of 0 were assigned about 50% of the time, scores of +/-1 about 45%, +/-2 less than 5 % of the time. Scores of +/-3 are rare and only one such score can be assigned per chart as explained in the electrodermal section (Bell et al. 1999).

For the peripheral vasomotor activity, the relative strength of the reactions is assessed by comparing the reduction in pulse amplitude (Figure 11). The source of this channel is a photoplethysmograph monitoring reduction in finger pulse amplitude. Numerical scores are based on the duration and degree of amplitude reduction.

Scores may be assigned when there is no difference in amplitude decrease but a discernable difference in duration of the reactions (Figure 12). Bell et al. noted in scoring finger pulse amplitude scores of 0 were assigned about 70% of the time and scores of +/-1 about 30%, (Bell et al. 1999).

The following graph shows the distribution of the numerical scores obtained during the survey by Bell et al. al (1999). As can be seen from the graph the majority of numerical scores assigned are zero or +/- 1 for most channels.

Figure 1. The distribution of scores from the Bell et al. survey.



Decision Criteria

The examiner proceeds through the charts and totals the score for each relevant question on each chart. The total score of each relevant question for the first three charts is then determined. For single-issue tests where the test subject must be truthful or deceptive to all of the relevant questions, the cutting score is ± 6 . In other words, when there is a grand total of $+6$ or greater, the result is truthful. A grand total of -6 or less would result in a determination of untruthful or deception indicated. Scores falling between -5 and $+5$ would result in a determination of inconclusive and the examiner would conduct an additional two charts as described above.

Following those two additional charts, the relevant question scores are once again totaled. The cutting scores of ± 6 remain the same for five charts.

The decision criteria are slightly different for multiple-faceted examinations where the test subject may be truthful to some, but not all, of the relevant questions. If the spot totals for all relevant questions are either all positive or all negative (ignoring spot scores of zero), use the ± 6 Grand Total rule described above for single-issue tests.

If any of the spots are opposite (some positive and some negative, again ignoring spot scores of zero), then use a Spot Score Rule (SSR) for each spot. The SSR is that each spot total must be $+3$ for a conclusion of no deception indicated (NDI), and any one spot total of -3 or less calls for a decision of deception indicated (DI) to the examination. However, if decisions are made on individual questions caution is warranted as research indicates when test subjects answer some questions truthfully and some deceptively the accuracy for calls on individual questions is reduced (see the discussion in Raskin & Honts, 2002.) The problem can be exacerbated when attempting to verify truthfulness to one or more questions when total scores for any relevant question have indicated deception (Raskin & Honts, 2002).

While there has been a consistent effort to evaluate the empirical validity of various cut scores, little emphasis has been placed on the determination of statistically determined cut scores, in the manner of a Gaussian signal detection model, as described by Barland (1985). Krapohl and McManus (1999), Krapohl (2002) and Nelson, Handler and Krapohl (2007) are an exception to this trend.

Utah Directed-Lie Test

DLC questions are those which the examiner instructs the test subject to answer falsely (Honts & Raskin, 1988; Raskin & Honts, 2002). DLC questions may offer some relief to potential problems identified in PLC versions of polygraph testing. Examiners may experience difficulty in standardizing comparison questions in the PLC version. Each test subject brings with them his or her own life experiences and idiosyncrasies that may hamper maintaining a rapport while attempting to lay foundation for and set the PLC questions. Test subjects who have prior polygraph experience or those who have researched polygraph techniques may not be naïve to the PLC principles. This sophistication could make laying the foundation for the comparison questions challenging. Non-naïve test subjects may acquiesce to the procedure in order to not seem obstreperous in which case the PLC questions become similar to DLC questions.

One theory behind the DLC approach is that it is similar to the PLC approach in that it is assumed the subject's cognitive and (possibly some emotional) attention will be focused more on the questions that pose the greatest concern of failing the test. Thus, the truthful subject will be more concerned with whether or not they are a suitable subject and are producing appropriate responses to the DLC questions to show when they are lying. Though they have permission to lie on these questions, the questions still serve to draw the examinee's attention during testing. The theory further proposes the guilty subject will remain primarily concerned about the relevant questions on the test and will thus produce the greatest reactions to them. In this sense, the DLC questions operate as a distracter item for the truthful subjects, who are more capable of being distracted away from the relevant questions than are the deceptive subjects.

DLC questions are easily standardized, require little psychological manipulation and have greater face validity. Standardization and simplification of any technique can serve to increase inter-rater and test-retest reliability, and both of these dimensions constrain the potential validity of a technique. Excessive variability in test administration or

interpretation will necessarily compromise the reliability and validity of any test method. Inter-rater reliability is a concern that will remain of paramount importance to questions about polygraph validity. When standardized practices are based on principles that are consistent with validated constructs and data obtained through the objective study of data, we can more reasonably anticipate that improvements will contribute meaningfully to the test design goal of criterion validity and decision accuracy.

There are far fewer field and laboratory studies that address validity of the DLC than the PLC. However, the results of existing studies (Barland, 1981; Barland et al., 1989; DoDPI Research Division Staff, 1997; DoDPI Research Division Staff, 1998; Honts & Raskin, 1988; Horowitz, Kircher, Honts & Raskin, 1997; Kircher, Packard, Bell & Bernhardt, 2001; Reed, 1994; Raskin & Kircher, 1990) suggest that the DLC questions perform as well or better than PLC questions. DLCs require less complex administration practices than those associated with the PLC approach and offer greater potential for standardization. Studies using DLC techniques (DoDPI Research Division Staff, 1997; Research Division Staff, 1998) suggest that a DLC approach and other improvements in test administration structure and decision policies contributed significantly to polygraph testing program objectives of sensitivity to deception and specificity to truthfulness.

There are certain caveats that attend the use of DLC testing. First, examiners with no familiarity with DLCs should seek instruction in their proper development and introduction. Second, there is some indication in the research data that at least some test subjects show unusual respiration responses with the DLC (see Horowitz et al. 1997; Kircher 2001). However, standard numerical scoring procedures in Horowitz et al.(1997) performed well with the DLC, although it may not be optimal and research exploring this issue is currently underway. Moreover, there are currently no computer algorithms available that have been trained on DLT data. Therefore, the results of those models should be viewed cautiously.

Examples of Scoring Criteria

Respiration

Figure 2 below shows an example of suppression of respiration amplitude.

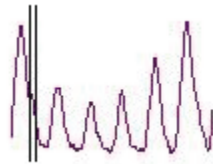


Figure 3 below shows an example of reduction in respiration rate.

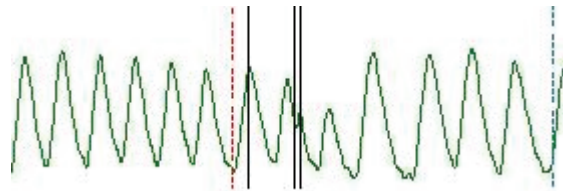
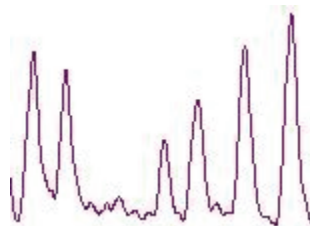
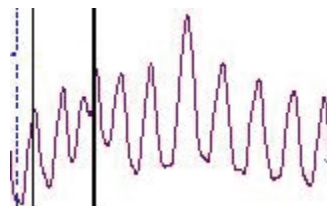


Figure 4 below shows an example of apnea occurring at or near exhalation.



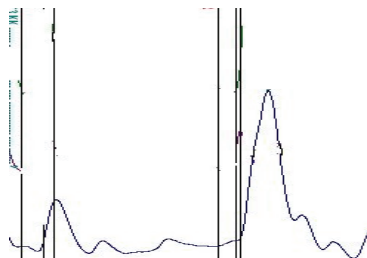
Note: The above three reaction criteria are those that are captured by the phenomenon known as RLL.

Figure 5 below shows an example of temporary baseline arousal.



Electrodermal Activity

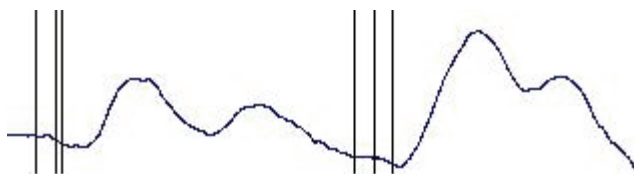
Figure 6 below shows an example of amplitude Increase.



Note the ratio of the above tracings is about 2.6:1 (26 mm vs. 10 mm in amplitude) with obviously longer duration and complexity

on the recovery side of the tracing. This ratio would qualify for a score of +/-2 based on the increased duration and complexity.

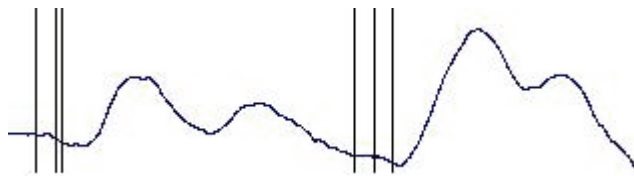
Figure 7 below shows an example of increased duration.



Note the ratio of these tracings is about 1.8:1 (18 mm vs. 10 mm in amplitude) with obviously longer duration and complexity on the recovery side of the tracing. The amplitude ratio does not exceed the 2:1

normally required for a score of +/-1. This would qualify for a score of +/-1 based on the ratio of at least 1.5:1 with greater duration and complexity.

Figure 8 below shows an example of complexity.



Note the ratio of the tracings in figure 8 is about 1.8:1 (18 mm vs. 10 mm in amplitude) with obviously longer duration and complexity on the recovery side of the tracing. The amplitude ratio does not exceed the 2:1

normally required for a score of +/-1. This would qualify for a score of +/-1 based on the ratio of at least 1.5:1 with greater duration and complexity.

Cardiograph

Figure 9 below shows an example of baseline arousal.

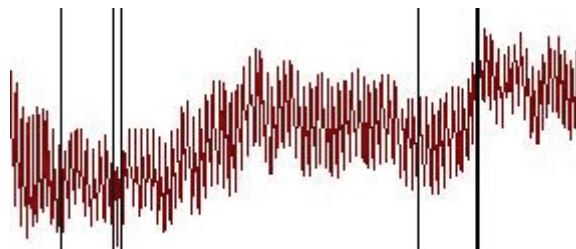
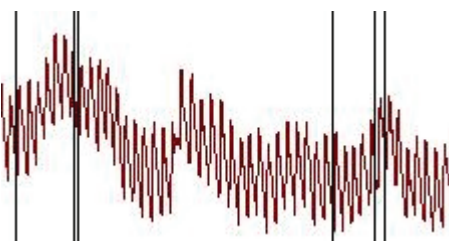


Figure 10 below shows an example of increased duration of response.



Note the ratio of the tracings in figure 10 is about 1.3:1 (8 mm vs. 6 mm in amplitude) with obviously longer duration and complexity on the recovery side of the tracing.

The amplitude ratio in figure 10 does not exceed the 1.5:1 normally required for a score of +/-1. This ratio would qualify for a score of +/-1 based on the greater duration observed in the first reaction.

Photoplethysmograph

Figure 11 below shows an example of amplitude reduction.

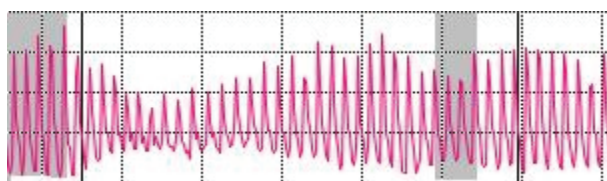
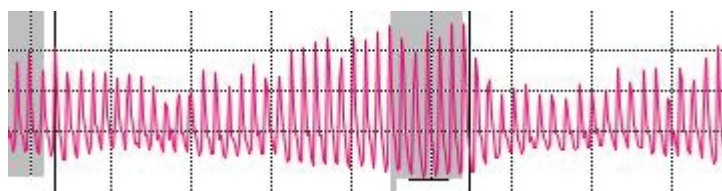


Figure 12 below shows an example of increased duration of amplitude reduction.



Conclusion

The Utah CQT was created by psychologist/examiners and founded upon known and proven principles of psychology and psychophysiology. The reliability and validity of the Utah CQT has been demonstrated in many peer-reviewed and published scientific studies (see the review in Raskin & Honts, 2002). A number of writings may be found in scientific journals and texts discussing the Utah-CQT (for example, Raskin & Honts, 2002; Bell et. al. 1999; Handler, 2006). We hope we consolidated some of those writings into a basic description of how to properly administer and evaluate the examination.

Those scientists who created and refined the technique took great pains to thoroughly research and assess the reliability and validity of the examination. This included numerous field and analog studies conducted over three decades. The Utah Scoring System

(Bell et al., 1999; Handler 2006) takes a somewhat conservative approach to assigning values. This ensures that scores are assigned to reactions that are clearly different in comparison and not arbitrarily assigned. Some argue this conservative approach may result in an inconclusive finding after three charts and thus require the additional two charts be conducted. From a scientific standpoint, more data is better and the additional two charts should serve to increase confidence in the results.

It is the sincere hope of the authors that others in the field of polygraphy will consider learning and using the Utah CQT. The more we move our profession toward techniques that employ scientifically validated principles, the more respect we will gain from others outside of the polygraph profession. As in any scientific field, progress can be made through the refinement of proven techniques already in place.

References

- Abrams, S. (1999). The role of polygraphy with the professional board. *Polygraph*, 28, 300-303.
- Amsel, T. T. (1999). Exclusive or nonexclusive comparison questions: A comparative field study. *Polygraph*, 28, 273-283.
- ASTM (2002) E2229-02 Standard Practices for Interpretation of Psychophysiological Detection of Deception (Polygraph) Data. ASTM International.
- ASTM (2005). E2324-04 Standard Guide for PDD Paired Testing. ASTM International.
- Barland, G. H. (1981). A validation and reliability study of counterintelligence screening test. Unpublished manuscript, Security Support Battalion, 902nd Military Intelligence Group, Fort Meade, Maryland.
- Barland, G. H. (1985). A method of estimating the accuracy of individual control question polygraph tests. Proceedings of IDENTA, 1985, 142-147.
- Barland, G.H., Honts C.R. and Barger, S.D. (1989) Studies of the Accuracy of Security Screening Polygraph Examinations. Department of Defense Polygraph Institute, Fort McClellan, Alabama.
- Bell, B.G., Raskin, D.C., Honts, C.R., & Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph*, 28(1), 1-9.
- Department of Defense. (2006). Federal Psychophysiological Detection of Deception Examiner Handbook. Retrieved on 03/13/2008 from; <http://antipolygraph.org/documents/federal-polygraph-handbook-02-10-2006.pdf>
- DoDPI (2006) Air Force Modified General Question Technique. Presentation at the Senior Polygraph Examiner Course, Austin, TX.
- DoDPI Research Division Staff. (1998). Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage. *Polygraph*, 27, 68-73.
- DoDPI Research Division Staff, (1997). A Comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope polygraph and the test for espionage and sabotage question formats. *Polygraph*, 26, 79-106.
- Handler, M. (2006). Utah Probable Lie Comparison Test. *Polygraph*, 35, (3) 139-148.
- Handler, M. and Honts, C. R. (2007). Psychophysiological Mechanisms in Deception Detection: A Theoretical Overview. *Polygraph*, 36 (4), 221-232.
- Hardy, W.L. and Murphy, V.T. (1996). Victims and Polygraph Examinations. *Polygraph*, 25 (4).
- Harris, J, Horner, A & McQuarrie, A (2000), An Evaluation of the Criteria Taught by the Department of Defense Polygraph Institute for Interpreting Polygraph Examinations Prepared under contract DABT02-96-C-0012 – CLIN0004-Phase 11 for: The Department of Defense Polygraph Institute
- Hilliard, D. L. (1979). A cross analysis between relevant questions and a generalized intent to answer truthfully question. *Polygraph*, 8, 73-77.

- Honts, C.R. (1996). Criterion development and validity of the control question test in field application. *The Journal of General Psychology*, 123, 309-324.
- Honts, C.R. (1999) The discussion of comparison questions between list repetitions (charts) is associated with increased test accuracy. *Polygraph*, 28, 117-123.
- Honts, C., Amato, S. & Gordon, A. (2004). Effects of outside issues on the Control Question Technique. *Journal of General Psychology*, 151, 53-74.
- Honts, C.R., Kircher, J.C., & Raskin, D.C., (1988) Patterns of activation and deception. *Psychophysiology*, 25, 455.
- Honts, C. & Raskin, D. (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. (1992). Effectiveness of control questions answered "Yes": Dispelling a polygraph myth. *Forensic Reports*, 5, 265-272
- 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.
- Horvath, F. S. (1988). The utility of control questions and the effects of two control question types in field polygraph techniques. *Journal of Police Science and Administration*, 16, 198-209.
- Kircher, J. C., Packard, T., Bell, B. G. & Bernhardt, P.C. (2001,). Effects of prior demonstration of polygraph accuracies on outcomes of probable-lie and directed-lie polygraph tests. DoDPI02-R-0002.
- Kircher, J.C., Kristjansson, S.D., Gardner, M.G., Webb, A. (2005) *Human and Computer decision-making in psychophysiological detection of deception*. Submitted to the Department of Defense Polygraph Institute.
- Krapohl, D. (2002). Short Report: An Update for the Objective Scoring System. *Polygraph*, 31, 298-302.
- Krapohl, D. (2006). Validated Polygraph Techniques. *Polygraph*, 35(3), 149-155.
- Krapohl, D. & Handler, M. (2006). New Ratios for Scoring Electrodermal Responses. *Polygraph*, 35(2), 65-70.
- Krapohl, D. & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, 28, 209-222.
- Krapohl, D., Stern, B. & Ryan, A. (2003). Exclusionary vs. nonexclusionary: A review of the evidence. *Polygraph*, 32 (4), 245-250.
- Matte, J. (2000). A critical analysis of Hont's study: The discussion (stimulation) of Comparison Questions. *Polygraph*, 29(2), 146-149.
- Nelson, R., Handler, M. & Krapohl, D. (2008). Brute-Force Comparison: A Monte Carlo Study of the Objective Scoring System version 3 (OSS-3) and Human Polygraph Scorers. *Polygraph*, 37(1), 185-215.
- Offe, H. & Offe, S. (2007). The comparison question test: does it work and if so how? *Law and Human Behavior*, 31, 291-303.

- Palmatier, J. J. (1991). Analysis of two variations of control question polygraph testing utilizing exclusive and nonexclusive controls. *Masters Abstracts*, 30, 380.
- Podlesny, J. A. & Raskin, D.C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, 15, 344-58.
- Raskin, D.C. & Honts, C.R. (2002). The comparison question test. In M. Kleiner (Ed.), *Handbook of polygraph testing*. London: Academic (1-49).
- Raskin, D.C. and Kircher, J.C. (1990). Development of a Computerized Polygraph System and Physiological Measures for Detection of Deception and Countermeasures: A Pilot Study. A preliminary report under contract 88-L655300-000, Scientific Assessment Technologies, Inc. Salt Lake City, Utah.
- Raskin, D.C., Kircher, J.C., Honts, C.R. and Horowitz, S.W. (1988) A Study of Validity of Polygraph Examinations in Criminal Investigation, Grant number 85-IJ-CX-0040. Salt Lake City: Department of Psychology, University of Utah.
- Reed, S. (1994). Briefing on the Test of Espionage and Sabotage (TES), Expansion and Validation Study: Retrieved on 03/13/2008 from: <http://antipolygraph.org/documents/tesexpansion-validation.shtml>.
- Senter, S., Dollins, A. & Krapohl, D. (2004). A Comparison of Polygraph Data Evaluation Conventions Used at the University of Utah and the Department of Defense Polygraph Institute. *Polygraph*, 33, 214-222.
- Swinford, J. (1999). Manually scoring polygraph charts utilizing the seven-position numerical analysis scale at the Department of Defense Polygraph Institute. *Polygraph*, 28(1), 10-27.
- Timm, H. W. (1982) Analyzing deception from respiration patterns. *Journal of Police Science and Administration*, 10(1), 47-51.
- Weaver, R.S. (1980) The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. *Polygraph*, 9, 94-108.