

VOLUME 37

Г

2008

NUMBER 4

Contents	
An Assessment of the Backster "Either-Or" Rule in Polygraph Scoring	240
Eldad Meiron, Donald Krapohl & Tzachi Ashkenazi	
A Comparison of the Objective Scoring System and Probability	250
Andrea K. Webb, Mark D. Handler, Donald J. Krapohl & John C. Kircher	
Respiratory Blood Pressure Fluctuations Observed During Polygraph Examinations Mark D. Handler & Joel M. Reicherter	256
Accuracy Demonstration of the Horizontal Scoring System Using Field Cases Conducted with the Federal Zone Comparison Technique	263
Donald Maponi, Nathan Gordon & Christopher Lombardi	
An Introduction to the APA's Panel on International Developments in Polygraphy Frank Horvath	269
History of Polygraphy in Venezuela Esteban Balthazar	271
Polygraph Research: Value, Shortcomings, and Perspectives Stuart M. Senter	277

An Assessment of the Backster "Either-Or" Rule in Polygraph Scoring

Eldad Meiron¹, Donald J. Krapohl² and Tzachi Ashkenazi³

Abstract

We investigated the Backster "Either-Or" Rule (EOR) using 100 Backster field cases which were blind scored by field examiners who used either the Backster scoring system or the US federal scoring system. The Backster scorers indicated those scores which were the product of the EOR. Without the EOR the Backster scorers made an average of 70% correct decisions, 3% errors, and 27% inconclusive results. With the EOR they rendered an average of 80% correct decisions, 3% errors, and 17% inconclusives. All of the increase in decision accuracy and reduction in inconclusive rate was attributable to improvement in classifying deceptive cases. There was no effect on truthful cases. The scores from the US federal system produced an average of 86% correct decisions, 5% errors, and 9% inconclusives. Decision accuracy, inconclusives and errors between the US federal and the Backster (with EOR) were not significantly different, though both performed better than the Backster method without EOR. The implications and limitations of these findings are discussed.

Among first scoring systems for the Backster's 7-position polygraph, scoring system (Backster, 1963) was to garner widespread acceptance and became the dominant approach to chart interpretation for the polygraph profession. It is the system from which virtually all other manual scoring systems are derived. Portions have been updated over the years, most notably changes in the decision rules in 1983 (Weaver, 1985) however much of the system has remained constant since 1963.

Despite the prominence of 7-position scoring, scientific investigation of the Backster methods are difficult to find. We located a single paper which reported the results of two laboratory experiments which employed the entire Backster method, both testing and scoring approaches (Honts, Hodes & Raskin, 1985). Combining the non-countermeasure data from the two studies, the Backster

system with 31 deceptive cases produced 90% correct calls, no errors and had 10% inconclusives. For the 31 truthful cases, the system made 45% correct calls, 32% error and 23% inconclusive. The overall inconclusive rate was 16%. The average decision accuracy for truthful and deceptive cases was 79% when inconclusives were excluded. Because these were laboratory studies, the very high accuracy with deceptive cases was not Some writers have expressed expected. reservations about laboratory studies because examinees in those studies have little to fear about getting caught (Iacono & Patrick, 1987). Nevertheless, without inconclusives decision accuracy for deceptive cases in the Honts et al study was 100%. What may be more inexplicable was the disappointing accuracy of truthful cases, especially given the relatively low jeopardy laboratory studies typically engender. Even excluding the inconclusive cases which accounted for nearly a third of

¹ Eldad Meiron is the Director of the Israeli Government Polygraph School.

² Donald J. Krapohl is an APA Past President (2006-2007) and a regular contributor to this publication.

³ Dr. Ashkenazi is a Research Associate of the Israeli Government Polygraph School.

We appreciate the assistance of the volunteer scorers: Johnnie Rodgerson, Jimmie Swinford, Esther Harwell, and the Israeli scorers in this study The views expressed are the authors', and do not necessarily represent those of the American Polygraph Association, US Government, US Department of Defense, or the Israeli government. Requests for reprints should be sent to Eldad Meiron at igpolyschool@012.net.il.

the truthful sample, accuracy was still only 58%. Taken together, these data show a promising sensitivity for the Backster method, but a deficient specificity. Absent any other published validity studies of the Backster system, questions regarding the generalizeability of the Honts et al findings cannot be answered.

Our present interest is turned toward the Either-Or Rule (EOR), a central tenet of the Backster scoring system and one for which we have not located any previously Of the many published investigations. variants of the 7-postion scoring system, Backster's is noted for its complexity (Weaver, 1980). It has 21 scoring rules divided into four categories: Primary, Secondary, Upgrading, and Tracing Oddity Rules (Backster, 1990). The first listed among the Primary Rules is the EOR, a rule not found beyond the followers of the Backster method but one staunchly defended by advocates in the field. It reads as follows (Backster, 1990):

"Either-Or" Rule

To arrive at an interim spot analysis tracing determination of (+2) or (-2)there must be a significant and timely tracing reaction in either the red zone or green zone being compared. Note: If the red zone indicates a lack-of-reaction it should be compared with the neighboring green zone containing the larger timely reaction. If the red zone indicates a timely and significant reaction it should be compared with the neighboring green zone containing no reaction or the least reaction. Rule Classification: Primary (and most important). Tracings Included: Breathing, G.S.R. and Cardio. (Authors' comment: Underline in the original).

For those unfamiliar with the Backster system, some explanation of the language might be helpful. "Red zones" take place during the testing phase of the examination, and mark the presentation of the relevant question. Similarly, "green zones" denote the period of testing at the presentation of the probable-lie comparison (PLC) questions. A "lack-of-reaction," as the name suggests, speaks to an absence of a phasic response after either a relevant or PLC question is presented during the test. In sum, Backster promotes a contingency rule based on reactivity to the relevant question. More plainly, if there is no significant reaction to a relevant question, that tracing is compared to the stronger of the two phasic responses to the adjacent PLCs. Conversely, if a relevant question evokes a significant and timely physiological response, that response is compared to the weaker of the two responses elicited bv the two adjacent PLCs. Significance of a response is judged by globally assessing the examinee's entire polygraph chart to gauge his or her reaction capability. Therefore, the significance is assessed within the context of the examinee's own characteristic reaction pattern rather than by a universal metric. As noted earlier, Backster considers the EOR "most important."

In order to determine the value of the EOR we engaged scorers who routinely use the Backster system in the blind interpretation of field cases for which ground truth had been established. We also recruited scorers who use the US federal scoring system to blind score the same cases. Our approach was designed to answer two questions:

1. Did the EOR improve the decision accuracy of those who used the Backster system?

2. Does the Backster system, with its use of the EOR, have a performance advantage over the more generally used US federal scoring system?

Methodology

Cases

One hundred cases were drawn from the archive of verified field cases collected by the Israeli government. Half were confirmed as deceptive cases and the other half were truthful. The 50 truthful cases comprised the entire archive of confirmed truthful examinees who were suspects at their time of testing since 2004. The confirmed conducted deceptive cases were randomly drawn from an archive of all confirmed deceptive cases conducted since 2004. A11 had been conducted using the Backster "You Phase" (single-issue format) polygraph technique (Matte, 1996). The charts were collected on Lafayette 4000 computer polygraphs. The number of charts varied by case: 19 cases had only 2 charts, 24 had 3 charts, and 57 had 4

charts. All identifying information was stripped from the charts, and they were produced in paper copy for the examiners to score.

Scorers

Three Israeli government polygraph examiners represented the Backster scorers. Similarly, three US federal scorers were also recruited to score the cases. All scorers were very experienced with the scoring methods they used in this study. The Israeli scorers were trained at the Israeli Government Polygraph School and were compliant with continuing education requirements. Thev learned the Backster method during in-house continuing education lectures from senior Israeli polygraph examiners who had attended the weeklong seminar held each December at the Backster School of Lie Detection. The US federal scorers had completed the 14-week US government polygraph school, and had met federal continuing education requirements of 80 hours each two years. Neither the Israeli nor the US government scorers were aware of the base rates, case facts, ground truth or test questions of the cases used.

Backster Scoring Method

In the Backster scoring system, examiners assign 7-position scores by comparing reactions to the relevant question to the stronger of the two adjacent probablelie comparison questions (each channel separately) except when the EOR is applied while the examiner evaluates a specific reaction in the relevant question as "significant and timely" response. This method is done channel by channel for the four data channels. For summing purposes in the pneumograph, the average of the two channels is calculated. All scores are summed by each chart (cumulative scores), and summed overall (total charts). Decision rules for a call of No Deception Indicated (NDI) the overall sum for two charts should be +5 or greater, three charts +7 or greater, and for four charts +9 or greater. Decision rules for Deception Indicated (DI) are that the overall sum for two charts should be -9 or greater, three charts -13, and for four charts -17. If the overall sum falls short of the threshold, the rule of "The two most productive charts" is applied in order to reduce the inconclusive

rate. All other scores result in a No Opinion (Inconclusive) result. The decision rules for the Backster system were based on two to four charts.

Federal Scoring Method

The federal system assigns scores to tracing features reported by Bell, Raskin, Honts and Kircher (1999) known as the Utah scoring system. In the federal system, examiners assign 7-position scores bv comparing reactions to the relevant question to the stronger of the two adjacent probablelie comparison questions. This method is done channel by channel for the four data channels. For summing purposes in the pneumograph, only the score furthest from 0 is used, unless the two pneumograph scores are on opposite sides of 0, in which case a 0 is used in the tabulation. All scores are summed by relevant question (spot scores), and summed overall (total scores). For a decision of No Deception Indicated (NDI) all spot scores must be greater than 0 and the total score must be +4 or greater for the You-Phase format with two relevant questions. If any spot score is -3 or lower or the total score is -4 or lower, the decision is Deception Indicated (DI). All other scores result in a No Opinion (Inconclusive) result.

Results

Table 1 shows the percentage of correct decisions for each one of the polygraph examiners at each one of the scoring methods. The percentage is calculated relative to all decisions (including erroneous and inconclusive). Table 2 shows the average percentage of correct decisions across the three examiners at each one of the scoring methods.

Both the overall percentages of correct decisions and the percentages of correct decisions within the truthful cases were significantly higher than 50% (chance percentage) for all examiners in all the scoring methods. However, in the deceptive cases, all three examiners using Backster (no EOR) did not reach a percentage of correct decisions that was significantly higher than 50%, while all other examiners using Backster (EOR) or federal scoring methods did so.

	Deceptive Cases	Truthful Cases	
	<u>(n=50)</u>	<u>(n=50)</u>	<u>Overall</u>
Backster (no EOR) examiner A	50%	96%	73%
Backster (no EOR) examiner B	58%	78%	68%
Backster (no EOR) examiner C	56%	84%	70%
Backster (EOR) examiner A	80%	92%	86%
Backster (EOR) examiner B	80%	76%	78%
Backster (EOR) examiner C	72%	82%	77%
Federal examiner X	82%	100%	91%
Federal examiner Y	82%	68%	75%
Federal examiner Z	90%	92%	91%

Table 1.	Percentages of correct decisions for each one of the polygrap	h
	examiners at each one of the scoring methods.	

Table 2. Average percentage of correct decisions (across the three examiners)at each one of the scoring methods

	Deceptive <u>Cases (n=50)</u>	Truthful <u>Cases (n=50)</u>	Correct <u>Overall</u>	Error <u>Overall</u>	Inc <u>Overall</u>
Backster (no EOR)	55%	86%	70%	3%	27%
Backster (EOR)	77%	83%	80%	3%	17%
Federal	85%	87%	86%	5%	9%

For the statistical significance analysis one can not use the average percentage of correct decisions (across the three examiners). Therefore, we computed another variable for each one of the 100 cases. We calculated the number of examiners who decided correctly at each of the scoring methods. Since there were three examiners at each scoring method, that variable could have one of the following four values: 0, 1, 2, 3 ranging from a case in which none of them were correct, to the case of all three examiners decided correctly.

Figure 1 shows the distribution of the number of examiners who decided correctly at each of the scoring methods.



Figure 1. The distribution of the number of examiners who decided correctly at each of the scoring methods

It can be seen at the above figure that the percentage of cases in which all three examiners correctly decided was lower in the Backster (no EOR) method (59%), and higher both in the Backster (EOR) (70%) and the Federal method (73%) who achieved similar results. It can also be seen that the Backster (no EOR) method had a relatively high percentage of cases in which none of the three examiners correctly decided (16%).

Significance analysis was conducted using either Sign Test or the Binomial distribution (in cases of small N or many ties). The Sign Test and the Binomial distribution analysis in this paper is based on comparing the number of cases in which there are more scorers who decide correctly in method A than in method B, to the number of cases in which there are more scorers who decide correctly in method B than in method A (H_0 : they are Analysis showed the following: (1) equal). Backster (EOR) scoring method is significantly more accurate than Backster (no EOR) method (p<0.05 in Binomial distribution); (2) The Federal scoring method is significantly

more accurate than Backster (no EOR) method (z=3.35 p<0.05 in Sign Test) and; (3) The Federal scoring method is not significantly different in accuracy from the Backster (EOR) method (z=1.28 p>0.05 in Sign Test).

In order to see whether the differences between the scoring methods are consistent in the deceptive cases as well as in the truthful ones, the above analysis was conducted separately for each of the two types. Figure 2 shows the distribution of the number of examiners who decided correctly at each of the scoring methods in the deceptive cases.

Focusing only in the deceptive cases one can see larger differences between the scoring methods. In contrast to a relatively lower percent of cases in which all three examiners decided correctly using Backster (no EOR) method (42%), one can see a relatively high percent of cases in which all three examiners decided correctly using Backster (EOR) method (68%) and the Federal method (78%) who are still similarly accurate. It can also be seen that the Backster (no EOR)



Figure 2. The distribution of the number of examiners who decided correctly at each of the scoring methods in the deceptive cases

method has relatively high percentage of cases in which none of the three examiners correctly decided (30%) compared to much lower percents (and relatively similar) in both other methods (14% and 10%).

Here also, significance analysis was conducted using either Sign Test or the Binomial distribution (in cases of small N or many ties). Analysis showed the following: (1) Backster (EOR) scoring method is significantly more accurate than Backster (no EOR) method (p<0.05 in Binomial distribution); (2) The Federal scoring method is significantly more accurate than Backster (no EOR) method (p<0.05 in Binomial distribution), and; (3) The Federal scoring method is not significantly different in accuracy than Backster (EOR) method (p>0.05 in Binomial distribution).

Figure 3 shows the distribution of the number of examiners who decided correctly at each of the scoring methods in the truthful cases.

When we focus only in the truthful cases we see much smaller differences between the scoring methods. The percentage of cases in which all three examiners decided correctly is similar in all three methods (it ranges between 68% and 76%). The percentage of cases in which none of the three examiners correctly decided is also similar across scoring methods (it ranges between 0% and 6%). The significance analysis using either Sign Test or the Binomial distribution showed no difference in the accuracy levels of the three methods.

Summary of the results: It appears from the above analysis that the Backster (no EOR) scoring method has the lowest accuracy levels (not different than chance) only among the deceptive cases, while both the Backster (EOR) method and the Federal method have high (and above chance) accuracy levels in general (both in the deceptive and in the truthful cases). This would permit the conclusion that both the Backster (EOR) method and the Federal method have higher



Figure 3. The distribution of the number of examiners who decided correctly at each of the scoring methods in the truthful cases

accuracy levels than the Backster (no EOR) scoring method with the deceptive cases. For the truthful cases there are no differences between the accuracy levels among all three methods. Consequentially, it was found that both the Backster (EOR) method and the Federal method have higher accuracy levels than the Backster (no EOR) scoring method also in the overall analysis (deceptive + truthful cases).

Inter-scorer agreement for the EOR

The use of the EOR entails an additional subjective step in the assessment of the physiological data, and the question as to whether this step would affect inter-scorer agreement for the application of the EOR was unknown. We tracked those scores identified by the Backster scorers as being where the EOR had been used, and compared the use of the EOR among the three Backster scorers. Scorer 1 used the EOR 156 times, Scorer 2 a total of 157 times, and Scorer 3 a total of 99 times. In tracking the individual scores affected by the EOR, we found that when one scorer used the EOR to assign a score the

likelihood of a second scorer also using the EOR at the same location averaged 32.9%, with a range of 25.3% to 40.4%. In other words, on average when a scorer invoked the EOR to assign a score, a second independent scorer used the EOR about a third of the time. Because these scorers were part of a government program with standardized training and quality control oversight, it is unknown whether the observed agreement among scorers in the use of the EOR will generalize to settings that did not have similar unifying conditions.

Discussion

The present study had two objectives: to assess the effect of the EOR within the Backster scoring system, and to compare the relative decision accuracy between the Backster and Federal scoring systems. With regard to the first question, the EOR did result in a significant increase in the correct classification of deceptive cases for the Backster scorers with no effect in the detection of truthful cases. This is an interesting finding, and warrants further investigation. If a replication of this study also finds that all of the benefit for the EOR is in the detection of deceptive cases, it might call for further refinements of the EOR. It may also suggest that the same effect may be achievable by simply adjusting the cutoff scores.

In that vein, we conducted a post hoc analysis of the present data to determine whether the accuracy found with the Backster EOR rule could be produced more parsimoniously with optimized cutoff scores. In an exploratory analysis we attempted to match the accuracy found with the Backster system with EOR by fixed cutoff scores to the data when the EOR had not been used. By considering the mean and variability of the scores of the truthful cases and deceptive cases of the Backster scores without EOR, we determined that similar accuracy could be achieved by using fixed cutoffs of -3/+6. (By fixed cutoffs we mean that the cutoffs remain the same regardless of the number of charts in a case.) Table 3 displays how these two approaches compare. There were no statistically significant differences between the Backster system with EOR using the Backster escalating cutoffs and the Backster system without EOR with fixed cutoffs. These findings suggest that the degree of complexity which attends the EOR and escalating cutoffs is not repaid with improved accuracy.

Table 3. Average percentage of correct, erroneous and inconclusive (Inc) decisions for deceptive and truthful cases for the traditional Backster system (including EOR and escalating cutoffs) and the same scores that employed fixed cutoffs and no EOR.

	<u>Deceptive</u>	<u>Truthful</u>	Correct <u>Overall</u>	Error <u>Overall</u>	Inc <u>Overall</u>
Backster Traditional	77%	83%	80%	3%	17%
Backster with fixed cutoffs and no EOR	83%	87%	85%	4%	11%

The finding that fixed cutting scores may deliver accuracy competitive with the standard Backster system, combined with evidence that scorers did not agree on when to apply the EOR most of the time, provide two challenges for the EOR. In balance, the EOR did improve decision accuracy for Backster scorers over not using the EOR, but we also observe that adjusted cutoffs were shown to be just as effective in boosting accuracy. It is important to make clear that our data do not argue that the traditional Backster scoring methods is ineffective, but merely that the addition of the EOR appears to send a scorer on the longer of two paths going to the same destination. If the value of the EOR over the non-EOR is confirmed in other research, examiners who prefer the EOR can continue to use it in good conscience. Examiners who favor simplicity in scoring can eschew the EOR and opt instead for fixed cutoffs. The cutoff scores used here, however, should not be adopted in the field for reasons articulated later under Limitations.

On the second research question, that of the relative accuracy of the Backster and federal scoring systems, we found no statistical difference. Without inconclusives Backster scorers were correct in 96% of their decisions (16% inconclusive) and the federal scorers were correct in 94% of their decisions (9% inconclusive). Polygraph examiners competent in the scoring of charts using the federal method did as well as those polygraph examiners competent in the Backster scoring system. The close similarity in the accuracy of the Backster and federal scorers bodes well for practitioners from both schools. If these same results are confirmed elsewhere, an established performance equivalency may allow field users to set aside professional disagreements regarding the superiority of their preferred methods. In terms of accuracy, errors and inconclusives, our research did not uncover any meaningful differences between two systems. the Arguably, given the equal accuracy the only real difference may be the degree of complexity: the federal system is simpler to learn and apply.

Limitations

1. The present study is only the first to examine the effect of the EOR and to directly compare decision accuracy between the federal and Backster systems. Consequently the findings should be considered tentative until more studies are published.

2. One of the criticisms of field studies is the strong potential for selection bias. In short, it means that confirmed field cases tend to be those in which the original polygraph results were right as opposed to those occasions when the polygraph decisions were wrong. Using only confirmed cases in one's analysis, therefore, might make a technique appear more accurate than it really is. Because the sample in this study did not include any cases where the field polygraph decisions had been incorrect or inconclusives, critics may point to a selection bias that could artificially inflate the accuracy estimate of the testing technique, including the blind scoring results. This argument may find support in our data: The fixed cutoffs that matched the accuracy for the traditional Backster system had an average score for truthful cases that were further from 0 than did deceptive cases. In the context of field scoring research these findings are anomalous. Indeed, the Backster scoring system itself has cutoffs that are further from 0 for DI decisions than the cutoffs of NDI decisions, which is in concert with most research. The possibility of nonrepresentative sampling here is acknowledged. That said, because our interest was the relative accuracy of three approaches to scoring rather than the absolute accuracy of the Backster technique, and because a sampling bias would not advantage а particular scoring method, we are more confident that our findings in that regard will replicate.

References

- Backster, C. (1963). Standardized polygraph notepack and technique guide: Backster zone comparison technique. Cleve Backster: New York.
- Backster, C. (1990, August 14). *Backster zone comparison technique: Chart analysis rules*. Paper presented at the 25th annual seminar of the American Polygraph Association, Louisville, KY.
- Bell, B.G., Raskin, D.C., Honts, C.R., and Kircher, J.C. (1999). The Utah numerical scoring system. *Polygraph*, <u>28</u>(1), 1-9.
- Honts, C.R., Hodes, R.L., and Raskin, D.C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology*, <u>70</u>(1), 177-187.
- Matte, J.A. (1996). Forensic psychophysiology using the polygraph. J.A.M. Publications: Williamsville, NY.
- Iacono, W.G., and Patrick, C.J. (1987). What psychologists should know about lie detection. In (Eds Hess & Weiner) *Handbook of forensic psychology*. pp 460-489. John Wiley: New York.
- Weaver, R.S. (1980). The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. *Polygraph*, <u>9(2)</u>, 94-108.
- Weaver, R.S. (1985). Effects of differing numerical chart evaluation systems on polygraph examination results. *Polygraph*, <u>14(1)</u>, 34-42.

A Comparison of the Objective Scoring System and Probability Analysis

Andrea K. Webb¹, Mark D. Handler², Donald J. Krapohl³, and John C. Kircher¹

Abstract

The Objective Scoring System (Krapohl & McManus, 1999) and probability analysis (Kircher & Raskin, 2002) are two methods for determining veracity from a polygraph examination. The purposes of the present project were to compare classification accuracies for the two methods and determine if the scoring rules for either method are biased. Both methods had high classification accuracies. The accuracies for OSS were slightly higher than for PA, but the difference was not statistically significant. Both methods are valid and scores from both can be easily obtained with a mouse click using available computer software.

Introduction

The Objective Scoring System (OSS) was developed by Krapohl and McManus (1999) as an objective method for scoring the physiological data obtained during a polygraph examination. Three physiological measures, respiration line length, skin conductance or resistance amplitude, and baseline increases in the cardiograph, are obtained for relevant and comparison questions. The measures are converted to ratios by taking the measurement for the relevant question and dividing it by the measurement for the comparison question. In the first version of OSS, the adjacent comparison question with the strongest response was used to compute the ratio for the first relevant question. For the second and third relevant questions, the preceding comparison question was used to compute the ratio. In the second version of OSS (Krapohl, 2002), the comparison question preceding each relevant question is used to compute the

ratio. These ratios are used to assign scores to each relevant question, which then are summed to make classifications about truthfulness or deception. Negative values are more indicative of deception, and positive values are more indicative of truthfulness. Decision thresholds can be adjusted according to an examiner's tolerance for errors and inconclusive decisions. Krapohl and McManus (1999) developed and validated their method with polygraph cases conducted using the Zone Comparison Technique (ZCT). They used the first three charts of the ZCT tests, each of which contained three relevant and three comparison questions.

Another method used to determine a subject's veracity is probability analysis (PA). This method uses the same physiological features as OSS, but instead of forming ratios of physiological measurements, it uses a discriminant function to determine the probability that a subject was truthful or deceptive (Kircher & Raskin, 2002). Different

³ Defense Academy for Credibility Assessment

Corresponding Author:	Andrea K. Webb
	University of Utah
	1705 E. Campus Center Dr. Rm 327
	Salt Lake City, UT 84112
	Email: Andrea.Webb@utah.edu
	Phone: (801) 585-0964 / Fax: (801) 581-5566

¹ University of Utah

² American Association of Police Polygraphists

cutoff values can be used to determine a person's veracity. In the present study, if the probability of truthfulness was .70 or larger, the subject was deemed truthful. If the probability of truthfulness was .30 or lower, the subject was deemed deceptive. A complete description of this method and justification for the decision cutoffs is given in Kircher and Raskin (2002). The CPS computer program (Version 4.20; Stoelting Company, Wood Dale, IL) can be used to score polygraph charts with both the OSS and PA methods. The development of CPS is described in Kircher and Raskin (2002).

Although both scoring methods have been used for many years, there has not been a study conducted to compare results from both methods. The purpose of the present project was to determine which method was more accurate. We compared OSS and probability analysis using a sample of 160 field cases to ascertain which method had higher accuracy and larger area under the receiver operating characteristic (ROC) curve.

Method

One hundred sixty cases from the database at the Defense Academy for Credibility Assessment were used. None of the cases used in the development of OSS or PA were used in the current study. Seventyfive cases were ZCT tests, 78 were Modified General Question Technique (MGQT) tests, and 7 were You Phase examinations. Examinees were mostly male, ranged from 14 to 60 years of age, and consisted of witnesses, suspects, and victims. The crime under investigation in 51% of the cases was theft, robbery, or larceny. Ground truth was determined by incontrovertible physical evidence and/or a confession by the subject or another person. The numbers of guilty and innocent subjects for each test type are presented in Table 1.

Table 1. Frequencies of guilty and innocent subjects for each test type.

Test type	Guilty	Innocent	Total
ZCT	40	35	75
MGQT	39	39	78
You Phase	1	6	7
Total	80	80	160

CPS used the second version of OSS when there were at least three relevant questions and each relevant question was preceded by a comparison question. If this criterion was not met, CPS used the first version of OSS. There were between two and four relevant questions per chart across the 160 cases. OSS total scores were used for statistical analysis.

Discriminant scores were computed for each case by obtaining respiration line length, skin conductance amplitude, and blood volume amplitude measurements for each relevant and comparison question for each chart. Measurements for relevant and comparison questions were averaged across charts and converted to z-scores within the subject. An index of differential reactivity for each measurement for each subject was computed by subtracting the mean z-scores for the relevant questions from the mean zscores for the comparison questions. Each index of differential reactivity was multiplied by a discriminant coefficient and the products were summed to calculate a discriminant score for each subject. These discriminant scores were used for the statistical analyses.

Results

Independent samples t-tests were conducted to test for differences between guilty and innocent subjects in discriminant scores and total OSS scores. As expected, these differences were present for both methods for two of the test types [ZCT_{PA}: t(73) = -7.79, p <.001; MGQT_{PA}: t(76) = -8.43, p < .001; ZCT_{OSS}: t(73) = -7.53, p < .001; MGQT_{OSS}: t(76) = -9.15, p < .001.] We were not able to perform a t-test for the You Phase examinations because there was only 1 person in the guilty group. The means and standard deviations for each test type and group are presented in Table 2.

Tost type	Probabilit	y Analysis	OSS		
lest type	Guilty	Innocent	Guilty	Innocent	
ZCT	-1.29 (1.00)	.51 (1.00)	-19.18 (22.02)	18.17 (20.71)	
MGQT	-1.59 (.86)	.19 (1.00)	-25.15 (18.96)	22.87 (26.75)	
You Phase	-1.86 *	67 (.77)	-18.00 *	11.33 (16.35)	

Table 2. Means (and standard deviations) for guilty and innocent subjects for PA and OSS.

*More than one subject is needed to compute a standard deviation.

OSS and discriminant scores did not differ by sex, race, number of comparison or relevant questions, status as a suspect, witness, or victim, or age. Significant analyses of variance indicated that both OSS and discriminant scores differed as a function of crime type (ps < .05). The mean for sexual offenses was larger than the mean for theft/robbery/larceny offenses, (OSS: $M_{diff} =$ 22.88, $SE_{diff} = 6.63$, t(108) = 3.45, p < .05; discriminant scores: $M_{diff} = 1.01$, $SE_{diff} = .27$, t(108) = 3.76, p < .001).

CPS was used to derive the probability of truthfulness and the OSS score for each subject. The decision made by CPS was compared to ground truth to determine the frequencies and percentages of correct, wrong, and inconclusive decisions. The percentage of correct decisions (CD) excluding inconclusive classifications also is presented. These values are presented in Table 3 for OSS and Table 4 for PA. Because the OSS and PA were developed and validated for ZCT tests and not for other test types, the results are presented separately for the three different test types.

Table 3. Frequencies (and percentages) of correct, wrong, and inconclusive decisions and percentages of CD excluding inconclusive classifications for OSS.

Test		Guilty				Inı	nocent	
type	Correct	Wrong	Inconclusive	CD	Correct	Wrong	Inconclusive	CD
7.07	29	6	5	80.0	27	4	4	07
ZCI	(72.5)	(15)	(12.5)	82.9	(77.1)	(11.4)	(11.4)	07
MCOT	34	1	4	07	30	4	5	00 0
MGQ1	(87.2)	(2.6)	(10.3)	97	(76.9)	(10.3)	(12.8)	00.2
You	1	0	0	100	3	1	2	75
Phase	(100)	(0)	(0)	100	(50)	(16.7)	(33.3)	75

Table 4. Frequencies (and percentages) of correct, wrong, and inconclusive decisions and percentages of CD excluding inconclusive classifications for PA.

Test	Guilty				Inr	nocent		
type	Correct	Wrong	Inconclusive	CD	Correct	Wrong	Inconclusive	CD
707	28	7	5	80	27	3	5	00
ZCI	(70)	(17.5)	(12.5)	80	(77.1)	(8.6)	(14.3)	90
MCOT	32	3	4	014	25	5	9	000
MGQ1	(82.1)	(7.7)	(10.3)	91.4	(64.1)	(12.8)	(23.1)	00.0
You	1	0	0	100	2	2	2	FO
Phase	(100)	(0)	(0)	100	(33.3)	(33.3)	(33.3)	50

An examination of Tables 3 and 4 demonstrates that OSS performed slightly better than PA⁴. OSS had fewer inconclusives and more correct decisions than did PA. The one exception was for innocent subjects who were given a ZCT test. PA had a higher percentage of correct decisions (90%) than did OSS (87%). Chi-square tests indicated that both methods had detection rates above chance for both ZCT tests and MGQT tests (ps < .05), suggesting that OSS and PA perform well across different test types. The chisquare tests for the You Phase examinations were not significant, but there were only seven examinations Phase and You strong conclusions cannot be made for this test type.

Table 5 demonstrates the agreement between OSS and PA in correct, wrong, and inconclusive decisions for the three test types. The two methods were in agreement for

Inconclusive

Γ

correct decisions approximately 72% of the time (excluding You-Phase examinations), but there were instances in which both methods misclassified cases. There also were instances where OSS was able to make a decision and PA was not and vice versa.

The area under the ROC curve was computed to provide a measure of accuracy (Bamber, 1975). There were too few You Phase cases for an accurate computation, so only the results for ZCT and MGQT are presented. For PA, the area under the curve was .894 for ZCT tests and .932 for MGQT tests. For OSS, the area under the curve was .887 for ZCT tests and .948 for MGQT tests. There were no significant differences in area under the curve between OSS and PA for ZCT and MGQT tests (ps > .16). Accuracy for the two methods is comparable for each test type.

099

2

707		055			
	201	Correct	Wrong	Inconclusive	
Correct		54	0	1	
PA	Wrong	0	6	4	
	Inconclusive	2	4	4	
MGQT		OSS			
		Correct	Wrong	Inconclusive	
	Correct	56	0	1	
PA	Wrong	1	3	4	

Table 5. Frequencies for correct, wrong, and inconclusive decisions between OSS and PA.

You Phase		OSS			
		Correct	Wrong	Inconclusive	
	Correct	3	0	0	
PA	Wrong	0	1	1	
	Inconclusive	1	0	1	

7

4

⁴ The finding that OSS generally performed better than PA was surprising. Linear and quadratic discriminant function analyses were performed to determine classification accuracy using the indices of differential reactivity for respiration, skin conductance, and cardiograph. Using the discriminant coefficients from this sample to predict guilt theoretically should maximize accuracy. For ZCT tests, 80.0% of guilty and 82.9% of innocent subjects were correctly classified with a linear function. For the quadratic function, 82.5% of guilty and 82.9% of innocent subjects were correctly classified. Classification was identical for the linear and quadratic functions for MGQT tests. For MGQT tests, 89.7% of guilty and 82.1% of innocent subjects were correctly classified. For You Phase tests, the one guilty subject and 83.3% of innocent subjects were correctly classified with a linear function. For the quadratic function, the guilty subject was correctly classified, but none of the innocent subjects were correctly classified. Results for the You Phase test should be interpreted with caution because only 7 subjects received a You Phase exam, and of those, only one subject was guilty. OSS tended to have higher percentages of correct decisions than PA even when the optimal situation for PA was used.

Conclusion

The results of this study suggest that both PA and OSS are valid and acceptable methods for scoring polygraph data. Both methods had high percentages of correct decisions. There were no significant differences between OSS and PA in area under the ROC curve for ZCT tests or MGQT tests. This suggests that the two methods are comparable in terms of accuracy. It should be noted that Krapohl and McManus (1999) used data collected in the field with Axciton polygraphs, whereas Kircher and Raskin (2002) based their model on data collected with scientific instruments. Classification rates could differ because of differences in instrumentation or characteristics of samples available for model development and validation.

The classification accuracies obtained in the present study generally were slightly lower than those seen in prior research. Krapohl and McManus (1999) achieved 93.3% and 84.6% correct decisions for innocent and guilty subjects, respectively, in one of their OSS validation samples, whereas in the present study, OSS correct decisions for ZCT tests were 87% and 82.9% for innocent and guilty subjects, respectively. In contrast, Gordon et al. (2006) achieved 100% correct decisions using OSS with a sample of 11 cases. Kircher and Raskin (2002) obtained correct classification percentages between 77 and 100 for guilty subjects and between 86 and 94 for innocent subjects when using PA. Correct decisions for innocent and guilty subjects who received a ZCT in the present study were comparable. Classification rates in the present study could differ from what previous research has found because of the quality of the physiological data collected or differences in test administration.

Although OSS was developed and validated with ZCT tests with three charts and three relevant and three comparison questions per chart, the present results suggest that it also may be a viable method for MGQT tests. The difference in accuracy between the two methods was not statistically significant. OSS also appears to be valid for You Phase tests, but strong conclusions should not be made because only one subject was given a You Phase test in the present sample. Further research is needed before conclusions are made regarding the efficacy of OSS for test types other than ZCT.

Both methods are implemented in the CPS software and can be obtained with a mouse click. As noted by Krapohl and McManus (1999), using the OSS rules to score a chart by hand takes time. Scoring charts by hand using PA is not a reasonable endeavor, so the choice of scoring method will be dictated in part by software and time availability.

References

- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, <u>12</u>, 387-415.
- Gordon, N. J., Mohamed, F. B., Faro, S. H., Platek, S. M., Ahmad, H., & Williams, J. M. (2006). Integrated zone comparison polygraph technique accuracy with scoring algorithms. *Physiology* & *Behavior*, <u>87</u>, 251-254.
- Kircher, J. C., & Raskin, D. C. (2002). Computer methods for the psychophysiological detection of deception. In M. Kleiner (Ed.), *Handbook of polygraph testing* (pp. 287-326). San Diego, CA: Academic Press
- Krapohl, D. J. (2002). Update for the Objective Scoring System. Polygraph, <u>31</u>, 298-302.
- Krapohl, D., & McManus, B. (1999). An objective method for manually scoring polygraph data. *Polygraph*, <u>28</u>, 209-222.

Respiratory Blood Pressure Fluctuations Observed During Polygraph Examinations

Mark D. Handler and Joel M. Reicherter

Abstract

The purpose of this paper is to propose possible explanations for what is truly being recorded on the cardiograph waveform during normal and deep breathing patterns. The basic principles underlying cardiovascular recordings monitored during a polygraph examination and what may be affecting the waveform as a result of the breathing patterns will be explored. The term "vagus roll," widely employed by the polygraph community to describe the undulating cardiovascular waveform, is physiologically meaningless and not used by the broader physiologic or medical scientific community. In lieu of this polygraph manufactured term, it is suggested the fundamentally more accurate and reflective expression, Respiratory Blood Pressure Fluctuation (RBPF) be used to describe the cardiovascular waveform undulations.

Introduction

The cardiovascular recordings psychophysiological observed during а detection of deception (polygraph) examination reflect a familiar pattern when the subject engages in deep breathing. Polygraph examiners often use the expression "vagus roll" to describe the phenomenon of a cyclical rising and falling of the cardiograph waveform synchronous with respiratory activity (Krapohl & Sturm, 2002; DoDPI, 2006). Explanations have been posited that include physical associated with respiration movement imparted onto the blood pressure cuff and neuronal influence mediated through the parasympathetic nervous system (Krapohl & Sturm, 2002).

A recent internet search for "vagus roll" did not reveal any use of the expression except by the polygraph community. To conform to the physiologic scientific community, respiratory blood pressure fluctuation (RBPF) is an expression which more aptly describes the origin of the cardiovascular oscillations observed during the polygraph examination.

The cardiovascular circulation is a closed system consisting of the heart muscle, arteries, capillaries, and veins. The purpose of the cardiovascular system is to transport nutrients, hormones, enzymes and oxygen to body tissues, and remove metabolic wastes and carbon dioxide.

In polygraphy, heart and blood vessel hemodynamics provide significant diagnostic data that can be used to assess subject historically. veracity. Presently. and cardiovascular recordings have been obtained with a partially inflated sphygmomanometer (blood pressure cuff). One of the diagnostic features often discussed in the polygraph literature is baseline arousal. Baseline arousal is a rise in the pulse waveform from a prestimulus level. Previous investigators have reported the primary cause of baseline arousal is an increase in blood pressure. (Geddes & Newberg, 1977; Handler, Geddes. & Reicherter, 2006). This paper will focus on the undulations of the cardiograph baseline observed during eupneic (normal) and deep or exaggerated breathing cycles and discuss possible underlying causes.

Acknowledgments

The authors are grateful to Dr. Leslie Geddes, Dr. Michelle Grenier, Mr. Ben Blalock, Mr. William Chittenden, Mr. Geoffrey Flohr, Mr. Don Krapohl, and Mr. Jerry Thomas for their thoughtful reviews and comments to an earlier draft of this paper. The views expressed in this article are solely those of the authors, and do not necessarily represent those of the Montgomery County Texas Sheriff's Office, or the APA. Questions and comments are welcome at polygraphmark@sbcglobal.net.

Background

Blood pressure is a measurement of force per unit of area exerted on a blood vessel wall and is typically expressed in units of millimeters of mercury (mmHg) (Marieb, 1999). Since blood pressure in the vascular system oscillates between contracting and cardiac the pressure relaxing cycles, recordings are expressed in systolic (contracting) and diastolic (relaxing) values.

Due to the pressure oscillations between the systolic and diastolic phases of the cardiac cycle, the mean arterial pressure (MAP) may provide enhanced criteria signal value but has not been used by current polygraph scoring systems. Mean arterial pressure is not merely the average between the systolic and diastolic values because the diastolic phase of the cardiac cycle lasts about twice as long as the systolic phase. Since systolic phase is about one-third of the cardiac cycle, the mean arterial pressure can be calculated by multiplying the pulse pressure (systole-diastole) by one-third and adding that value to the diastolic pressure. For instance, a blood pressure of 120/80 would have a mean arterial pressure of 93 mmHg. Calculation: 120-80 = 40; $40 \ge 1/3 =$ 13; 13+80 = 93.

Blood flow rate is a function of pressure differentials and resistance factors which follow the principles of Ohm's law, (I =V/R). The I symbol of the Ohm's equation in hemodynamic measurements becomes the flow rate which can be represented by the symbol F (flow rate) or V (velocity). The systemic blood pressure source is generated by the contraction of the left ventricle. The power or force component in the Ohm's equation, represented by \boldsymbol{V} (voltage) can be replaced by **P** (pressure). Change in pressure (delta P or P1-P2) is actually the more precise value but for the sake of the fundamental concept, **P** can be used. The symbol **R** (resistance) in hemodynamics is primarily determined by total vessel length with vessel diameter and viscosity contributing to the overall R value. With Ohm's law modified to: F = P/R, hemodynamic measurements as they apply to polygraph recording can be evaluated and contribute to the criteria for decision making.

Hemodynamic recordings are based on measurements as a function of time. The oneminute time unit is the customary time bar for this parameter. Cardiac output volume is calculated by multiplying the stroke volume (SV) or ejection volume (EV) of one cardiac cycle **x** the number of cycles generated in one minute. For example, a stroke volume of 80 ml **x** 75 cycles per minute would generate a volume of 5.6 L.

The cardiac volume output can be altered by a change in stroke volume capacity, heart rate or both. Stroke volume capacity can vary according to venous blood volume availability and heart muscle contractile strength. According to the Frank-Starling Law of the Heart principle, when venous return increases, the heart chambers respond by stretching to increase the end diastolic volume (EDV) or preload. On the next cardiac cycle, the systolic contractile force will eject a larger stroke volume which, in turn, increases the systemic pressure. If the EDV continues to increase, the stroke volume will usually continue to increase for several more cycles until the heart reaches a maximum cardiac output efficiency. Total cardiac output is a function of the Frank-Starling mechanism and the sympathetic nervous system influence on cardiac contractile cells. Since there is always some blood volume left in the ventricles after systole, the final stroke volume can be calculated by subtracting the end systolic volume from the end diastolic volume. SV = EDV-ESV.

Venous return is affected primarily by the respiratory pump and the muscular pump. During inspiration, the contraction of the diaphragm will cause an increase in abdominal pressure, and simultaneously decrease thoracic pressure. This combination of cavity pressure changes will act as respiratory pump and will cause more blood to return to the heart and stimulate the Frank-Starling mechanism. Within limits, the greater the inspiration dynamic, the greater the effect will be on the cardiovascular response.

The muscular pump coupled with the respiratory pump, also aids in venous return by squeezing certain veins strategically positioned between the muscles. Exercise such as walking and running compress the veins in the legs which greatly aid the cardiovascular system return of blood to the heart.

The Valsalva maneuver can significantly alter blood flow dynamics and raise blood pressure. Compressing the abdominal muscles while exhaling through a partially closed glottis (opening of the voice box) can prevent cerebral blood loss and pooling in the legs. Combat pilots employ this maneuver to prevent black-outs from negative G forces. Variations of this activity can be used as a polygraph countermeasure.

Bainbridge (1915) observed that right atrial distention produced an increase in heart rate. He found the reflex arc responsible for tachycardia was mediated through an increase in the sympathetic effect and a decrease in the parasympathetic effect (Brownly, Hurwitz & Schneiderman, 2000).

Respiratory Sinus Arrhythmia (RSA) is a phenomenon that was first described by Ludwig in 1847 (Porges, McCabe & Yongue, 1982). The mechanisms responsible for RSA include CNS influence from the cardiac and respiratory centers, afferent feedback from stretch receptors in the lung (Hering-Breuer reflex), ventilation dynamics and barorecpetors in the aortic and carotid sinuses. (Porges, McCabe & Yongue, 1982)

In summary, there are several physiological mechanisms which affect cardiovascular dynamics. Cardiac output and blood pressure can be influenced by central and peripheral sensory receptors. Vasomotor function such as venous return mechanisms and arterial resistance factors can have a major influence on cardiovascular changes during the polygraph examination experience.

Observation

During polygraph examinations, the cardiovascular waveform normally maintains a relatively stable baseline. There are times, however, when the waveform undulates. As discussed above, this undulating pattern has been erroneously referred to as a "vagus roll."

Figures 1 and 2 show examples of a respiratory effect on the cardiovascular channel. Both figures are from the same subject and the sensors were not moved between chart recordings. While collecting the data shown in Figure 1, the subject was instructed to answer "yes" or "no" to the test questions. Note the peaks of the respiration lead the peaks of the cardiograph waveform by approximately two seconds. Figure 2 is a screen shot of data collected using a "Silent Answer Test." The undulations of the cardiograph waveform have decreased markedly during the more typical 14-16 eupneic breathing cycles.

Figure 3 is from a different subject in a laboratory setting. This subject was instructed to sit quietly and not attempt to control breathing cycling. The blood pressure cuff was placed in contact with the subject's chest during the polygraph recordings. A matched peak-to-peak synchrony between the respiration and blood pressure waveform cycles can be observed. Note the difference in the peak-to-peak timing between the two waveforms when comparing Figure 1 and Figure 3.

Discussion

There appears to be a difference in undulating waveforms produced when the cuff is in contact with the subject and when the cuff is not in contact. Respiratory influenced undulations can be caused by cuff to subject contact but also can occur when there is no such contact. The former results in a in waveform which the peak-to-peak synchrony is very closely matched in time. The latter produces a waveform in which the peaks are more delayed. Differentiating between the causes of the undulations then becomes a matter of comparing the peaks of the waveforms.

During breathing, vasoconstrictor neurons are activated in the inspiratory phase leading to rhythmic vasoconstriction of blood vessels controlling blood pressure (Janig, Bursts of sympathetic activity in 2006). human muscle vasoconstrictor neurons are pulsatile unloading generated bv of baroreceptors (Janig 2006). Increased vasomotor constriction results in increased blood pressure. Increasing the depth of breathing can exacerbate this phenomenon because of increased effect on the baroreceptors.



Figure 1. The respiratory effect on the cardiovascular channel (the second waveform from the bottom) can be observed. For the purpose of this figure, the respiration channel sensitivities were adjusted to match the settings of those in Figure 2 for comparison. The rise in relative blood pressure follows behind the inspiration cycle. Also, a decrease in finger blood pulse amplitude (bottom waveform) can be seen just prior to the rise in the blood pressure. Note: The EDA channel has been hidden to allow a more clear view of the data of interest.



Figure 2. The same examinee during a Silent Answer Test administered shortly after the waveforms shown in Figure 1 were collected. The components are in the same location and the examinee is in the same body position. The respiratory effect on the cardiovascular channel has been reduced significantly. Note the difference in the respiration rate and depth. The respiration channel sensitivities are the same for Figures 1 & 2. Note: The EDA channel has been hidden to allow a more clear view of the data of interest.



Figure 3. Data collected while purposely placing the blood pressure cuff in contact with the subject's upper body. This allowed the movement associated with breathing to be imparted onto the blood pressure cuff. Note the matching (in time) of respiration and blood pressure waveform peaks.



Figure 4. Measurement of the heart rate during inspiration. RSA can be observed in the measurements taken using the calipers. During the 4.0 second time period of inspiration, the average heart rate value (expressed in beats per minute) was 105 BPM.

Additionally, deep breathing results in negative pressure in the venae cavae resulting in increased blood flow. Increased blood flow results in a larger "pre load" (the amount of blood returning to the right side of the heart) or end diastolic volume which leads to increased blood pressure. RSA has been shown to be directly affected by consciously controlled deep respiration. Increased RSA results in increased heart rate during the inspiration cycle. Figure 4 provides examples of RSA measured using a modern polygraph instrument (Lafayette Instrument Company, Lafayette, In.). Using the calipers provided in the software and fixing them at four seconds, we were able to measure the heart rate during inspiration and expiration. Note the heart rate is greater during inspiration than it was during expiration. Increased heart rate results in increased cardiac output which, in turn, results in increased blood pressure. When an examinee engages in a deliberate pattern of deep and slow breathing we can expect to see a cyclic waveform in the cardiovascular channel. The synchronous rise and fall of the relative blood pressure we may observe is quite possibly a result of any combination of these physiological factors discussed earlier.

The cause of the respiratory/heart rate interaction is the interconnection between the cardiac and respiratory centers in the medulla. Even at rest, there is an increase in heart rate during inspiration and a decrease during expiration (Geddes, 1998). With slow, deep breathing these events are more prominent in the blood pressure record. During polygraph examinations in which the examinee breathing is eupneic, the cardiovascular channel often shows a slight undulation that follows normal breathing. Deep, slow breathing merely increases the variation in rate and pressure.

Conclusion

It is not our intention to suggest that vagal tone does not play a part in the undulating blood pressure wave form. It is suggested, however, the scientifically baseless term "vagus roll" be replaced by a more universally accepted and understood expression such as RBPF to describe the undulating cardiograph waveform.

Polygraph professionals should not use terminology that is in conflict with conventional scientific terminology and understandings. We should invite outside scientists to join us in our efforts to define and understand human physiologic responses in polygraphy. Collaboration with outside experts from sister disciplines can help our field avoid "jargonization" of concepts which have established terms in mainstream science.

References

- Bainbridge, F.A. (1915). The influence of venous filling upon the rate of the heart. J. Physiology, 50, 65-78.
- Brownly, K.A., Hurwitz, B.E. and Schneiderman, N. (2000). Cardiovascular Psychophysiology. In Cacioppo, J.T., Tassinary, L.G. and Berntson, G.G. (Eds.), *Handbook of Psychophysiology*. (p. 234). Cambridge: Cambridge University Press.
- DoDPI (2006). Test data analysis: DoDPI Numerical Evaluation Scoring System.
- Des Jardins, T.R. (1997). Cardiopulmonary Anatomy and Physiology, 3rd Edition, Albany, NY: Delmar Publishers.
- Geddes, L.A. (1998). Medical Device Accidents, with Illustrated cases, 2nd Edition. Boca Raton, FI: CRC Press.
- Geddes, L.A. & Newberg, D.C. (1977). Cuff pressure oscillations in the measurement of relative blood pressure. *Polygraph*, <u>7</u>(2), 113-122.
- Handler, M.D., Geddes, L.A., and Reicherter J. M. (2007) A Discussion of Two Diagnostic Features of the Polygraph Cardiovascular Channel. *Polygraph*, <u>36</u>(2), 70-83.
- Janig, W (2006). The Integrative Action of the Autonomic Nervous System: Neurobiology of Homeostasis. New York: Cambridge University Press.
- Krapohl, D.J. & Sturm, S. (2002). Terminology reference for the science of psychophysiological detection of deception. *Polygraph*, <u>31(3)</u>, 154-239.

Marieb, E.N. (1999). *Human Anatomy & Physiology*. Old Tappan, NJ: Benjamin Cummings Publishers.

Porges, S.W., McCabe, P.M., & Yongue, B.G. (1982). Respiratory-heart rate interactions: Psychophysiological implications for pathophysiology and behavior. In Cacioppo, J.T. & Petty, R.E. (Eds.), *Perspectives in Cardiovascular Psychophysiology*. (pp. 223-259). New York: Guilford Press.

Accuracy Demonstration of the Horizontal Scoring System Using Field Cases Conducted with the Federal Zone Comparison Technique

Donald Krapohl¹, Nathan Gordon², and & Christopher Lombardi³

Abstract

The Horizontal Scoring System (HSS) is a rank ordering method for the analysis of polygraph charts. In the present study 100 confirmed field cases were blindly scored using the HSS. Applying the HSS cutoff scores for multiple-facet testing formats, 86% and 68% of the deceptive and truthful cases were correctly classified, respectively, with an overall inconclusive rate of 9%. Average accuracy without inconclusives was 84%. Using total scores along with asymmetrical decision rules proposed in other research for single-issue techniques, 82% and 84% of the deceptive and truthful cases were correctly classified, respectively, with an overall inconclusive rate of 4%. Accuracy averaged for truthful and deceptive cases, excluding inconclusives, was 86%, with all of the accuracy improvement coming from truthful cases.

The Horizontal Scoring System (HSS, Gordon & Cochetti, 1987) was devised with the aim of reducing the level of subjectivity in manual scoring of polygraph examinations conducted using a comparison question technique (COT). Rather than assigning scores based on the differences in reaction intensity between question pairings as is done in the 7-position scoring method, Gordon and Cochetti suggested that a ranking strategy across an entire chart could afford a more objective assessment of the data. To further ensure objectivity, assignment of ranking to tracing features was based on measurements. The HSS is taught as part of the Integrated Zone Comparison Technique for which there is published research (Gordon, Fleisher, Morsie,

Habib, & Salah, 2000; Gordon, Mohamed, Faro, Platek, Ahmad, & Williams, 2006).

The idea of approaching chart analysis from a ranking perspective is not new. Lykken proposed an abbreviated ranking system to score his Guilty Knowledge Test as 1959. Japanese also early as The experimented with ranking systems for the electrodermal channel (Suzuki, Watanabe, Ohnishi, Matsuno, & Arasuna, 1973; Suzuki, Ohnishi, Matsuno, Arasuna, 1974) and Timm (1982) applied a ranking system to respiration line length. The HSS (Gordon and Cochetti, 1987) expanded ranking to all of the polygraph channels and added decision rules for multiple-issue, multiple-facet and singleissue formats.

¹ Donald J. Krapohl is Past President of the American Polygraph Association (2006-2007) and regular contributor to this publication.

² Nate Gordon is Director of the Academy for Scientific Investigative Training, an American Polygraph Association accredited school since 1980.

³ Christopher Lombardi is a detective with the Morris County (NJ) Prosecutor's Office.

The views expressed in this report are those of the authors, and do not necessarily represent those of the US Government, Department of Defense, the American Polygraph Association, or the Morris County (NJ) Prosecutor's Office. Request for reprints should be directed to the first author at dkrapohl@aol.com.

The original Gordon and Cochetti (1987) article did not identify which features were ranked. The HSS uses measurements of specific features. These features are listed below. All measurement units, including the time dimension, are in millimeters.

<u>Respiration</u>: Two possible methods.

1. Respiration line length (Timm, 1982) for 20 seconds, or;

2. The duration (D) of four respiration cycles divided by the sum of heights (H) of four respiration cycles squared. The formula is $D/(H_1 + H_2 + H_3 + H_4)^2$. Three cycles may be permitted for occasions where four cycles cannot be obtained. However, the three cycle consistency must prevail throughout that particular chart.

<u>Electrodermal</u>: Amplitude (A) squared times the duration (D). The formula is $(A)^2$ X D.

<u>Cardiovascular</u>: Amplitude measured by the increase of blood pressure as shown by the change of the bottom of the tracing from beginning of question to its highest rise within 50mm (where 1 second = 2.54mm.)

Each channel (respiration, electroderand cardiovascular) is ranked mal individually, with the largest reaction within each channel receiving the highest rank. In the typical Zone Comparison Technique (ZCT) there are three comparison and three relevant questions, for a total of six questions to be ranked. In this example, the greatest reaction in each channel would receive a "6", the second largest a "5" and so forth to a "1" going to the least reaction. The sign of the rank depends upon the type of question: the ranks of relevant questions are given a negative (-) sign, while a positive sign (+) is for ranks to comparison questions. The ranks for the two pneumograph channels are averaged and then added to the ranks of the EDA and cardiovascular channels by question, by chart, and for the entire examination. Decision cutoffs for a spot analysis on multiple-faceted examinations are +/- 4.5 for three charts, and +/- 6 for four charts. Decision cutoffs for single issue examinations are +/-13 for three charts, and +/-18 for four charts.

The HSS has not been subject to research previously independent of the Integrated Zone Comparison Technique. The research question of interest was whether it, when combined with the Federal Zone Comparison Technique, would deliver sufficient accuracy to meet the APA standards for evidentiary (90%), paired testing (86%), or investigative (80%) examinations as well as remain below the maximum inconclusive rate of 20% (APA Standards of Practice, 2007).

Method

Data Source

One hundred cases previously selected for the development of Evidentiary Decision Rules were used (Krapohl & Cushman, 2006). All were Federal Zone Comparison Technique examinations conducted in the field. Each had three relevant questions, three probablelie comparison questions, and three charts. The cases were delivered in electronic form to the scorer.

Scorer

A scorer (third author) who was very experienced with the HSS performed the analysis of the 100 cases using the HSS under the guidance of the second author. The scorer was kept blind to ground truth, base rates, and the decision of the original examiner. The scorer submitted an electronic data sheet that included his decisions by question and by case. Only the case decisions are considered here.

Decision Rules

The Federal ZCT has two very focused and nearly identical relevant questions that cover the central test issue, and one relevant question that may address broader participation in the crime under investigation, knowledge of the participants, or the examinee's connection to a piece of evidence. These characteristics in the strictest sense make the Federal ZCT a multiple-facet examination in that the examination covers one crime but the relevant questions address different aspects of the crime. As such, the HSS multiple-facet rules would apply. The creator of the HSS, the second author, teaches that the multiple-facet rules should be used with the Federal ZCT, and therefore we used these decision rules.

The multiple-facet decision rules are based on the total score for individual test questions. If the total score for any relevant question is lower than -4.5, the decision is DI. When every relevant question has a score of +4.5 or more, the call is NDI. All else is inconclusive. Because the Federal ZCT has two relevant questions that cover the exact same issue, the scores of those two questions can be added together, and the decision cutoffs doubled to a combined +/-9.

Results

The HSS correctly called 84% of the deceptive cases and 68% of the truthful cases,

a difference which was not statistically significant (z = 1.62, ns). Errors for the deceptive and truthful cases were 14% and 16%, respectively, a difference which also did not exceed chance (z = 0.44, ns). Inconclusives were found in 2% of the deceptive cases and 16% of the truthful cases, which did prove to exceed chance expectancy (z = 5.80, p < .001). The accuracy averaged for the truthful and deceptive cases when inconclusives were excluded was 84%. Overall accuracy was significantly above chance both counting inconclusives (z = 3.89, p < .001) and excluding inconclusives (z = 4.96, p < .001). Figure 1 shows the percentages of correct decisions, errors and inconclusives.

Figure 1. Percentage of correct decisions, errors and inconclusives using the HSS with multiple-facet decision rules on 100 Federal ZCT examinations.



Krapohl, Dutton and Ryan (2001) previously conducted an evaluation of a related rank ordering method using again Federal ZCT cases, but applying single-issue decision rules. They proposed alternate cutoff scores, -13/-1, a method which reduced the proportion of inconclusive decisions in that study. We applied these decision rules to the total score of the cases in the current study. Using these alternate cutoff scores, the HSS correctly called 82% of the deceptive cases and 84% of the truthful cases, accuracies which were not significantly different from one another (z = 0.27, ns). Accuracy averaged for the truthful and deceptive cases without Errors for the inconclusives was 86%. deceptive and truthful cases were 16% and 10%, respectively, which was a non-significant difference (z = 0.89, ns). Inconclusives were found in 8% of the deceptive cases and none of the truthful cases, a difference which was significant (z = 2.04, p < .05). Accuracy averaged for truthful and deceptive cases was significantly above chance both when counting inconclusives (z = 5.10, p < .05) and when inconclusives were excluded (z = 5.49, p< .05). Figure 2 shows the percentages of correct decisions, errors and inconclusives.

Figure 2. Percentage of correct decisions, errors and inconclusives using the HSS with proposed cutoff scores of -13/-1 on 100 Federal ZCT examinations.



Tests of proportion (Bruning & Kintz, 1997) found no statistical differences in overall accuracy between the traditional HSS cutoffs and the alternate cutoffs, in the overall error rate or the overall inconclusive rate. See Table 1.

	Truthfu <u>Correct</u>	l Cases (<u>Error</u>	N=50) <u>Inc</u>	Deceptiv <u>Correct</u>	re Cases <u>Error</u>	(N=50) <u>Inc</u>
Traditional Multiple-Facet Rules	34	8	8	42	7	1
New Decision Rules (-13/-1)	42	8	0	41	5	4

Table 1. Correct, erroneous and inconclusive totals for the Traditional HSS decision rules and the
New Decision Rules for truthful and deceptive cases.

Discussion

In this assessment of the HSS it was found to produce high accuracies. Using both the traditional multiple-facet decision rules and new decision rules the HSS had accuracies high enough to meet investigative purposes (80%). HSS accuracy fell below that necessary for paired testing (86%) and 2007). evidentiary testing (90%) (APA, Excluding inconclusive results, HSS had an accuracy of 84% with the multiple-facet decision rules and 86% for the new decision rules. By way of comparison, the average accuracy without inconclusives using the 7position scorings of these same cases in Krapohl and Cushman (2006) was 87% for the Evidentiary Decision Rules. Though 7position scoring and Evidentiary Decision Rules combined to produce accuracy slightly higher than that of HSS, that difference was not statistically significant (z = 0.32, ns).

In future research there may be profit in the exploration of optimization of the processes within the HSS that could improve its accuracy. One area could be the weighting of the electrodermal channel, an approach found useful in algorithms such as CPS, PolyScore, and OSS but not used with the HSS. It may also be useful to further refine the features used in the ranking for HSS. Additionally, it would be worthwhile to explore scoring ZCT examinations which were formulated with pure, single issue relevant questions i.e.; strict adherence to ZCT R10 questions evidence connecting while adjustments in cutoffs may also improve its accuracy. Once these and other optimization methods have been explored, it would be informative to replicate the present research using not only the HSS, but also the other scoring methods from the field to allow a head-to-head comparison of these methods.

References

American Polygraph Association. (2007). APA Standards of Practice.

- Bruning, J.L., & Kintz, B.L. (1997). Computational Handbook of Statistics (4th Ed.). Addison Wesley Longman: Reading, MA.
- Gordon, N.J., & Cochetti, P.M. (1987). The Horizontal Scoring System. Polygraph, 16(2), 116-125.
- Gordon, N.J., Fleisher, W.L., Morsie, H., Habib, W., & Salah, K. (2000). A field validity study of the Integrated Zone Comparison Technique. *Polygraph*, <u>29</u>(3), 220-225.
- Gordon, N.J., Mohamed, F.B., Faro, S.H, Platek, S.M., Ahmad, H, and Williams, J.M. (2006). Integrated zone comparison polygraph technique accuracy with scoring algorithms. *Physiology and Behavior*, <u>87(2)</u>, 251-254.
- Krapohl, D.J., & Cushman, B. (2006). Comparison of evidentiary and investigative decision rules: A replication. *Polygraph*, <u>35(1)</u>, 55-63.
- Krapohl, D.J., & Dutton, D.W., & Ryan, A.H. (2001). The Rank Order Scoring System: Replication and extension with field data. *Polygraph*, <u>30</u>(3), 172-181.
- Lykken, D.T. (1959). The GSR in the detection of guilt. *Journal of Applied Psychology*, <u>43</u>(6), 385-388. Reprinted in *Polygraph*, <u>7(2)</u>, (1978), 123-128.
- Suzuki, A., Watanabe, S., Ohnishi, K., Matsuno, K., & Arasuna, M. (1973). The objective analysis of GSR in the detection of deception: An analysis of GSR amplitudes in terms of rank scores. *Reports of the National Research Institute of Police Science*, <u>26</u>(4), 237-245.
- Suzuki, A., Ohnishi, K., Matsuno, K., & Arasuna, M. (1974). Amplitude rank score analysis of GSR in the detection of deception: Detection rates under various examination conditions. *Reports of the National Research Institute of Police Science*, <u>27</u>, 142-148. Reprinted in *Polygraph*, <u>8</u>(3), 242-252.
- Timm, H.W. (1982). Analyzing deception from respiration patterns. *Journal of Police Science and Administration*, <u>10</u>(1), 47-51.

An Introduction to the APA's Panel on International Developments in Polygraphy

Frank Horvath, Ph.D.

In 2005, at the APA seminar in San Antonio, the first-ever APA "International" panel was organized. The purpose of the Panel was to serve as a forum for discussion of contemporary events in Polygraphy and Credibility Assessment in countries outside of the U.S. The advent of the internet, changes in social, political and legal areas, the menace of terrorism and the growing problem of transnational crime have raised the need for an awareness of developments in the field of Polygraphy. The panel was the APA's initial step in that direction.

The International Panel presentations, hopefully, will be a continuing feature of the annual seminar. Generally the organization will be as follows: Each Panel will consist of three or four presenters, each from a different country. Panelists will make a 20-30 minute presentation, after opening remarks from the After the featured "country" moderator. presentations, the moderator will summarize and integrate the important points. That will be followed by a question and answer session, with questions posed by audience members to the panelists. This organization is intended to promote greater interest in international issues and a better understanding of how practices and policies in other countries are related to those in the U.S.

In addition to a presentation at the seminar, each panelist also agrees to prepare a more detailed paper, in a relatively consistent way, that will be submitted to the APA's Editor for publication consideration. Examples of items that are to be covered in each of the papers include: Who is credited with the initial development of polygraph testing in the country? When? Who uses polygraph testing? How many examiners are there and how are they selected and trained? What kind of instrumentation is used? What are the dominant procedures ("techniques") in What are the legal issues of most use? concern? What is the public perception of Polygraphy?

In issue 37(2) of *Polygraph* a paper from the second International Panel was published. In this issue of we are pleased to publish a paper from the second International Panel in 2006. In this paper the author describes Polygraphy in Venezuela, a country widely discussed in the U.S. popular media, primarily because of the views and actions of its political leaders. How Polygraphy develops in that country is something that examiners ought to watch closely.

About the author

Esteban Balthazar Haden earned a college degree in mechanics from the Instituto Universitario de Technologia in Caracas, Venezuela in 1979. In 1993 he graduated from Universidad Católica Andrés Bello, also in Caracas, with a degree in law. He completed the Venezuelan bar examination in 1993. From 1979 to 1991 Esteban was a security advisor and investigator for the Detecta Organizacion Organizacion у Centurion, C.A. From 1991 to today he has been a Director for Consultores Securitas, C.A., a company dedicated to security assessments and consulting. In addition, he has had a legal practice in Venezuela since 1993 and has been a Director for Soluciones Perimetrales, C.A. since 1996.

Esteban received his basic polygraph training from Dr. Gabriel Gazso in Caracas in 1983. He has attended a large number of training seminars in the U.S. and Latin American, including the ALP seminars in 2004 and 2005, the Florida Polygraph Association in 2005, and the American Polygraph Association in 2006. He also completed a number of advanced training courses including an advanced chart evaluation course in Columbia in 2005. He completed the APA accreditation test in 2004 and has served as an instructor in basic polygraph training courses offered by the The Academy of Polygraph Science in Caracas, Venezuela, specializing in instrumentation operations and legal aspects of Polygraphy.

Aside from his training in Polygraphy, Estaban has completed a large number of courses and seminars in management, security management, security systems and other security-related topics. He maintains an interest in a wide range of security management issues and has been a member of the American Society for Industrial Security since 1983. He has served as the Vice-President of the local ASIS chapter since August 2005. He is also a member of the Venezuelan Association of Security Executives, a membership held since 1990. He is a full member of the American Polygraph Association and an affiliate member of the Florida Polygraph Association.

I wish to thank Esteban for his time, effort and sincere interest in sharing this information with the APA membership.

History of Polygraphy in Venezuela

Esteban Balthazar

Beginnings of polygraphy in Venezuela

Polygraphy in Venezuela was born in early seventies, during Dr. Rafael the Caldera's administration. At that time, a small group of police professionals from the Venezuelan National Intelligence Service -DISIP, some of them had already finished the FBI training course in Quantico -- decided to learn polygraphy. They intended to use it for counter-intelligence investigations. The promoter of this idea was Dr. Gabriel Gazsó, at that time assistant director of the DISIP. He contracted with a polygraphist, whose name was Collazo, to give him and Commissioners Francisco Rotundo and Árpád Bangó, a basic polygraph training course in the DISIP headquarters.

In DISIP the tests were mainly applied to police applicants for DISIP and in criminal investigations. It is important to point out that at that time the country had passed through 10 years of armed conflict with leftist guerrillas who wanted to take power by force. They were supported by the Soviet Block and by Fidel Castro in Cuba.

After some time the polygraph was used in the DISIP for two main purposes. The first one was for their own internal investigations. Here testing was used to detect possible infiltrators, concealment of information, or sabotage by agents who were employed in intelligence and counterintelligence tasks. The second purpose for the testing was auxiliary tool for as an investigators to verify the veracity of confessions and statements obtained during interrogations.

During the government of President Caldera, the armed conflict with the guerrillas diminished substantially by means of a political process called "the pacification." This process tried to incorporate guerrilla members into society, granting them a pardon if they were willing to give up their guns. Many of them supported this process sincerely, but some used it to infiltrate the police and once inside, became double agents loyal to the guerrillas, not the government.

Training of the first polygraph examiners

As far as I could determine the first group of examiners received training only in a personalized manner. The classes were given on convenient schedules according to the student's availability. There were no formal classes; nor did the students have to register in an Academy. Nevertheless, the students dealt directly with the Instructor in both theoretical and practical classes. This type of flexible schedule was needed because the obligations and responsibilities these people had inside the police agency were very important. They couldn't leave their jobs for weeks or months.

Later on, in the 1980's, Dr. Gabriel Gazsó and Francisco Rotundo, already working in the private field, trained Eng. Arno Horvath, Ismael López and myself. This training was also given in personalized way, covering both the theory and the practice of testing.

At the end of the 1980's, Dr. Gazsó and I participated in the training of four civil officers of the Technical Judicial Police, PTJ. (The PTJ is the Venezuelan equivalent to the FBI dedicated primarily to carrying out criminal investigations). This training was supported by Commissary Rojas Ochoa, at that time Chief of the Technical Room and Microanalysis of the PTJ. He acquired a fivechannel analog Lafayette instrument and created for the first time a physical space in the Judicial Police headquarters dedicated to polygraphy. This was called the "polygraph room." These four officers attended a 10-week theoretical and practical course prepared by Dr. Gazsó. They had six training hours per week. Unfortunately, this group diminished rapidly and then completely disappeared due to internal problems in the PTJ.

At the same time, we also gave magisterial classes of approximately two academic hours for three consecutive years to the students in their last year of Law at the Catholic University. Andrés Bello introduced them to the instrument and made several tests to show its efficiency and to induce the students to use testing later on as a tool in their practice of criminal law.

At the beginning of the 1990's, a group of approximately five civil officers of the DISIP attended a polygraph training course in Mexico. Later on, more people might have been trained since it is known that at that time the political police had 10 polygraph instruments. Nevertheless, and as is always the case, this information was managed with much discretion. It was practically secret to the public light. The polygraph continued to be a "low profile" topic in Venezuela.

Today we have in Caracas the Academy of Polygraph Science, founded in 2002 with the patronage of Dr. Richard Poe from Florida and Eng. Arno Horvath in Venezuela. This academy already has given three 400-hour basic polygraph training courses. Each course has been ten weeks in length, with five 8-hour classes per week. The Academy has already graduated eight polygraphists and has done advanced training courses in Colombia.

Polygraphists in Venezuela

The first important fact here is that the exact number of examiners that work in Venezuela is unknown. We do not have much information about the governmental field. As far as can be determined the Judicial Police (PTJ), an institution that should be using the polygraph, has neither the instruments nor people dedicated to the science.

Regarding the DISIP we were able to contact a group of polygraphists that work in this institution. I was told that the institution has three active polygraph examiners and three that are not. They were all trained in Mexico in 2000 and they work with two computerized Lafayette instruments and one computerized Stoelting instrument.

On the other hand, and as expected, we have more information about the private

sector. We know approximately 15 people involved with polygraphy. Nevertheless, there are only eight persons dedicated exclusively to this field. Four of them are female. These eight were trained in the private field. The remainder of the 15 were with the police, but they are no longer civil officers. Of the eight privately trained. four females are psychologists, while among the men we have one engineer, one attorney and one is a student of the last year of a university career; the remaining examiner is not a university graduate at this time.

Despite the lack of requests for polygraphy in Venezuela we know that three years ago a polygraphist living in the USA was contracted by a North American airline to come to Venezuela to conduct preemployment and routine (periodic) tests. This person stayed in Venezuela for two days and conducted between 12 and 16 tests in that period of time. Make your own conclusions about his professionalism and the results of the tests that were done.

We can conclude that polygraphy is still a science or area of expertise in Venezuela that has not yet obtained enough support or acceptance from the governmental or the judicial sectors, though it is a tool that is being used with a lot of success by some private companies. We also have come to know that in Venezuela some unscrupulous people call themselves polygraph examiners and offer their services to determine truth or lies with a small instrument that supposedly turns on a red light when a suspect lies and a green light when the truth is told. I guess we all know what gadget is being used here.

This is one of the reasons why this year some Venezuelan polygraphists have become interested in creating the Venezuelan Polygraph Association. This association will have bylaws concerning minimal requirements for becoming a member and it will regulate polygraph practices. With this Association there is interest in improving the polygraph field and making it a profession officially supported by law. At this time, though, the Association is not a reality. We hope that will change in the coming year or so.

Polygraph testing in Venezuela

Due to the lack of information and contact with police or governmental examiners it is impossible for us to determine what happened with their instruments and for what and where they are being used. We also do not know what happened to the rest of the examiners. The only information we have is that today's police examiners are mainly using the polygraph to test bodyguards for embassies.

In the private sector there are three companies dedicated exclusively to personnel selection and polygraphy or combinations of the two. According to information obtained from the Directors of these companies they each do approximately between 100 and 150 tests each month. Of these, about 70% are pre-employment tests, 20% routine screenings and 10% specific issue tests.

Countermeasures in Venezuela

Countermeasures made by examinees are not yet a big problem in Venezuela. Two of the companies use a movement sensor to detect physical countermeasures. This problem does not affect us like it does examiners in the USA and elsewhere due to the fact that many examinees have no computer or access to the Internet. They do not have much knowledge about polygraphy or access to the anti-polygraph web sites that advertise countermeasures.

The main countermeasures we have observed in the recent years in Venezuela are: body movements, constant eye movements, for example, looking at or watching the ceiling or staring at the floor, movements of arms and legs, ingestion of tranquilizers or alcohol, lack of sleep and excessive irritability. In some cases a complete rejection of the testing is done. Depending on the know-how of the examiner these behaviors are usually detected and proper counter-countermeasures are applied.

Polygraph techniques and test formats used in Venezuela

At its beginnings in the 1970's up through the decade of the 1990's, the PE (or the inquiry test) was used for pre-employment testing. The applicants received a 10-page personal record form containing questions about their life. For example, personal data, such as name, current and former addresses, nicknames, telephone numbers, place and date of birth, marital status and the same information about their spouse and children was asked for. There were also questions about parents, though with fewer details. Afterwards, we required complete information about studies, and previous jobs, indicating company, address, telephones, salaries, name the last boss, reason of resignaof tion/dismissal and working period. Applicants also had to provide information about their personal financial balances. indicating properties, credit cards, bank accounts, vehicles, hobbies, alcohol consumption and drug use. With regards to their physical condition they were asked if they were suffering or have ever suffered any of the diseases listed on a table, in which they had to answer with a "yes" or a "no." In the case of an affirmative answer, they had to go into detail. A space was also left in order that they could mention any other disease not listed. They were also asked if they were ever hospitalized, if they had any surgery, or if they had any psychological treatment in the past 10 years. They were also asked about previous police or military experience and about their religion.

Once the form was duly filled out, the applicant was interviewed regarding the information submitted in the form. After that, the polygraph was applied to the subject and a string of 18 questions asked. The questions were related to the form and included two extra questions. The first one of these was if he/she had omitted any information that could harm his work in the company. The second one was if he/she had told the truth in the interview regarding his real intentions to get the requested job.

The polygraph investigation tests were a sort of MGQT (Modified General Question Test) format. The question list included irrelevant questions mixed with relevant ones. A comparison question was placed at the last position of the string.

As a specific issue test format the Peak of Tension Test was also used when its application was possible. As you know, in this test the examinee is confronted with a great deal of facts (items of evidence, such as the weapon used, the method of entry into a home or room, and so forth) that only the victim, the criminal and the investigator could know. In this case seven questions were asked first in ascending and then descending order. The first question might be: "Have you used weapon Nr. 1 to kill Mrs. xxxx?" The second question: "Have you used weapon Nr. 2 to kill Mrs. xxxx?", etc. The test was repeated twice, and in case of any doubt, three times.

Today the Modified General Question Technique and the Zone Comparison Technique are the most used specific issue testing formats. The Peak of Tension test may be used on occasion when it is appropriate.

Research and the polygraph instruments used in Venezuela

Due to the few publications about polygraph testing and the low acceptance of it in the country, polygraphists have neither the resources nor the time, whether private or governmental, to develop their own research. Nevertheless, we are making great efforts to improve the quality of our work. Many of us attend at least once a year seminars in the USA, Colombia or the Dominican Republic. This helps to keep us updated on some of the new research, technology and techniques.

Among the eight graduates of the Academy of Polygraph Science-Caracas, founded in 2002, some of them already have attended continuous education courses and seminars. Moreover, they have fulfilled or are fulfilling the supervision phase following their graduation from the Academy, in order to conduct examinations so the instructors can evaluate their work and observe their development.

The Venezuelan Army, through its Commander, has shown a special interest in polygraphy. At this time five Marine Officers have been sent to the Academy of Polygraph Science – Caracas. All of them have completed satisfactorily the basic polygraph training course. Now they are all in the practical supervision stage and they regularly use the polygraph in their Divisions. With regard to the instruments used in Venezuela, there is a great variety. In the public sector, the Army has two Axciton instruments and they are about to acquire two new instruments with countermeasure sensors. The Political Police use two Lafayette instruments and one Stoelting. We presume that the Lafayette acquired in the decade of the 1980's by the Judicial Police burned in a fire approximately 12 years ago. It has never been replaced.

In the private sector, there is a company that has five Axciton instruments, three of them with a countermeasure sensor. Another company has one or two computerized Stoelting instruments. A new polygraph company acquired an Axciton last year. Another independent polygraphist uses a computerized Lafayette. I myself still get my fingers dirty and my shirts and ties blue with I use an analog Stoelting ink spots. UltraScribe instrument.

Political, social and legal aspects of polygraphy in Venezuela

The common Venezuelan is a mixture of the Indian, Black and Spanish races, but sometimes, one of these races predominates. One of the cultural features inherited from our Spanish ancestors is the figure of the "nobleman," a character of noble lineage, who was part of the Castilian aristocratic hierarchy during the Middle Age in Spain. The main characteristics of this nobleman were to be generous, decent and honorable. Nevertheless, in the Spanish-American trans-culturization the titles remained, but much was lost in its essence.

In Venezuela, when you suggest to a common citizen that in order to verify the veracity of his statements a polygraph examination will be asked for, the first thing he asks is: What is that? Once you tell him that the polygraph is commonly known as the "lie detector", the reaction that follows is rejection of the test, and the subject exclaims some of the following phrases: "My mother taught me to tell the truth. No device is going to measure my honesty!" Or, "I was told that this machine is not good and the result is altered due to the nervousness of the examinee." and some others dare to protest saying: "This machine is an attempt against Human Rights. Its use is prohibited by Law!" All these arguments are the typical behavior of that Spanish nobleman, whose honor could never be doubted, and much less, be tested by some sensors that are connected to a machine. We can conclude that in the country, people do generally reject polygraph testing. In the end, when they do not agree to be tested, the phrase most commonly exclaimed is: "I am innocent, and if you do not believe me, go and look for other proofs, but I will not undergo a polygraph test."

Usually, we can recognize innocent subjects. After the general features of the polygraph are explained to them, the instrument is shown and, if necessary, a Stim Test is applied to demonstrate the reliability of the instrument, they are then less reluctant. After this, we finish the interview with a very popular phrase in Venezuela: "If you don't owe it, you are not afraid of it." In 99.9 % of the cases the subject will agree to undergo the examination and in cases when we have to convince a whole group, the innocent persons seem to accept rapidly. Sometimes guilty subjects "go with the flow," so they won't raise any suspicions.

In the private field, no professional, serious polygraphist undergoes the risk of testing someone without previous written consent. This prevents unnecessary lawsuits. In the DISIP (Political Police), in the 1970's and 1980's, applicants had to sign a consent form in which they stated that they agreed to a polygraph test at any time. Afterwards, the same method was used in companies that routinely applied the polygraph for preemployment, screening and specific issue testing. In many cases, we have to convince first the director of a union in order to start an investigation and later on to use the polygraph.

In my 25 years of experience in the field, with approximately 4,000 examinations to my credit, in only two cases have I been judicially denounced. In both cases the subject, who claimed to have been attacked during the exam, showed up with a District Attorney protesting that his rights were trampled during the test. In both cases, we informed the District Attorney about the polygraph, about what we did during the exam and showed him the consent form his defendant signed before the test. This was enough to dismantle any attempt of being sued.

In Venezuela regarding the Criminal / Legal matter until the year of 1998 the expertise and criminal law procedures were regulated by the former Venezuelan Code, later on substituted with the Venezuelan Criminal Procedures Code. In the former Code of Criminal Prosecution the polygraph was not mentioned as a legal testing procedure; nevertheless, in some particular cases, the Judge allowed the inclusion of a polygraph test as additional evidence, but the polygraph results were not conclusive for the sentence. If the result of the polygraph test was consistent with other evidence and proofs, the Judge included it in his assessment; but if it was not the same, he rejected the test.

In 1994 we had in Venezuela the "Attorney Landaeta Case." Landaeta was a well known private lawyer in Venezuela. He was accused of strangling a very attractive young female attorney in her apartment, with her baby in the same room. The Judge in charge, who knew of the case in Appeal, asked to be tested by the polygraph so she would be able to appraise the polygraph by herself, before she included the case test as evidence in the process. Landaeta failed the polygraph exam, and that fact was considered by the Court as additional circumstantial evidence.

As the new Code of Criminal Procedure entered into force, the spectrum of admitted evidence (kind of proofs) in court was enlarged. Though the polygraph continues without being mentioned, the Code allows the admission in Court of any lawful testing, as a part of (or like) the expertise field. Today, if the Judge or the District Attorney decides it, it is possible to include the polygraph results as "circumstantial evidence" in the set of expertise evidence. Besides this, the expert's qualifications are regulated during the judgment.

In spite of these new conditions favorable to the incorporation of the polygraph in the legal system, the common practice in Venezuela is to avoid taking a case into Court. In the private field, when a suspect is discovered, the Company's Security Department prefers to negotiate the resignation of the person. This avoids a court trial and is also convenient for the criminal because he will avoid criminal records. On the other hand, the company's owner will also be benefitted, because he will save money, time and trouble. It is important to point out that in Venezuela it is extremely rare to get a confession after a polygraph examination is failed. If this occurs the suspect has to ratify his statement in Court or in the presence of a District Attorney, although it has been previously recorded on a tape. If this does not occur, the tape will not be included as evidence. The general case is that the suspect denies his guilt or remains quiet.

The Judicial field should take advantage of polygraphy since the Code of Criminal Procedures offers benefits in certain type of crimes for those who confess their crime or give to the District Attorney useful important information and for the investigation. Nevertheless, and we repeat it, that as far as we could find out, the polygraph is not being used regularly as a form of evidence in court.

Another sector of society that has not helped polygraphy in Venezuela, probably due to ignorance or lack of information, is the mass media. The press and the television have mentioned polygraphy in the past, but generally in a negative way. In the year 1995, a journalist confused the polygraph with the Reid Interrogation Technique and belittled its effectiveness. In the "Attorney Landaeta Case," in the year 1994, they referred to the polygraph without expressing any value judgment or its results. One other thing that was seen on Venezuelan television last year was the weekly Spanish program, which faces people in conflict accusing each other in front of the cameras. Then a polygraphist, on the air, tests them to determine which one of the two is lying! At this moment, we think that

this is not the kind of advertising that projects polygraph testing in a positive way.

I would like to finish my presentation highlighting the positive side of Venezuelan polygraphy. The efforts made in the last five years, under the patronage of Eng. Arno Horvath, Dr. Gabriel Gazsó and Dr. Richard Poe from Florida, allowed the creation of the first polygraph academy in our country. I am absolutely sure that eight polygraphists have already been trained to a high academic and ethical level. Among them we have several university professionals like attorneys and psychologists that will hopefully be the seeds of a new impulse of polygraphy in Venezuela.

The planned creation of the Venezuelan Polygraph Association will also become important for boosting professional polygraphy in Venezuela. Its main promoter in the country is Mr. Carlos Müller, trained by Eng. Horvath in 2000. He took the APA test 2002 and in passed with excellent qualifications. If the Venezuelan association becomes a reality, it will undoubtedly benefit the polygraph community. We are hopeful the association will be formally announced soon.

Last but not least, we have to point out the relevance of the Venezuelan examiner participation in overseas seminars. Every year, at least one Venezuelan polygraphist attends one of the seminars hosted by the Florida Polygraph Association (FPA), the American Polygraph Association (APA), or the Latin-American Polygraphists Association (ALP). The ALP's 2005 Annual Meeting took place in Margarita Island, Venezuela, in which we had the honorable participation and presence of Dr. Frank Horvath who, as always, contributed with new knowledge and his usual and pleasant way of teaching.

Polygraph Research: Value, Shortcomings, and Perspectives

Stuart M. Senter

Polygraph is a unique field that encompasses an interesting and dynamic niche in our society. Polygraph examiners are trained to accomplish a task that, in the mind of the public, should only be made possible through rapid advances in seemingly futuristic technological equipment or through the weaving of mystical powers thought to be wizards proffered by and magicians. Whichever the perspective one might choose to take, we operate in a field that draws both great interest and fascination, in addition to skepticism and criticism. Lie detection or its more encompassing contemporary term, 'credibility assessment' likely draws such a polarized interest due to the excitement that the potential for flawless lie detection exists, on the one extreme, and the disgust and cynical amusement that anyone would fall for such a preposterous notion on the other. As practitioners and researchers in the polygraph field, we must understand that both ends of the spectrum are flawed. The clear path toward producing and providing a more moderate and realistic view of the polygraph field is through increasing and continually updating the body of knowledge about polygraph. This paper briefly explores, in general terms, the mechanisms we have at our disposal to achieve this end.

Before proceeding any further, let us be clear about the role and importance of research, in very broad terms. As with any scientific endeavor, court case, or decision with minor or major implications, there should be an accumulation of *evidence* before a decision is reached. In fact, it is the consideration of the evidence that typically produces the decision. In the area of science, we collect data to support or weaken support for a hypothesis. All scientists are taught that we do not *prove* hypotheses, or on a grander scale, that we do not prove theories. Instead, through the collection of data, we accumulate evidence for or against a particular hypothesis or theory. It is also understood within science that all data collection efforts are *inherently* flawed, and are expected to include some degree of error in the sampling process or in some other component of the scientific effort. Objective scientific efforts make strides to illustrate. address, and mitigate such weaknesses, though it is clearly accepted that perfection or flawlessness in science is impossible. For example, imagine that I hope to conduct a study using a sample of 1000 individuals. from which I will make conclusions about the opinions of а population of millions. I can work hard to make sure that my sample is representative of the population, in terms of demographic and socioeconomic ratios, but there is always a presumed rate of error, in spite of my diligence. To summarize, science can be conceptualized as the collection of data to be used as evidence to make a decision about a particular hypothesis, with the general acceptance that each and every scientific effort is flawed to some degree.

Next, let's move on to the context or goals of the research. Is the purpose of the research to accumulate evidence as it pertains a theoretical model of psychological to functioning? Is the purpose to provide evidence as it pertains to a real-world policy? These two possibilities help to delineate two ends of a continuum that can be used to categorize research. On one end of the continuum or spectrum is basic research. This work focuses on development and testing of theories, and to a degree with evaluating the basic assumptions of a variety of concepts and applications. This type of research is vital in that it advances our general knowledge

This article previously appeared in the 2008 American Association of Police Polygraphists (AAPP) Digest, and is reprinted here with the kind permission of the author and the AAPP.

base of a variety of scientific arenas. It helps us to understand how things work, what are the key underlying processes, and what are the critical relevant factors. Without basic research, we are left without a true understanding of why things happen.

On the other end of this spectrum is what we can call applied research. This is work that focuses on real world problems and issues, and may or may not integrate theoretical knowledge. The goal is often to develop evidence for a solution or approach that will be used in the real world.

It is important to understand that the complete continuum of research is vital to a particular body of knowledge. It is important to have a good grasp of the underlying principles and vital factors that relate to a particular problem, and also an understanding of how those elements will impact the real world.

Within the polygraph field, much of the research has focused on the applied end of the spectrum described here. That is, there is a sizeable body of work that focuses on polygraph applications, either using real world laboratory-based data. or with crime simulations. However, less work exists on the foundations of polygraph principles, such as why it works, what factors are critical, what are the key underlying processes, etc. These are largely reported in experiential fashion, or presented in the form of hypotheses that are rarely subject to scrutiny in the form of scientific evaluation. testable This is unfortunate, for there is a great deal of existing applied polygraph research (e.g., reporting accuracy rates, etc.), but little work in understanding the underlying factors that contribute to accuracy and the diagnostic value of the polygraph process.

Another issue that confronts polygraph research, and more importantly, the entire field of credibility assessment, is capturing the accuracy of a given technique in a real world environment. As we'll see, this problem represents an elusive 'Holy Grail' of sorts. There are two approaches that can be used to assess, establish, or investigate the accuracy rate of polygraph procedures; laboratory studies and field studies. Both of which, by the way, exist on the applied end of the continuum, as the goal of both types of study is to provide an estimate of polygraph accuracy under real world conditions.

Laboratory polygraph research can be broadly characterized as a situation where participants are (ideally) randomly assigned to participate in either 'deceptive' or 'truthful' roles during a given time period, and then subject to a polygraph examination to discern their role. For example, a subset of participants in a given study may be instructed to steal some money, and another subset of participants in the study will be assigned to be innocent, not committing any crime. The key strength of laboratory polygraph research is the fact that, within the confines and limitations of the study, accuracy rates can be clearly and precisely determined. In other words, because the exact role completed by each participant (commonly known as ground truth) is known by the researcher, the precise accuracy rate produced by the polygraph decision process can be determined.

The primary weakness of laboratory polygraph studies lies in concerns over external validity, or whether these simulated situations actually capture the real world circumstances in which polygraph is actually implemented. In real world polygraph situations, examinees are faced with an array of consequences, based on the outcome of the polygraph test. They may be subject to interrogation relevant to specific crimes, complications may be added to the employment process, and other negative reallife outcomes. In the laboratory context, while participants may be rewarded for truthful results on the polygraph test, there are really no punishments or negative consequences that can be exacted to capture the anxiety that exists in the real world.

Field polygraph research typically involves doing an archival collection of polygraph data from an agency, and aligning the polygraph results with their corresponding criminal case records. Generally speaking, those polygraph tests that are affiliated with criminal cases that have been resolved and a guilty individual identified via confession and ideally other corroborating evidence are considered to be 'confirmed' with respect to ground truth. Field polygraph research overcomes the primary criticism of laboratory polygraph research in that it captures polygraph as it exists in the real world; real polygraph examiners are testing real examinees, using real polygraph testing formats, with real consequences hanging in the balance.

However, field polygraph research suffers from an inability to establish solid ground truth. In other words, it is very difficult to establish whether a given examinee is *actually* truthful or deceptive in a real world polygraph examination. For example, if one or more polygraph examinations are administered relevant to a particular criminal case, and the case is never solved, ground truth is not likely to be established for any of the polygraph examinations. In addition, even if a confession is obtained following a polygraph examination, if no other evidence is produced to corroborate the confession, then it is difficult to presume ground truth with absolute certainty. Ultimately, the problem of establishing ground truth presents a formidable obstacle to establishing accuracy rates using polygraph data obtained in the field.

So we are left with what appears to be an impasse; these two sources represent our only possibilities for estimating polygraph accuracy under real world conditions, and each possesses a critical flaw. Some would argue that these flaws are fatal, in the sense that they render the research efforts useless. However, from a purely scientific perspective, this reasoning runs opposite to basic scientific principles. As suggested above, there is no perfect scientific theory, no study that has been undertaken that is completely free of flaws or random variance that could offer alternative explanations to the conclusions produced in the research effort. A key component of the scientific process, especially published research articles, is the in recognition of the flaws and weaknesses inherent in the study. As such, it is certainly important for us, as practitioners and researchers in the polygraph field to recognize the weaknesses inherent in our research tools. By doing so we can work together to improve our research efforts. In addition, it is important to recognize the polygraph research for what it is; evidence. Like research products in other areas of science, we should accept that polygraph research is certainly not perfect, and possesses problems that we should continue to mitigate and overcome. It is highly probable that polygraph research will never be able to truly capture the perfect combination of real world jeopardy and unquestionable ground truth. However, no scientific study has ever been rejected because it lacked perfection. In addition. to categorically reject polygraph research for these flaws, is inherently unscientific. In this respect, polygraph research should be viewed and considered in similar fashion to other scientific research; as evidence for the accuracy of polygraph processes under real world field conditions...or at least our best estimate, given the inherent limitations of the scientific tools with which we are equipped. Keep in mind that to clinch our eyes shut to the weakness inherent in polygraph research efforts would render us just as closed-minded as those who completely reject all polygraph research.

So, rewinding a bit, and hopefully bringing this thing full circle, it is vital for the polygraph community to recognize the significance of research efforts in our field. provides evidence Research for the effectiveness of a variety of polygraph testing formats. Research helps us to understand the basic principles impacting the effectiveness of the polygraph process. Research efforts will continue to provide evidence for continuing advances in our field. Ultimately and ideally polygraph research should provide the rationale behind why we follow the processes that we do in polygraph. Research provides an objective forum for the documentation of such efforts that will stand the test of time and provide a solid foundation for policy and practice for years to come.

We consider also must the consequences of not conducting research. Without research, we would truly know nothing about our polygraph procedures, beyond what our frail memories conveyed to us about past successes and failures. Without research, there would be a total severance of scientific ties with the and academic communities. Without research there would be a complete loss of legal credibility. Without research there would be no basis for polygraph standardization of testing procedures. When queried about reliability accuracy rates of polygraph and

examinations, the profession could only spout fabricated estimates. The polygraph profession would descend into turmoil amid a scattered mass of personal preferences and improperly formed beliefs regarding what constitutes an effective polygraph procedure.

One final point that is vital to drive home. Collectively, we stand as the greatest obstacles for the future success of polygraph as a field, and our greatest weaknesses are overconfidence and complacency. Research in our field gets outdated quickly, as practices, technologies, examinees and other aspects of the testing context continue to evolve and change. In the blink of an eye, scores of research studies can become irrelevant and inapplicable, as they may grow to represent procedures and techniques of a distant era. What's my point? Research must be ongoing, continual, and never ending – particularly in our field. Without constantly looking at and advancing ourselves through the scrutiny of research, we will surely fall by the wayside, as an outdated field that, in the distant past, enjoyed great success, but then grew lazy, complacent, and failed to stand the test of time.

My goal is not to terrify you, or cast darkness on what was an otherwise pleasant day, but rather to motivate you. There are hungry, impassioned individuals in our field who are working tirelessly, each day, often deep into the night, to move us forward, to learn more about what we do and why it seems to work, exploring both the basic and applied ends of the research continuum. Seek these people out. Feed off of them. Work with them. Contribute to our collective efforts. Read the research. Understand where we've been so that we can continue to advance. We need as much evidence as we can get our hands on.