REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*
July 2000

2. REPORT TYPE
Final Report

3. DATES COVERED *(From - To)*
December 1996 - March 2000

4. TITLE AND SUBTITLE
Comparison of Utah and DoDPI Scoring Accuracy: Equating Veracity Decision Rule, Chart Rule, and Number of Data Channels Used

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Stuart M Senter, Ph.D., Andrew B. Dollins, Ph.D., and Donald J. Krapohl, M.A.

5d. PROJECT NUMBER
DoDPI97-P-0005

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DoD Polygraph Insituste
7540 Pickens Avenue
Fort Jackson, SC 29207

8. PERFORMING ORGANIZATION REPORT NUMBER
DoDPI00-R-0001

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DoD Polygraph Insititute
7540 Pickens Avenue
Fort Jackson, SC 29207

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Public release, distribution unlimited

20000925 155

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The performance of scorers using the University of Utah and Department of Defense Polygraph Institute (DoDPI) physiological detection of deception chart evaluation rules were compared to discover if differences in laboratory-based decision accuracy rates are due to chart evaluation rules. Four scorers (two based at the DoDPI, two based at the University of Utah) evaluated the charts from 100 polygraph examinations (50 deceptive, 50 nondeceptive). We attempted to isolate scorer ability by equating the rules for making veracity decisions, number of charts used, and number of data channels considered. There was no evidence, when these variables were held constant, that scorers differed on the proportion of correct, incorrect, or no opinion decisions rendered. Results suggest no differences in chart scoring ability among scorers based at the two institutions. Observed differences in accuracies for Utah and DoDPI scoring systems may be due to differences in veracity decision rules, the number of charts evaluated, the inclusion of the photo-plethysmograph data channel, or a combination of these factors. The greatest accuracy was obtained by all scorers using the University of Utah chart evaluation rules.

15. SUBJECT TERMS
polygraph, chart evaluation, veracity, decision accuracy, psychophysiological detection of deception

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Andrew H. Ryan, Jr, Ph.D.
Unclassified	Unclassified	Unclassified	UL	23	19b. TELEPHONE NUMBER *(Include area code)* (803) 751-5867

DTIC QUALITY INSPECTED 4

DoDPI00-R-0001

Comparison of Utah and DoDPI Scoring Accuracy: Equating Veracity Decision Rule, Chart Rule, and Number of Data Channels Used

Stuart M. Senter, Ph.D.
Andrew B. Dollins, Ph.D.
Donald J. Krapohl, M.A.

July 2000

Department of Defense Polygraph Institute Research Division
7540 Pickens Street
Fort Jackson, South Carolina, 29207

Director's Foreword

DoDPI has an abiding interest in developing and testing scoring and decision rules in the field of psychophysiological detection of deception (PDD), with the objective of disseminating best practices to field examiners in the government. As part of this continuing venture, the present study evaluated the DoDPI rules and the University of Utah rules on PDD recording collected in a laboratory study. Within the context of analog study data, the results suggest that the systems do not perform equally. However, the data only indicate that there were differences in performance, though the ultimate source of the differences will require more focused further study. With the present findings as a foundation, other investigations can be undertaken to ultimately determine the optimal scoring and decision rules.

Michael H. Capps
Director

Acknowledgments

The authors wish to thank the scorers who contributed the data for this study: William Gary, Shirley Sturm, Sheila Thomas, John Kircher, and David Raskin. We would also like to thank Ms. Charlene Stephens for assisting with the data collection. Funds for this research were provided by the Department of Defense Polygraph Institute under project number DoDPI97-P-0005. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Defense or the U. S. Government.

Abstract

SENTER, S. M., DOLLINS, A. B., and KRAPOHL, D. J. Comparison of Utah and DoDPI scoring accuracy: Equating veracity decision rule, chart rule, and number of data channels used. July, 2000, Protocol No. DoDPI00-R-0001. Department of Defense Polygraph Institute, Fort Jackson, SC 29207--The performance of scorers using the University of Utah and Department of Defense Polygraph Institute (DoDPI) physiological detection of deception chart evaluation rules were compared to discover if differences in laboratory-based decision accuracy rates are due to chart evaluation rules. Four scorers (two based at the DoDPI, two based at the University of Utah) evaluated the charts from 100 polygraph examinations (50 deceptive, 50 nondeceptive). We attempted to isolate scorer ability by equating the rules for making veracity decisions, number of charts used, and number of data channels considered. There was no evidence, when these variables were held constant, that scorers differed on the proportion of correct, incorrect, or no opinion decisions rendered. Results suggest no differences in chart scoring ability among scorers based at the two institutions. Observed differences in accuracies for Utah and DoDPI scoring systems may be due to differences in veracity decision rules, the number of charts evaluated, the inclusion of the photo-plethysmograph data channel, or a combination of these factors. The greatest accuracy was obtained by all scorers using the University of Utah chart evaluation rules.

Keywords: polygraph, chart evaluation, veracity, decision accuracy, psychophysiological detection of deception

Table of Contents

Clear accuracy discrepancies have been reported by scorers at the University of Utah and the Department of Defense Polygraph Institute (DoDPI) when evaluating psychophysiological detection of deception (PDD) examinations. Table 1 summarizes the accuracy of laboratory-based specific issue studies conducted at the DoDPI and the University of Utah over the last 22 years. Total accuracy does not represent the average of DI and NDI accuracy as different numbers of no opinion (NO) decisions across the two categories often produced different sample sizes in each category. Thus, the total accuracy for each study represents the weighted mean of the DI and NDI accuracies. Excluding NO decisions, the weighted mean accuracy of the DoDPI scorers ($\underline{M}_{\text{DoDPI}}$ = 79.3%) is much lower than that calculated for affiliates of the University of Utah ($\underline{M}_{\text{Utah}}$ = 91.8%). The same trend exists when NO decisions are included ($\underline{M}_{\text{DoDPI}}$ = 64.1% versus $\underline{M}_{\text{Utah}}$ = 79.3%).

Table 1
Accuracy Rates and Sample Sizes for Specific Issue Examinations conducted by the DoDPI and Utah University

		N		Accuracy(%)			
Study	Affiliation	D	ND	DI	NDI	Tot	TotNO
Barland & Honts (1990)	DP	20	20	74	78	76	70
Blackwell (1994)	DP	60	60	86	73	80	72
Honts et al (1989)	DP	40	20	83	83	83	78
Honts & Barland (1990)	DP	44	44	91	62	79	64
Honts (1992)	DP	72	79	58	90	75	53
Ingram (1996a)	DP	10	10	92	75	84	53
Ingram (1996b)	DP	15	15	97	83	92	60
Honts et al (1987)	UT	10	10	100	78	88	75
Honts et al (1994)	UT	20	20	78	88	83	73
Kircher & Raskin (1988)	UT	74	74	97	95	96	87
Podlesny & McGhehee (1987)	UT	72	24	74	90	87	78
Podlesny & Truslow (1991)	UT	48	24	92	84	90	73
Podlesny & Truslow (1993)	UT	72	24	94	64	89	73
Rovner (1986)	UT	12	12	100	90	96	88
Raskin & Hare (1978)	UT	24	24	91	100	96	88
Raskin et al (1988)	UT	69	36	95	96	95	77

Note. DP = Department of Defense Polygraph Institute, UT = Utah University, D = Deceptive, ND = Nondeceptive, DI = Deception Indicated, NDI = No Deception Indicated, Tot = accuracy excluding NO decisions, TotNO = accuracy including NO decisions as incorrect.

Among the factors that could cause the discrepancy in accuracy rates obtained by scorers at the University of Utah and the DoDPI are: participant characteristics, participant manipulation methods, physiological tracing quality, scoring system, chart evaluation rules, and efficiency of applying scoring system and evaluation rules. In order to make the task manageable, the scope of this study was limited to determining whether differences exist in the abilities of scorers trained at the two institutes to evaluate the physiological data when applying the DoDPI or University of Utah scoring systems. In addition, the accuracy of the chart evaluation rules used by the two institutions will be compared. For the purposes of this paper, scoring system will be defined as the rules used to assign numbers corresponding to reactions in the physiological tracings. Also, for the purposes of this paper, chart evaluation rules are defined as encompassing veracity decision rule, chart usage rule, and number of data channels used. These elements are described in subsequent sections.

Scoring Systems

During a PDD examination, the participant is asked a series of questions while physiological reactions are recorded digitally or on paper charts. The questions are usually categorized as irrelevant (e.g., "Is today Thursday?"), comparison (e.g., "Before the age of 18, did you ever take anything of value from someone who trusted you?") or relevant (e.g., "Did you steal that money from the bank?"). Test format refers to question syntax, the number of questions, their presentation order, and the number of times each question is presented.

Most examiners measure thoracic and abdominal respiration, electrodermal activity using either resistance or conductance, and cardiovascular activity using a blood pressure cuff (the auscultatory cuff method). Investigators at the University of Utah typically measure cardiovascular activity using photoplethysmography, in addition to the blood pressure cuff.

When the physiological data have been recorded they are evaluated manually or by computer. During manual evaluation of PDD examinations that use comparison questions, examiners compare the reaction following a comparison question to that following a relevant question for each physiological channel. A comparison and relevant question pair are typically presented at least three times during a PDD examination. A score indicating the size of the difference for each pair of reactions is assigned for each of the three question pairs and each physiological channel. The assigned scores are between -3 and +3 or between -1 and +1 (inclusive), depending on whether the 7- or 3-position scale is used. For descriptions of the scoring systems various organizations use to evaluate physiological data, refer to Swinford (1999), Bell, Raskin, Honts, and Kircher (1999), and the Federal Psychophysiological Detection of Deception Handbook (1999).

Veracity Decision Rules

Decisions regarding participant veracity are made using the assigned scores. Both the DoDPI and University of Utah use total score cutoff criteria. That is, the scores assigned to each pair of reactions and physiological channel are added together and a decision of deception indicated (DI), no deception indicated (NDI), or no opinion (NO) is made, depending on the total. If the assigned total is -6 or less the decision is DI, if the total is +6 or greater then the decision is NDI, if the total is between -6 and +6 then the decision is NO (Bell et al., 1999; Federal Psychophysiological Detection of Deception Handbook, 1999; Swinford, 1999).

The DoDPI also uses a "spot score" rule (Federal Psychophysiological Detection of Deception Handbook, 1999). All of the scores assigned to a comparison and relevant question pair are summed over repeated presentations of the question pair, producing a spot score. Because the DoDPI uses three relevant questions, a typical examination will have three spot scores. The DoDPI spot score rule dictates that a participant must have a +1 or greater on all spot scores and a total score of +6 or greater to be classified as NDI. A participant with a -3 or less in any spot score or a total of -6 or less is classified as DI. Examinations that do not meet either the DI or NDI criteria are assigned a decision of NO.

Chart Usage Rules

The DoDPI teaches that three charts should be recorded during specific issue examinations. A fourth chart may be recorded if a question in the earlier series cannot be evaluated. The three chart rule may have been adopted because it was believed that data produced after three question series was less diagnostic or useful due to habituation (Balloun & Holmes, 1979; Suzuki & Hikita, 1964). However, recent work has shown that the strength and diagnosticity of the data signals do not degrade with additional presentations (Dollins, Cestaro, & Pettit, 1998; Elaad & Ben-Shakhar, 1997; Nakayama & Kizaki, 1990; Yankee & Grimsley, 1987). The Utah approach uses either three or five charts. If a decision of deceptive or nondeceptive is assigned after the first three charts, then only three charts are used. If a NO decision is reached after the first three charts, two additional charts are evaluated and a decision is made using five chart totals. The +6 and -6 cutoff scores are used whether 3 or 5 charts are scored.

Inclusion or Exclusion of Photo-Plethsymograph Data Channel

The DoDPI approach uses three data channels; respiration, cardiograph, and skin conductance. The Utah approach uses these three channels plus a measure of peripheral vasoconstriction in the index finger garnered with a photo-plethysmograph. Research has shown that this channel represents a useful predictor of participant veracity, though inferior in terms of diagnosticity to one or more of the other data channels (Cutrow, Parks, Lucas, & Thomas, 1972; Podlesny, Raskin, & Barland, 1976; Suzuki, 1965; Thackray & Orne, 1968).

Method

Lab data collected from 50 deceptive and 50 nondeceptive participants (Kircher & Raskin, 1988) were evaluated by five different scorers. Five charts were collected from each participant and the scorers evaluated each of the charts a single time by completing a score sheet (Appendix A). Two of the scorers used the University of Utah scoring system (Bell et al., 1999), two used the DoDPI scoring system (Swinford, 1999), and one used the Backster scoring system (The Backster School of Lie Detection, San Diego, CA). The scores collected from the two scorers using the University of Utah scoring system were those described by Kircher and Raskin. It is noted that only one of the two scorers was originally trained using the Utah scoring system. The other Utah scorer, while originally trained at the Backster School, was one of the developers of the Utah scoring system. Thus, we considered both of these scorers to be representative of those who use the Utah scoring system. Each scorer completed a score sheet by assigning a numerical value between -3 and +3, inclusive, to each data channel (i.e., respiration, skin conductance, cardiograph, and photo-plethysmograph), for each relevant question. The instructions given to the scorers are included in Appendix B.

The numerical scores collected from each of the 5 scorers were then totaled using the University of Utah chart evaluation rules (+6 and -6 cutoffs, 3 or 5 charts, and inclusion of photo-plethysmograph data channel), then the DoDPI chart evaluation rules (spot scores, 3 charts only, and exclusion of photo-plethysmograph data channel). The data for the scorer who was trained at the Backster School of Lie Detection were omitted from analysis because the Backster scoring procedures differ from those taught by both the University of Utah and the DoDPI. The data for the scorer using the Backster scoring system are included for informational purposes.

Chi-square analyses were calculated for each set of comparisons, and used as a global test of differences. If any individual chi-square result was significant, Cochran's Q (Siegel & Castellan, 1988) analyses were calculated to test for differences in the proportion of correct, incorrect, and NO decisions as a function of scorer. The Cochran's Q is a nonparametric test for use with related or repeated measure categorical binomial data. This test is used to detect differences in response proportions across conditions or coders (scorers in this case). Because there were three categories (correct, incorrect, NO) and the Cochran's Q requires binomial data, the two categories that were not the focus of a given test were collapsed together to create a binary data set. The significance criterion for the Chi-square test was set at .05. However, a Bonferroni correction was used when sets of Cochran's tests were calculated because each test was calculated on interrelated frequency proportions. This correction decreases the significance level by dividing it by the number of overlapping tests (Keppel, 1991). Therefore, the significance level for the Cochran's tests was set at .017 (.05 divided by 3).

The first analysis compared proportions of correct, incorrect, and inconclusive decisions calculated for scorers trained at the University of Utah using the University of Utah chart evaluation rules with those calculated for scorers trained at the DoDPI using the DoDPI chart evaluation rules (typical). The second analysis compared accuracy proportions of scorers trained at the University of Utah using the DoDPI chart evaluation rules with those calculated for DoDPI scorers using the University of Utah chart evaluation rules (reversed). The third analysis compared proportions between the two groups of scorers, all using the DoDPI chart evaluation rules (DoDPI scoring rules). The fourth analysis compared proportions between the two groups of scorers, all using the University of Utah chart evaluation rules (University of Utah scoring rules).

Results

Typical Scoring

Table 2 shows the frequencies of scorers' correct, incorrect, and NO decisions as a function of participant veracity, calculated according to the DoDPI and Utah chart evaluation rules. Correct decision (accuracy) rates for each scorer were produced by summing the number of DI decisions for deceptive participants and the number of NDI decisions for nondeceptive participants. Incorrect decision rates were produced by summing the number of NDI decisions for deceptive participants and the number of DI decisions for nondeceptive participants. Inconclusive rates were calculated by summing the number of inconclusive decisions for both deceptive and nondeceptive participants.

Table 2
Frequency of Correct, Incorrect, and No Opinion Decisions as a Function of Scorer Calculated Using the DoDPI Chart Evaluation Rules and the Utah Chart Evaluation Rules

	Deceptive (N=50)			Nondeceptive (N=50)		
Scorer	Correct	Error	NO	Correct	Error	NO
			DoDPI Rules			
DP-1	42	1	7	26	10	14
DP-2	43	3	4	30	7	13
UT-1	36	1	13	26	4	20
UT-2	37	2	11	26	3	21
BK*	38	3	9	33	3	14
			Utah Rules			
DP-1	39	6	5	41	3	6
DP-2	41	5	4	42	4	4
UT-1	39	2	9	41	2	7
UT-2	44	3	3	43	3	4
BK*	40	4	6	41	3	6

Note. NO = No Opinion Decision. DP = DoDPI affiliated scorer, UT = Utah affiliated scorer. BK = Backster affiliated scorer. * = data not included in analyses.

Consistent with previous research, the mean accuracy (with NO decisions excluded) for the Utah scorers (94.4%) following the Utah chart evaluation rules was higher than that for the DoDPI scorers (87.0%) following the DoDPI chart evaluation rules. If NO decisions are included, the mean accuracy for the Utah scorers (83.5%) was only slightly higher than that for the DoDPI scorers (81.5%).

The results of the chi-square analysis conducted on decision frequencies resulted in a significant difference as a function of scorer, $\underline{X}^2(6) = 13.77$, $\underline{p} < .05$. Data from the four scorers were then analyzed with three separate Cochran's Q tests. The Cochran's test for the proportion of correct decisions was significant, $\underline{CQ}(3) = 19.9$, $\underline{p} < .017$, suggesting differences among the proportions of correct decisions of the four scorers. Because the proportions of correct decisions by DoDPI scorers (.68 and .73) showed no overlap with the proportion of correct decisions by Utah scorers (.80 and .87), we infer that this difference was largely attributable to the difference in the proportions of correct decisions between DoDPI and Utah scorers. The test for the proportion of errors calculated for the four scorers was not significant, $\underline{CQ}(3) = 7.42$, $\underline{p} > .017$. The test for the proportion of NO decisions calculated for the four scorers was also not significant, $\underline{CQ}(3) = 9.74$, $\underline{p} > .017$.

Reversed Scoring

Under reversed scoring, the mean accuracy calculated for the two groups of scorers drew closer together. When scored according to the DoDPI chart evaluation rules, accuracy with NO decisions excluded for the scorers trained at the University of Utah dropped slightly in comparison to when scored according to Utah chart evaluation rules (92.6% vs. 94.4%). A large decrease in mean accuracy occurred for the Utah scorers when NO decisions were included (62.5% vs. 83.5%). Conversely, accuracy with NO decisions excluded for the scorers trained at the DoDPI showed a slight increase with Utah chart evaluation rules relative to when scores were combined according to the DoDPI chart evaluation rules (90.1% vs. 87.0%). However, a large decrease in mean accuracy for the DoDPI scorers also occurred when NO decisions were included (70.5% vs. 81.5%).

The chi-square conducted on these frequencies indicated significant differences across the proportion of decisions as a function of scorer, $\underline{X}^2(6) = 32.8$, $\underline{p} < .05$. The Cochran's test indicated strong differences in the frequency of correct decisions across scorers, $\underline{CQ}(3) = 54.9$, $\underline{p} < .017$. The proportions of accurate decisions for the two pairs of scorers showed a dramatic crossover. The effect detected by the Cochran's test appears to be largely attributable to the comparison of the proportion of correct calls calculated for the

DoDPI scorers (.80 and .83) and those calculated for the Utah scorers (.62 and .63).

The Cochran's test for the frequency of errors was not significant, CQ(3) = 2.91, p > .017. The Cochran's test conducted on the frequency of NO decisions was significant, CQ(3) = 65.4, p < .017, and reflected a clustering of inconclusive proportions between the two pairs of scorers. The proportion of NO decisions calculated for the DoDPI scorers was small (.11 and .08) relative to that calculated for Utah scorers (.33 and .32).

Both Scored Using DoDPI Rules

Analyses in this section used the decision frequency data with evaluations equated using the DoDPI chart evaluation rules. The chi-square results were not significant, $X^2(6)$ =12.5, p > .05, suggesting no differences in the proportion of decision frequency as a function of scorer. Thus, no subsequent tests were conducted.

Both Scored Using University of Utah Rules

Analyses in this section used the decision frequency data with evaluations equated using the University of Utah chart evaluation rules. The chi-square test calculated on the decision frequencies as a function of scorer was not significant, $X^2(6) = 7.64$, p > .05.

Proportion of Agreement

To provide a further comparison of results calculated when using the DoDPI chart evaluation rules and the Utah chart evaluation rules, the reliability across scorer using both types of chart evaluation rule was examined. Table 3 shows the proportion of agreement among scorers when decisions were derived using the DoDPI chart evaluation rules and the Utah chart evaluation rules. This is the proportion of time that the same decision (DI, NDI, and NO) was obtained for both scorers for each participant's chart. Thus, higher values indicate a greater degree of reliability across scorers. Comparing the proportions of agreement calculated using the two chart evaluation rules shows that the average proportion of agreement using the DoDPI chart evaluation rules (.728) is lower than that using the Utah chart evaluation rules (.838).

Discussion

For the typical and reversed scoring analyses, the frequency of correct decisions calculated for the pair of scorers using the Utah chart evaluation rules was significantly greater that that calculated for the pair of scorers using the DoDPI chart evaluation rules. However, when the frequency of correct decisions for the two pairs of scorers was generated using the same chart evaluation rules, no significant differences arose between the two pairs of scorers. Thus, the lower decision accuracy of the DoDPI scorers was due the different chart evaluation rules, and not to a difference in ability to code physiological tracings produced in the PDD examinations.

Table 3

Proportion of Agreement between Scorers Using the DoDPI Chart Evaluation Rules and the Utah Chart Evaluation Rules

	DoDPI Rules Scorer				Utah Rules Scorer			
Scorer	DP-1	DP-2	UT-1	UT-2	DP-1	DP-2	UT-1	UT-2
DP-2	.66				.75			
UT-1	.78	.70			.94	.78		
UT-2	.72	.67	.75		.86	.79	.84	
BK	.66	.68	.77	.89	.88	.80	.90	.84

Note. DP = DoDPI affiliated scorer, UT = Utah affiliated scorer. BK = Backster affiliated scorer.

The results of the Cochran's Q tests may have come out differently had the significance criterion been set at .05 rather than .017. While our decision to adopt the reduced significance criterion does increase the possibility of committing a Type II error (basically, categorizing a finding as unreliable when it is actually reliable), the possibility of committing a Type I error (categorizing a finding as reliable when it is actually unreliable) due to the linear dependency of the decision categories is reduced. It is our belief that the commission of a Type II error is less of a concern than the commission of a Type I error. In a similar vein, we note that the values of the chi-square tests were similar for the typical scoring analyses and that where both groups were scored using DoDPI rules. However, one value was just above the significance criterion and the other just below. Therefore, even though the two chi-square tests produced results of similar magnitude, one was judged as significant and the other was not. A more liberal interpretation of this result might classify the results of both chi-square tests to be statistically significant. However, we have chosen to adhere to the .05 criterion, again to avoid the commission of a Type I error. Thus, we have taken a conservative position in evaluating the results of this study, primarily to avoid commission of Type I errors.

The proportion of NO decisions calculated for the two groups of scorers was significantly different when proportions for the Utah group were derived using the DoDPI chart evaluation rules. This effect may have occurred because the DoDPI chart evaluation rules only use data from three charts, while the Utah chart evaluation rules use three charts only if a decision of deceptive or nondeceptive has been reached. If no decision has been reached, two additional charts are evaluated, for a total of five charts. Thus, scorers trained at the University of Utah may have a tendency to postpone conclusive decisions after only three charts have been examined, with the expectation that additional charts will be examined, providing stronger evidence for a conclusive decision. This tendency to suspend judgment after three charts could also explain the significant increase in the proportion of inconclusive decisions calculated for Utah scorers relative to DoDPI scorers under the DoDPI chart evaluation rules.

Overall, the combined results from the four sets of analyses suggest one very clear implication. The differences that exist in laboratory-based evaluation accuracy between DoDPI and the University of Utah appear to be attributable to numerical combination differences following the scoring of the physiological data, the inclusion or exclusion of data from the photo-plethysmograph, and to the number of charts used in evaluation, and not due to differences in the scoring systems used between the Utah and DoDPI scorers. The isolation of which of the three factors contributes the greatest amount to the observed difference in performance is unclear at this juncture, and remains a topic of ongoing investigation.

The magnitude of the accuracy difference between the two groups of scorers in the current study is less than that produced by the weighted means of the studies displayed in Table 1. This casts some concern upon the generalizability of the results of the current study to previous results. However, the findings that accuracy rates improved for DoDPI scorers and that proportional differences in decision types were eliminated when evaluations for all scorers were coded using the same chart evaluation rules does suggest that the different chart evaluation rules used by the two institutions contribute to the historical difference in laboratory-based evaluation accuracy.

In conclusion, the results suggest that a number of shifts in chart evaluation protocol may improve laboratory-based decision performance produced with the DoDPI chart evaluation rules, ideally removing the historical disadvantage in decision accuracy relative to that of affiliates of the University of Utah. Furthermore, the examination of scorer agreement using the two chart evaluation rules indicates that shifting to the Utah chart evaluation rules may improve reliability across scorers. Because our results indicate that these performance differences are not primarily due to pure evaluation ability, further research is necessary to determine the contributions attributable to the different veracity decision rules, number of charts used, and use of the photo-plethysmograph.

References

Balloun, K. D., & Holmes, D. S. (1979). Effects of repeated examinations on the ability to detect guilt with a polygraphic examination: A laboratory experiment with a real crime. Journal of Applied Psychology, 64, 316-322.

Barland, G. H., & Honts, C. R. (1990). A laboratory study of the validity of the ZOC: An executive summary (DoDPI89-P-0001). Fort McClellan, AL: Department of Defense Polygraph Institute.

Bell, B. G., Raskin, D. C., Honts, C. R., & Kircher, J. C. (1999). The Utah numerical scoring system. Polygraph, 28, 1-9.

Blackwell, N. J. (1994). An evaluation of the effectiveness of the Polygraph Automated Scoring System (PASS) in detecting deception in a mock crime analog study (DoDPI94-R-0003). Fort McClellan, AL: Department of Defense Polygraph Institute.

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

Dollins, A. B., Cestaro, V. L., & Pettit, D. J. (1998). Efficacy of repeated physiological detection of deception testing. Journal of Forensic Science, 43, 1016-1023.

Elaad, E., & Ben-Shakhar, G. (1997). Effects of item repetitions and variations on the efficiency of the guilty knowledge test. Psychophysiology, 34, 587-596.

Federal Psychophysiological Detection of Deception Examiner Handbook. (1999, May). Fort Jackson, SC: Department of Defense Polygraph Institute.

Honts, C. R. (1992). Counterintelligence scope polygraph (CSP) test found to be a poor discriminator. Forensic Reports, 5, 215-218.

Honts, C. R., & Barland, G. H. (1990, Jan). A laboratory study of the validity of the MGQT: An executive summary (DoDPI88-P-0002). Fort McClellan, AL: Department of Defense Polygraph Institute.

Honts, C. R., Barland., G. H., & Barger, S. D. (1989). The relative validity of criminal and screening approaches to the detection of deception [Abstract]. Psychophysiology, 26(Suppl. 4A), 533.

Honts, C. R., Raskin, D. C., & Kircher, J. C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. Journal of Psychophysiology, 1, 241-247.

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

Ingram, E. M. (1996a). Test of a mock theft scenario for use in the psychophysiological detection of deception: I (DoDPI96-R-0003). Fort McClellan, AL: Department of Defense Polygraph Institute.

Ingram, E. M. (1996b). Test of a mock theft scenario for use in the psychophysiological detection of deception: III (DoDPI97-R-0003). Fort McClellan, AL: Department of Defense Polygraph Institute.

Keppel, G. (1991). Design and analysis: A researcher's handbook. Englewood Cliffs, NJ: Prentice Hall.

Kircher, J. C., & Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. Journal of Applied Psychology, 73(2), 291-302.

Nakayama, M., & Kizaki, H. (1990). Usefulness of the repeated presentation of questions on the psychophysiological detection of deception. The Japanese Journal of Psychology, 60 (6), 390-393.

Podlesny, J. A., & McGehee, C. M. (1987). Investigative detection of deception: Role discrimination with general question technique [Abstract]. Psychophysiology, 24(5), 605.

Podlesny, J. A., Raskin, D. C., & Barland, G. H. (1976). Effectiveness of techniques and physiological measures in the detection of deception (Report No. 76-5). National Institute of Law Enforcement and Justice, Law Enforcement Assistance Administration, U.S. Department of Justice (Contract No. 75-NI-99-0001) University of Utah.

Podlesny, J. A., & Truslow, C. M. (1991). Simulated crime role classification with an expanded issue control question technique [Abstract]. Psychophysiology, 28(6), S44.

Podlesny, J. A., & Truslow, C. M. (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. Journal of Applied Psychology, 78(5), 788-797.

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

Raskin, D. C., Kircher, J. C., Honts, C. R., & Horowitz, S. W. (1988). Validity of control question polygraph tests in criminal investigation [Abstract]. Psychophysiology, 25, 476.

Rovner, L. I. (1986). The accuracy of physiological detection of deception for subjects with prior knowledge. Polygraph, 15(1), 1-39.

Siegel, S., & Castellan, N. J. (1988). Nonparametric statistics for the behavioral sciences (2nd ed.). New York: McGraw-Hill.

Suzuki, A. (1965). The plethysmograph as an index of the lie detector. Reports of the National Research Institute of Police Science, 18, 288-293.

Suzuki, A., & Hikita, Y. (1964). An analysis of response on polygraph: A diminution of responses. Polygraph, 10(1), 1-7.

Swinford, J. (1999). Manually scoring polygraph charts utilizing the seven-position numerical analysis scale at the Department of Defense Polygraph Institute. Polygraph, 28, 10-27.

Thackray, R. I., & Orne, M. T. (1968). A comparison of physiological indices in detection of deception. Psychophysiology, 4, 329-339.

Yankee, W. J., & Grimsley, D. L. (1987). The effect of a prior polygraph test on subsequent polygraph tests [Abstract]. Psychophysiology, 24, 621-622.

Appendix A

Score Sheet

Subject: __________ **Date:** __________ **Examiner:** __________

Decision __________ **Reviewed by** ______ **Date:** __________

Comments:

Chart 1	5	7	10		
Upper Pneumo					
Lower Pneumo					
Electrodermal					
Cardiovascular					
Plethysmograph					

Chart 2	5	7	10		
Upper Pneumo					
Lower Pneumo					
Electrodermal					
Cardiovascular					
Plethysmograph					

Chart 3	5	7	10		
Upper Pneumo					
Lower Pneumo					
Electrodermal					
Cardiovascular					
Plethysmograph					

Chart 4	5	7	10		
Upper Pneumo					
Lower Pneumo					
Electrodermal					
Cardiovascular					
Plethysmograph					

Chart 5	5	7	10		
Upper Pneumo					
Lower Pneumo					
Electrodermal					
Cardiovascular					
Plethysmograph					

Upper Pneumo __________
Lower Pneumo __________
Electrodermal __________
Cardiovascular __________
Plethysmograph __________

Appendix B

Instructions to Scorers

1. Investigators at the University of Utah (Drs. David Raskin & John Kircher) report studies in the literature which indicate very high accuracy rates using laboratory Mock Crime Scenarios. Investigators at the Department of Defense Polygraph Institute and elsewhere have been unable to consistently duplicate these high accuracy rates. There have, thus, been some questions regarding methodological differences among the various research groups.

2. We are attempting to open some lines of communication between Dr. John Kircher at the University of Utah and the Institute to determine what the methodological differences are. Dr. Kircher has been kind enough to send us original charts of 100 examinations reported by Kircher and Raskin (1988). We would like for you to score each of these charts, using the attached guidelines, so we can determine if there are differences among the DoDPI and University of Utah scoring procedures.

3. These are not the usual Lafayette or Stoelting charts, but were collected using a Beckman laboratory-grade polygraph. Each chart page is divided by an easily-torn perforation. In addition, the charts are at least fifteen years old, so they are very fragile. The charts are the personal property of Dr. Kircher and he has asked that they be returned. I have given him my personal assurance that I would take every measure to ensure that the charts are returned in the same condition as they were loaned. Poor handling of these charts will, in addition to unnecessarily damaging someone else's property, cause the loss of a valuable collaborator and PDD supporter. PLEASE exercise the utmost care when handling the charts.

4. Ms. Charlene Stephens (Building 3195 Room I-106, ex-4297) will organize this project. Please contact her to obtain charts and score sheets--and return them to her. There are 100 examinations. I have established a suspense date, in consultation with Mr. Broadwell, of 16 December 1997. You will only be given a few charts at a time and everyone must score and return every chart. Be sure to allow for complications and delays when scheduling the time needed to complete this project. Please be courteous to others scoring the charts and don't keep the charts if you are not actively scoring them. If a chart is accidentally torn, please ask Ms. Stephens to assist you in repairing the damage. Finally, I ask that you not talk to others regarding specific charts.

5. Please refer any questions regarding scoring to me. The instructions for scoring the charts are attached.

Thank you.

Each examination is composed of five separate charts. Each chart contains 10 tracings. From top to bottom the tracings are:

1. Stimulus Marker
2. Thoracic (Upper) respiratory activity
3. Abdominal (Lower) respiratory activity
4. Electrodermal (Skin Conductance)
5. Cardiovascular Cuff
6. Finger pulse amplitude
7. Finger pulse volume
8. Cardiotachometer
9. Electrocardiogram (EKG)
10. Stimulus Marker

The questions asked were:

1. (Buffer) Do you understand that I will ask only the questions that we have discussed?

2. (Buffer) Regarding the theft of the ring and your basic honesty, do you intend to answer all of the questions truthfully?

A. (Neutral) Is your first name Richard?

4. (Control) During the first 24 years of your life, did you ever take something that didn't belong to you?

5. (Relevant) Did you take that ring?

B. (Neutral) Do you live in the United States?

6. (Control) Between the ages of 10 and 24, did you ever do something dishonest or illegal?

7. (Relevant) Did you take that ring from the desk?

C. (Neutral) Were you born in the month of January?

9. (Control) Other than what you told me, prior to 1981 did you ever deceive someone?

10. (Relevant) Do you have that ring with you now?

Use an accompanying score sheet for each examination. Complete the blanks for Subject number and examination Date on the first line of the score sheet using information from the charts. Complete the Reviewed by and Date blanks on the second line with your last name and the current date. Leave the spaces labeled Examiner and Decision blank.

Assign a separate numerical score between -3 and +3 for each channel of every relevant question on all charts (i.e., 5, 7, and 10). NOTE that all pens on the Beckman polygraph are the same length. The channels to be scored include:

1. Upper and Lower Pneumograph (decreases in the amplitude and rate of respirations and increases in respiration baseline)
2. Skin Conductance Response (SCR - amplitude, duration, and complexity of the SCR)
3. Cardiovascular Cuff (increases in blood pressure)
4. Finger Pulse Amplitude and Finger Blood Volume (decreases in both are considered to be responses)

Score the response to each Relevant question relative to the response to the preceding Control Question. Assign a negative value when the Relevant question produces the stronger response and a positive value when the Control question produces the stronger response. Assign a 0 when the reactions to the Relevant and Control questions are of similar strength. Do not score the 1st (Stimulation/Peak-of-Tension) chart, but score the subsequent five charts. Do not attempt to make a decision regarding the subjects veracity. Include any and all comments you care to make on the score sheet for that chart.