

## **A Comparison of the Backster Scoring and Traditional Federal Scoring Using Computerized Analysis**

**Keith Hedges, George Deitchman and Dharam Samra**

### **Abstract**

Reid's introduction of a revised technique in lie detection tests was essentially the birth of the comparison question and set the stage for the semi-objective scoring of polygraph charts. Backster was the first to apply a positive and negative scoring system comparing the relevant questions against the comparison questions. Backster's numerical scoring technique has been modified to score Federal and Utah polygraph examinations and almost certainly by individual examiners and unpublished formats taught at little known polygraph schools. The immediate question that comes to mind concerns why modifications to Backster's techniques were thought necessary. The Federal You-Phase always scores relevant question tracings against the stronger bracketing comparison question tracings. The Backster You-Phase uses the Either-Or Rule (EOR) to select the bracketing comparison tracing to score the relevant question against. In essence the EOR permits scoring relevant question tracings against a stronger bracketing comparison tracing only when the relevant tracing reaction is weaker than both bracketing comparisons. Unless this condition is met the comparison is usually scored against the weaker comparison tracing. The EOR leads many examiners to assume an increased probability of a false-positive result. Computer simulated semi-objective examiner scoring was used to compare the Federal You-Phase, Backster You-Phase, modified Federal ZCT and Utah PLC. Statistical analysis showed the Backster system did produce a tendency toward larger negative values and a dramatic imbalance in conclusive examination scores. However, there was no practical difference in the false positive rates of the scoring techniques tested.

### **Introduction**

There have been many contributions to the Psychophysiological Detection of Deception and three individuals stand out for their concepts or actions. Leonarde Keeler's (Keeler, 1983) early and widespread promotion of polygraphy and opening the Keeler Polygraph Institute greatly enhanced the early development of "lie detection" in criminal investigations. Reid's (1946) introduction of a revised questioning technique in "lie detection" tests was the birth of the comparison question. Abrams (1977) documents that Backster's semi-objective scoring of polygraph

charts produced an enhanced level of objectivity to chart tracing analysis.

All of these contributions have evolved. Keeler's analog polygraph became a computerized polygraph. Reid's "comparative response" question evolved to the control question (either time-inconclusive or time-barred from the relevant questions) and then to the comparison question (typically time-barred from the relevant questions). Backster's numerical scoring has been revised to accommodate and enhance the Federal Zone Comparison Test and the Utah Probable Lie Test.

---

Keith Hedges is an APA member, lecturer, and regular contributor to the journal *Polygraph*. He developed a computerized program to confirm examiner scoring in 1997 and marketed the program until selling it in 2005. He considers the APA David L. Motsinger Award as his greatest achievement.

George Deitchman, PhD., LMFT is a therapist residing in Florida. He has been treating sex offenders since the early 1990s. He became a polygraph examiner in 2009 and is a member of the APA.

Dharam Samra is a 34-year veteran of local and state law enforcement. He has conducted thousands of interviews and interrogations in a variety of criminal, administrative, and civil cases. He is a graduate of the Backster School of Lie Detection and an APA member.

The semi-objective numerical scoring of polygraph charts sets standards for comparing tracing differences. The electrodermal tracing is the simplest and unquestionably the most objective tracing analyzed. The vertical deflection always results in a ratio of the comparison and relevant questions. The Backster and Federal Schools teach that a greater than or equal to 2:1 ratio results in a score of one, a greater than or equal to 3:1 ratio results in a two and a greater than or equal to 4:1 ratio results in a three. A ratio less than 2:1 results in a zero. The value is negative if the relevant question reaction is greater than the comparison question reaction. Scoring the respiration and cardiograph tracings is not quite as straightforward. Shorter respiration line lengths are considered to be of greater significance than a longer line length. The cardiograph tracing is viewed in a manner similar to the electrodermal tracing. Greater deflection or rise and duration is given greater comparative significance. Both the respiration and cardiograph tracings have historically relied upon descriptive terms rather than measurement to define comparisons. A "subtly" greater difference might describe a value of one. An "obviously" greater difference might describe a value of two and "dramatically greater difference" could describe a value of three. The Federal School currently teaches a method of ratios and secondary rules scoring the cardiograph. Examiner numerical scoring is effective though semi-objective. Spend an hour with a room full of polygraph examiners scoring single relevant and comparison questions on a seven-position scale and with the exception of the electrodermal tracing the examiners rarely if ever all agree on a given score though they rarely make opposing scores (negative versus positive). However, provide a relevant question that may be scored against two comparison questions and the scores are nothing less than chaotic and classification agreement diminishes. One method of reducing the magnitude of the opposing differences is to score on a three-position (-1,0,1) rather than a seven-position scale. Fewer scoring options limit diversity and provide greater consistency among scorers.

Backster (2001) teaches a comparison question format known as the You-Phase. Using the abbreviations IR for irrelevant

questions, SR for Sacrifice Relevant questions, SY for Symptomatic questions, C for Comparison questions and R for Relevant questions, the You-Phase has the following sequence as taught by the former Department of Defense Polygraph Institute (DODPI), now called the National Center for Credibility Assessment (NCCA):

IR1 SR2 SY3 C4 R5 C6 R7 C8 SY9

The Backster You-Phase question sequence is slightly different, reversing the positions of SR2 and SY3. This is where the similarity ends abruptly as the two schools teach extremely different methods of chart analysis. Although both Backster and the Federal School semi-objectively score each of the two relevant questions in the above chart sequence (three charts are normally required in an examination) against the bracketing comparison questions, the two methods could not be more opposed when it comes to comparison question tracing selection. This is where the controversy begins. The Federal School teaches scoring relevant question tracings against the stronger bracketing comparison tracing, while Backster uses the Either-Or Rule (EOR) to select the comparison tracing for scoring. Under the EOR when the subject produces a relevant question tracing reaction *weaker* than both of the bracketing comparison tracings, the relevant question tracing is scored against the stronger adjacent comparison tracing. If the relevant question tracing is stronger than or equal to either bracketing comparison tracing it is scored against the weaker comparison question tracing. To examiners accustomed to scoring against the stronger available tracing, the EOR would seemingly tend to produce an excessive number of false positives, especially when the Federal Polygraph School and probably the majority of other polygraph schools teach scoring relevant question reactions against the stronger bracketing comparison question. Backster has an additional rule known as the Green Zone Abuse (GZA) rule that states if a comparison question reaction is four times the magnitude of a relevant question, it is an abuse not to compare the relevant question to the tracing meeting GZA standards.

Samra further simplifies the either-or rule as follows:

- 1) If the relevant question tracing in question (for example the electrodermal) is weaker than both of the bracketing comparison question tracings the relevant question is scored against the stronger bracketing comparison tracing.
- 2) If the relevant question tracing is equal to both of the bracketing comparison tracings the relevant tracing is scored against either of the comparison tracings.
- 3) If the relevant question tracing is greater than or equal to the smaller of the two bracketing comparison tracings and less than the greater of the comparison tracings the relevant question tracing is scored against the weaker comparison tracing.
- 4) If the relevant question tracing is equal to the greater comparison tracing and greater than the other bracketing comparison tracing the relevant tracing is scored against the weaker comparison tracing.
- 5) If the relevant tracing is greater than both of the bracketing comparison tracings the relevant tracing is scored against the weaker comparison tracing.
- 6) The respiration and cardiograph tracings meet GZA requirements if they justify a +3 score.
- 7) The conclusive values for the Backster three-chart You-Phase are +7 or greater for non-deceptive classifications and a -13 or less for a deceptive classification.

The Backster EOR can be statistically tested without scoring the examinations. A computer program can be used to count the number of times the relevant question tracings are weaker than the bracketing comparison tracings in an examination, justifying scoring against the stronger adjacent comparison tracing. If deceptive examinations are statistically less likely than non-deceptive examinations to score against the stronger comparison tracings then EOR scoring may be valid and comparable to other scoring techniques.

A method of testing the EOR is to count the number of times the EOR criteria

select the strongest comparison in deceptive and non-deceptive examinations. If the EOR is valid it should detect more instances of scoring against the strongest comparison question for non-deceptive examinations thus providing greater positive scores for the non-deceptive than deceptive examinations. This project will score the examinations with the following scoring methods modified to score the You-Phase on a three-position scale: Federal You-Phase, Federal Zone Comparison Test, Preceding Comparison and EOR. The EOR electrodermal tracing (EDT) will be scored with Backster ratios and again with an EDT "bigger is stronger" rule. The relevant questions will also be scored against the weakest bracketing comparison question on a three-position scale to illustrate a worst-case scenario. In this paper the EOR with GZA is the only method scored on a seven-position scale.

## Method

### Relevant / Comparison Question Scoring

The Federal You-Phase relevant and comparison sequence is restated as follows:

IR1 SR2 SY3 C4 R5 C6 R7 C8 SY9

All of the mentioned techniques mandate the comparison of the respiration, electrodermal and cardiograph polygraph tracings. Our computerized program scores the abdominal respiration, electrodermal and cardiograph tracings. The Federal Polygraph School You-Phase is the only technique requiring all positive spot scores for a truthful conclusion. In a previous paper Hedges and Deitchman (2012) displayed the potential consequences negative spot scores can have on classifying as deceptive an otherwise correct truthful classification. The Federal School scores the You-Phase by comparing each relevant question tracing to the stronger bracketing comparison question. This method results in 12 comparisons per chart where R5 = 6 and R7 = 6. This is arrived at by scoring two relevant questions (R5, R7) against a possibility of three comparison questions (R5 versus C4 or C6 and R7 versus C6 or C8) by three tracings across three charts for a total of 36 comparisons in a three-chart examination. The examination has conclusive boundaries of plus and minus four and requires positive spot-scores for a truthful classification. This

method of comparison classified by total examination score is expected to produce the most correct truthful classifications but has the potential of producing false negatives due to the truthful bias the large number of comparison options provide. In this case it is predicted the Federal Scoring protocol of using negative spot scores will reduce false negative classifications, which is absolutely necessary where false negatives represent the error type of greatest concern. The results of Federal You-Phase classifications will be calculated twice, once based upon total examination scores and additionally using the negative spot score criteria so the two scoring techniques may be compared.

The Federal Zone Comparison Test contains three relevant questions per chart and permits multiple comparisons to the first relevant question in each chart. This method of analysis applied to the Federal You-Phase with only two relevant questions will result in nine comparisons per chart,  $R5 = 6$ ,  $R7 = 3$ , a total of 27 comparisons per three-chart examination. It is expected this method will produce less correct truthful classifications and more correct deceptive classifications than the Federal You-Phase method of analysis based on the total examination score.

The Utah Probable Lie Test scores relevant questions to the preceding comparison question. This method applied to the You-Phase will result in six comparisons per chart ( $R5 = 3$ ,  $R7 = 3$ ) or 18 tracing comparisons in a three chart examination. It is anticipated the Utah style of scoring will produce fewer false-negatives than either of the Federal methods.

The following steps must be taken to attain the noted research objectives:

1. Create measurements containing visually perceptible differences for the respiration, electrodermal and cardiograph recordings/tracings and a computer program to extract these measurements from computerized polygraph examinations.
2. Create a computer program to score the examinations according to each of the mentioned scoring techniques.
3. Test the program and various comparison methods on Federal You-Phase examinations for accuracy based on the

total score of the relevant questions with the exception of the Federal You-Phase scoring method.

### **Computerized Data and Scoring**

The data extraction program was written in 2007 by Fred Vater (deceased). The following criteria for each measurement were used by Vater during program creation.

Timm's (1981) respiration line length measurement was subjected to the abdominal respiration tracing from question onset through 15 seconds. If less than 15 seconds was available for measurement the measurement was discarded. The shorter line length is deemed the stronger of C and R lengths compared. The respiration tracings for all questions in a chart were summed and each question measurement was divided by the sum to produce a percentage for each question. This approach would hypothetically produce an average value of 11.11 ( $100 / 9 = 11.11$ ) if there were no differences in the measured lengths. Since shorter values are deemed the strongest, each calculation of respiration measurements was Relevant - Correctly Selected Comparison = Difference. If the difference was  $\geq 0.5$  or  $\leq -0.5$  it was operationally defined as visually significant since  $(0.5 / 11.11) = .045$  times 100 = 4.5%. The value was rounded to 5% for this project. Differences  $> -0.5$  and  $< 0.5$  were assigned a score of zero. Differences  $\geq 0.5$  or  $\leq -0.5$  were scored as a  $\pm 1$  respectively, differences  $> 1.83$  or  $< -1.83$  were scored  $\pm 2$  respectively and differences  $> 3.08$  or  $< -3.08$  were scored  $\pm 3$  respectively. The values 0.5, 1.83 and 3.08 were selected with error comparisons deleted to give approximately 17% assigned zeros, 60% assigned ones, 30% assigned twos and 10% assigned threes for non-deceptive comparisons.

The electrodermal tracing degree of reaction was measured from a baseline established at the point of tracing rise to the highest point achieved by the reaction. If an electrodermal deflection was less than one chart division the measurement was set at one chart division. The electrodermal tracing was considered eligible for measurement provided the rise began between question onset and six seconds following question end and there was a minimum of 15 seconds between question onset for the question being measured and the

question onset of the following question. The EOR was tested three times, once following the Backster ratio requirement of 2:1 to score a  $\pm 1$ , a second time with a "bigger is stronger" rule for a  $\pm 1$ , and a third time using both the EOR and GZA rules on a seven-position scale. The bigger is stronger rule was used for all other scoring techniques.

The cardiograph area under the curve was measured between the diastolic tips and a baseline established at the point of rise if the tracing rise began between question onset and not later than six seconds following question end. In Vater's data extraction program of 2007, if the diastolic tips returned to or fell below a previously established baseline and recovered in less than 1.97 seconds, a new baseline was established at the recovery point and the additional area(s) were summed for a total area below the diastolic tips. Measurement of any question ended 15 seconds after question onset. In the event a rise in the cardiograph tracing surpassed 15 seconds, measurement ended at 15 seconds, a vertical line was made to baseline and the area was calculated. The cardiograph tracing for all questions in a chart were summed and each question measurement was divided by the sum to produce a percentage for each question. This approach would hypothetically produce an average value of 11.11 ( $100 / 9 = 11.11$ ) if there were no difference in the measured areas. Since greater values are deemed the strongest each cardiograph calculation was (Selected Comparison - Relevant = Difference). If the difference was  $\geq 5.84$  or  $\leq -5.84$  it was operationally defined as visually significant. A difference of  $< 5.84$  and  $> -5.84$  was scored zero. A difference  $\geq 5.84$  or  $\leq -5.84$  was scored  $\pm 1$ , a difference of  $> 15.46$  or  $< -15.46$  was scored a  $\pm 2$  respectively and a difference  $> 26.24$  or  $< -26.24$  was scored a  $\pm 3$  respectively. The scoring values were set to meet the same percentage of positive scores as the respiration tracing.

#### **Data Selection Criteria**

All examinations from the CD labeled "Department of Defense Polygraph Institute, Psychophysiological Detection of Deception Field Data, Accurate: 03May02" were extracted from all folders and sorted to deceptive and non-deceptive folders. Vater's measurement extraction program produced

the data to analyze 43 deceptive and 25 non-deceptive Federal You-Phase examinations.

Examinations were restricted to those with a comparison and relevant sequence (excluding all other questions) of CRCRC. Relevant question labels were R5 and R7, the first comparison question was labeled "C4" and there was a minimum of two correctly sequenced charts in an examination.

#### **Inconclusive Zone**

The obvious method of forming an Inconclusive zone for a total examination score is to find the mean and standard deviations of the deceptive and non-deceptive groups and subtract 1.65 SD from the non-deceptive mean and add 1.65 SD to the deceptive mean. This will result in a 95% probability of the truthful or deceptive score falling in that range based on the normal distribution. Scores equal to or falling between the resulting two values are classified inconclusive. If a total examination score is greater than or less than the inconclusive boundaries one could state that a given examination total score has a .95 probability of not being in the opposite group of deceptive or non-deceptive determinations. The small number of non-deceptive examinations could produce a skewed view of classifications so a Jackknifed cross validation will also be used to compare scoring technique results.

### **Results**

#### **Stronger Comparison Selection Frequency**

Relevant question tracings were weaker than bracketing comparison question tracings thus qualifying for comparison to the stronger adjacent comparison tracing for non-deceptive examinations ( $M = 5.96$ ,  $SD = 2.89$ ) significantly more often than for the deceptive examinations ( $M = 2.07$ ,  $SD = 2.26$ ),  $t = -6.16$ ,  $p < .001$ . EOR scoring would not be feasible without this type of relationship.

#### **Scoring Techniques**

Multiple t-tests with a Bonferroni correction were used to test for statistically significant differences between deceptive and non-deceptive examination scores for the primary scoring methodologies. Statistical results must be less than 0.01 for each individual t-test to be significant. The degrees of freedom (df) for all tests were 66 and Table

1 contains the means and standard deviations for each scoring methodology tested. All results are based on the total examination score. The Federal ZCT scored on a three-position scale with a bigger is stronger EDT calculation produced a significant result ( $t = -8.53$ ,  $p < 0.001$ ). The preceding comparison (Utah Style) scored on a three-position scale with a bigger is stronger EDT calculation also produced a significant difference ( $t = -8.60$ ,  $p < 0.001$ ) as did the Backster EOR with the

GZA rule scored on a seven-position scale and 2:1, 3:1 and 4:1 EDT ratios ( $t = -7.73$ ,  $p < 0.001$ ). Finally, the Federal You-Phase style total examination scoring on a three-position scale with a bigger is stronger EDT calculations was also statistically significant ( $t = -8.44$ ,  $p < 0.001$ ). All of the t-tests produced statistically significant differences between deceptive and non-deceptive examinations for the above scoring methodologies.

**Table 1. Deceptive and non-deceptive means and standard deviations listed by scoring method.**

Scoring Method	Deceptive		Non-deceptive	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
EOR 2:1	-14.60	8.75	0.56	5.56
EOR $\pm^*$	-10.51	4.87	-1.88	3.91
UTAH $\pm^*$	-6.60	4.50	2.44	3.56
FED YOU-PHASE $\pm^*$	-3.91	4.69	5.48	3.92
FED ZCT $\pm^*$	-5.33	4.53	3.72	3.60
WEAKER C $\pm^*$	-10.95	4.59	-3.12	4.02
EOR 2:1 & GZA 7 Pt	-8.88	4.87	-0.36	3.38

\* Electrodermal bigger is stronger

Table 2 lists the Discriminate Analysis Jackknifed cross validation for the listed scoring methodologies. A Jackknifed cross validation was used due to the small number of non-deceptive examinations (25). It should be noted that a Jackknifed classification may produce an overly optimistic view of the results. That being noted, the only two methods sanctioned by polygraph schools produced three false positive errors (Backster EOR & GZA) and four false positive errors

(Federal spot scores). There are no practical differences between the two techniques based on false positive errors.

Note the consistency of the three Backster EOR scoring methods. All three had a non-deceptive classification bias and a higher classification error rate for deceptive examinations than for truthful examinations. The EOR 2:1 & GZA seven-position, Utah, Federal You-Phase and Federal ZCT modified

**Table 2. Discriminate analysis jackknifed cross-validation by scoring methodology**

Scoring Method	Jackknifed Classification					
	Deceptive			Non-deceptive		
	<i>Correct</i>	<i>Errors</i>	<i>Total%</i>	<i>Correct</i>	<i>Errors</i>	<i>Total%</i>
WEAKER C*	35	8	81%	20	5	80%
EOR±*	35	8	81%	23	2	92%
EOR EDT 2:1	35	8	81%	22	3	88%
EOR 2:1 & GZA 7 Pt	33	10	77%	22	3	88%
UTAH*	34	9	79%	22	3	88%
FED You-Phase*	35	8	81%	22	3	88%
FED You-Phase**	38	5	88%	21	4	84%
FED ZCT*	37	6	86%	21	4	84%
Mean	35.3	7.8	81.1%	21.6	3.4	86.5%

\* Electrodermal bigger is stronger

\*\* Federal You-Phase using spot scores as predictor variables

scoring methods all produced the same number of Total Examination Score (12%) non-deceptive errors. The experimental Weaker Comparison scoring method produced the best classification balance with 81% deceptive versus 80% non-deceptive classifications. The average error rate for each scoring method is calculated by adding the number of deceptive errors plus the number of non-deceptive errors divided by the total number of examinations. The Weaker Comparison method produced 19% errors, EOR EDT bigger is better method 15% errors, EOR EDT 2:1 method 16% errors, EOR EDT 2:1 GZA scored on a seven-position scale method 19% errors, Utah previous comparison method 18% errors, Federal You-Phase total examination score method 16% errors, Federal You-Phase spot scores 13% errors and the Federal ZCT method 15% errors.

Jackknifed classification consistency between the Federal You-Phase total examination score and Backster EOR/GZA scoring techniques was calculated for deceptive and non-deceptive examinations. The techniques agreed on 31 correct deceptive examinations, 3 incorrect classifications and disagreed on 9 classifications. The techniques agreed on 19 correct non-deceptive classifications and disagreed on 6 classifications. The techniques agreed on  $(34 + 19) / 68 = (53 / 68) = 78\%$  of the examinations scored.

Recall the calculation for a 95% probability is Deceptive Method Mean + (SD \* 1.65) and Non-deceptive Method Mean - (SD \* 1.65). Scores falling between or equal the two values are classified inconclusive/no opinion. Table 3 displays the classification rate requiring a 95% probability of group classification.

**Table 3. Classification results by scoring method and .95 probability**

Scoring Method	Deceptive			Non-deceptive		
	<i>Correct</i>	<i>Errors</i>	<i>Inc</i>	<i>Correct</i>	<i>Errors</i>	<i>Inc</i>
WEAKER C*	31	4	8	15	1	9
EOR±*	32	5	6	16	1	8
EOR EDT 2:1	32	3	8	18	2	5
EOR 2:1 & GZA 7 Pt	33	3	7	14	1	10
UTAH*	31	2	10	18	2	5
FED YOU-PHASE*	32	2	9	18	3	4
FED ZCT*	31	1	11	19	1	5

\* Electrodermal bigger is stronger

Classification agreement of the Backster You-Phase EOR with GZA scored on a seven-position scale and the Federal You-Phase scored on a three-position scale (both classified by Total Examination Score and a 95% probability) was examined by calculating the number of classifications in agreement, the number of classifications opposing and the number of classifications where one method made a classification and the other method was inconclusive. There were 29 deceptive examinations in agreement and 14 deceptive examinations where one of the methods was inconclusive. There were 12 non-deceptive examinations with classification agreement, 2 non-deceptive examinations with classification disagreement and 11 examinations where one of the two methods was inconclusive. The results are a worst-case scenario as the Total Examination Score was incorrectly used to calculate the Federal

You-Phase results. Absolute opposite classifications occurred just two times ( $2 / 68$ ) = 3% disagreement. It must also be noted that the impact of an inconclusive classification by either the EOR or Federal You-Phase scoring was  $(14 + 11) / 68$  = 37%, obviously reducing the number of classification disagreements.

Table 4 contains the inconclusive values based on a 95% probability. Notice the balance or lack thereof between the deceptive and non-deceptive scores required for a decision. The non-deceptive numbers are all negative except for the Utah scoring method (barely) while the Federal ZCT and Federal You-Phase methods non-deceptive numbers are both further from zero (positive). The Federal ZCT method has the best inconclusive balance of all of the scores produced by a scoring method.



**Table 4. Inconclusive boundaries**

Scoring Method	Deceptive	Non-deceptive
WEAKER COMPARISON	< -9.75	> -3.38
EOR $\pm^*$	< -8.33	> -2.47
EOR EDT 2:1	< -8.61	> -.16
EOR 2:1 & GZA 7 Pt	< -5.94	> -.84
UTAH*	< -3.43	> .83
FED YOU-PHASE*	< -.99	> 3.13
FED ZCT*	< -2.22	> 2.14

\* Electrodermal bigger is stronger

## Discussion

The Either-Or Occurrences t-test and the results displayed in Tables 2 and 3 all refute the assumption that the EOR is biased against non-deceptive examinees. Actually the exact opposite of this belief is displayed. All three methods of calculating the EOR result in a non-deceptive bias rather than deceptive. The three methods of scoring the EOR produced non-deceptive classifications averaging from Table 2,  $(92 + 88 + 88) = (268 / 3) = 89.3\%$ , while the deceptive classifications averaged  $(81 + 81 + 77) = 79.6\%$  which is exactly the opposite of what would be expected of a scoring technique with a false-positive bias. The number of non-deceptive cases is very small (25) but the authors must conclude the results are correct based upon other methods analyzed. One would expect the Federal You-Phase scoring methodology using the Total Examination Score to produce greater non-deceptive accuracy than deceptive, and that is exactly what happened. There were (Table 2) 88% correctly classified non-deceptive and 81% correctly classified deceptive cases. To base Federal You-Phase results upon the total examination score would be an error for this type of scoring as there is just too much

cherry picking comparing each relevant question against two bracketing comparison questions. The classification results using spot scores was 84% correct non-deceptive and 88% correct deceptive classifications. It is the authors' joint opinion that in law enforcement investigations the last result desired is a false negative. The Federal ZCT total examination score method utilizes less cherry picking than the Federal You-Phase scoring, producing 86% correct deceptive classifications and 84% correct non-deceptive classifications. However, close inspection of Tables 1 and 2 reveals no practical difference in the accuracy rates of the methods used.

It would be a mistake to form global conclusions about the Weaker Comparison, Utah and Federal ZCT accuracy rates as applied to the You-Phase. The Utah and Federal ZCT scoring techniques were developed for examination formats other than the You-Phase and were calculated in this project to demonstrate they all have the same basic accuracy. The Jackknifed classification for the Federal ZCT method error rate was  $(10/68) = 15\%$  and the Utah method error rate was  $(12/68) = 18\%$ . Compare these to the EOR on a seven-position scale error rate  $(13/68) = 19\%$ . Examination of Backster's

conclusive scores of +7 or greater for non-deceptive and -13 or less for deceptive classifications and the EOR means listed in Table 1 and Inconclusive Boundaries in Table 4 suggests and supports a negative scoring bias for the EOR. Note the Table 1 non-deceptive means are all very close to or less than zero, and almost double the non-deceptive examiner score requirement of +7 required for a deceptive classification of -13. This suggests a large number of inconclusive results. The seven-position scale (Table 3) produced  $(7 / 43) = 16\%$  deceptive inconclusives and  $(10 / 25) = 40\%$  non-deceptive inconclusives. Hedges acknowledges the proportions of positive and negative scores

assigned may not reflect the true state of nature and could be responsible for the excessive number of inconclusives in this project. A normal distribution was assumed but the data are negatively skewed. Given the consistency of projected expectations and confirmation by the results we are confident the EOR's negative scoring bias does not cause a false positive bias due to the results of our calculated inconclusive zone. Backster's deceptive requirement of -13 for a numerically scored conclusion of deception further confirms our findings. The authors agree or concede the Backster EOR You-Phase did not show a deceptive bias.

## References

- Abrams, S. (1977). *A Polygraph Handbook for Attorneys*. Lexington, Massachusetts: Lexington Books, D.C. Heath and Company.
- Backster, C. (2001). Backster Zone Comparison Test Variations. Backster School of Lie Detection.
- Hedges, K. and Deitchman, G. (2012). Does spot scoring and relevant and comparison question order help or hurt the examinee? A computer analysis of ground truth verified Army and Air Force MGQT and Federal ZCT Exams. *Polygraph*, 41(3), 156-169.
- Samra, D. (2005, PE-168). Standardized Polygraph Note Pack and Technique Guide, 1979 Edition. Backster School of Lie Detection, San Diego, CA.
- Keeler, E. (1983). *The lie detector man: The career and cases of Leonarde Keeler*. Telshare Publishing: West Palm Beach, Florida.
- Reid, J.E. (1946). A revised questioning technique in lie detection tests. *Journal of Criminal Law and Criminology*, 37(6), 542-547.
- Timm, H.W. (1982). Analyzing deception from respiration patterns. *Journal of Police Science & Administration*, 10, 47-51.