Survey of Reliability and Criterion Validity of Backster Numerical Scores of You-Phase Exams from Confirmed Field Investigations

Raymond Nelson, Mark Handler, Greg Adams, and Cleve Backster

Abstract

A cohort of seven scorers, all trained at the Backster School of Lie Detection and familiar with the use of the Backster numerical scoring system, provided blind scores for a sample of confirmed You-Phase examinations. Criterion accuracy was significantly greater than chance with a moderate to high pairwise correlation between the numerical scores of the participants. A dimensional profile of criterion accuracy is shown. Results of this study support the validity of the Backster numerical scoring system for You-Phase exams.

Introduction

The You-Phase technique is an eventspecific single-issue comparison question format for psychophysiological detection of deception (PDD) examinations, first described by Cleve Backster (1963), as a modification of the comparison question technique (CQT) first described by Reid (1947). A family of two- and three- question Zone Comparison Techniques (ZCTs) has evolved from the basic structure of the Backster You-Phase technique, including both generic (Department of Defense, 2006) and boutique (Gordon, Mohamed, Faro, Platek, Ahmad, & Williams, 2005; Matte, 1978; Matte & Reuss, 1989) variants of the ZCT.

Two versions of the You-Phase technique exist today: the US Federal You-Phase format, taught by the US Department of Defense (2006), and the version originally developed by Backster (1963). This technique was taught by Cleve Backster at the United States Army Military Police Polygraph School (USAMPS) at Fort Gordon, GA and Fort McClellan, AL. These two versions currently differ in their scoring features, transformation rules, decision rules, and cutscores. Both versions of the You-Phase technique are

Raymond Nelson is a research specialist with the Lafayette Instrument Company (LIC) and an elected member of the American Polygraph Association (APA) Board of Directors. The views expressed in this work are those of the author and not the LIC or the APA. Mr. Nelson is a psychotherapist, polygraph field examiner, developer of the Objective Scoring System, version 3 (OSS-3) scoring algorithm, and is the author of several publications on various polygraph topics. Mr. Nelson has no proprietary or financial interest in the Backster You-Phase technique. Unrestricted use of this work is granted to polygraph training programs accredited by the APA, or recognized by the American Association of Police Polygraphists or the National Polygraph Association. For information contact raymond.nelson@gmail.com.

Mark Handler is an experienced police examiner and polygraph researcher who helped develop the OSS-3 and the Empirical Scoring System. Mr. Handler is the research chairperson for the American Association of Police Polygraphists, and has published numerous articles and studies on many aspects of the polygraph. Mr. Handler has no proprietary or financial interest in the Backster You-Phase technique. He can be reached at polygraphmark@gmail.com.

Greg Adams is a retired Police Chief, and the Deputy Director and APA certified primary instructor at the Backster School of Lie Detection, which has a proprietary interest in the Backster You-Phase technique.

Cleve Backster is the founder and director of the Backster School of Lie Detection, the developer of the Backster You-Phase technique and has a proprietary interest in the technique.

intended for event-specific single-issue diagnostic or investigative examinations for which the relevant questions are non-There are differences in the independent.¹ rules for question formulation.² exact effect of these esoteric However, the differences is not examined in this study.

The Backster You-Phase technique is scored using grand totals and physiological features defined by Backster, as described by Matte (1996) and Weaver (1980), using the system of 21 rules shown in Appendix A. Examinations conducted with the Federal You-Phase technique are scored using both grand total and sub-total scores, with features similar to those developed at the University of Utah (Bell, Raskin, Honts & Kircher, 1999; Kircher, Kristjiansson, Gardner & Webb, 2005; Podlesny & Raskin, 1978; Raskin & Hare, 1978; Raskin, Kircher, Honts & Horowitz, 1988).

Table 1 shows the sequence of test questions for the Backster and Federal You-Phase formats. Cleve Backster reversed the order of the Symptomatic and Sacrifice Relevant in 1979. The order of questions for the USAMPS version of the You-Phase technique, now referred to more generally as the Federal You-Phase technique, has remained unchanged though the question wording has been slightly modified from the original Backster language. Responses to each relevant question (RQ) of Federal Youexaminations are compared to Phase responses of the strongest of reactions to the nearest comparison question (CQ) using the seven-position and three-position test data analysis rules described by the (Department of Defense, 2006). Seven-position scores assigned to reactions to RQs of Backster You-Phase examinations are compared to the weaker of reactions to nearby CQs, using the Either-Or-Rule, unless the magnitude of

	Backster You-Phase	Federal-You-Phase	
1.	Neutral Question	Neutral Question	
2.	Symptomatic Question	Sacrifice Relevant Question	
3.	Sacrifice Relevant Question	Symptomatic Question	
4.	Comparison Question	Comparison Question	
5.	Relevant Question	Relevant Question	
6.	Comparison Question	Comparison Question	
7.	Relevant Question	Relevant Question	
8.	Comparison Question	Comparison Question	
9.	Symptomatic Question	Symptomatic Question	

Table 1. Backster and Federal You-Phase formats

¹Independence, in scientific testing, refers to whether the criterion variance of each individual test stimuli is thought to influence or be influenced by the criterion variance of the other stimuli. Assumptions of non-independence, imply that the criterion variance of all stimuli do affect the other stimuli, and that the test is designed to address a single uniform target for which the examinee will pass or fail as a whole.

 $^{^2}$ For example: Backster introduced the Sacrifice Relevant Question to the field that requires a narrow scope, and the relevant questions that follow cannot exceed that scope.

response to the stronger of nearby CQs produces a ratio of 4:1 or greater, using the Green-Zone-Abuse-Rule.³ Numerical scores of Backster You-Phase examinations are assigned using a seven-position rubric based on linear ratios⁴ and a system of 21 rules. Studies have also shown the effectiveness of scoring Federal You-Phase exams with the Empirical Scoring System (Nelson, In press; Nelson, Handler, Blalock & Cushman, In press).

Subtle differences between Backster and Federal systems exist in linguistics and semantics for the Sacrifice Question, Symptomatic Questions, Relevant Questions and Comparison Questions. These have been proposed by proponents and adherents of various PDD techniques as being related to differences in criterion accuracy. Technical questions are not included in numerical scoring, and are not addressed by the scoring rules shown in Appendix A.

Previous studies by Nelson (In press) and Nelson, Handler, Blalock and Cushman, (In press) have shown the Federal version of the You-Phase technique to provide criterion accuracy that differentiates deception from truth-telling at rates that are significantly greater than chance. The Backster You-Phase technique was used by Honts, Hodes and Raskin (1985) in a series of countermeasure studies, also reported in Honts and Hodes, (1983), and by Meiron, Krapohl, & Ashkenazi (2008) who studied the Either- Or Rule. Neither of these studies is satisfactory as a study of generalizable criterion validity of the Backster You-Phase technique.

In research designed to investigate the Backster Either-Or Rule, Meiron et al. (2008), studied a highly-selective, non-random, and non-representative sample of Backster You-Phase (N = 100) examinations. Examinations with erroneous or inconclusive results were excluded from that sample,⁵ and the sampling distribution was therefore systematically devoid of error variance and not useful to represent generalizable estimates of accuracy and error rates.

Honts et al., (1985) used standard field components for recording electrodermal and pneumograph data but used a non-standard sensor instead of a blood pressure cuff to measure cardiovascular activity. As a result, cardiograph data and total scores from the Honts et al. (1985) study may vary in unknown ways from the distributions of scores conducted using standard field testing equipment.

The present study is intended to help fill a gap in the published literature regarding criterion validity of the Backster numerical scoring system and the You-Phase technique.⁶ The hypothesis was that blind scores obtained using the Backster numerical scoring system can discriminate confirmed deceptive from confirmed truthful You-Phase exams at rates that are greater than chance.

³ Backster considers it incorrect to score to the weaker CQ if the ratio of response magnitude difference is 4:1 using the stronger CQ. Although the Either-or-Rule is considered the main scoring rule, the Green-Zone-Abuse-Rule supercedes it.

⁴ Handler et al (2010) described that electrodermal responses, like many human physiological responses, are nonlinear, and often log-linear, and that linear assumptions are false regarding electrodermal reactions recorded during PDD examinations. Nevertheless, linear ratios have been used traditionally in PDD test data analysis models.

⁵ As described in the presentation of the study data at the annual conference of the American Polygraph Association in Indianapolis (August 2008).

⁶ Criterion validity of PDD techniques is dependent on a combination of 1) the test questions sequence, including rules for target selection, question formulation and stimulus presentation, and 2) the method for test data analysis. PDD validity may also be affected by other variables, including examinee suitability, examiner competence, and the contextual details of test administration, all of which are not the focus of this study.

Method

Participants

Seven examiners provided blind scores for this study. Ages of the participants ranged from under 30 years to over 60 years. All participants were trained at the Backster School of Lie Detection between 1981 and 2011. Two of the participants were female, five were male. Six of the participants were Polygraph American members of the Association (APA). Three were members of the American Association of Police Polygraphists (AAPP). Three participants had two-year two college degrees, had four-year undergraduate degrees, one had completed a graduate degree, and the remaining participant listed six vears of college education and no degree. Six of the participants were caucasian, and one was American Indian. Three were members of the ASTM International (American Society for Testing and Measures). Five were members of multiple professional organizations. One participant was not a member of any professional organization. The third author (GA) remained blind to the criterion status of all cases, managed the data collection activities, and also scored the cases as a participant in the study.

Sample data

The data selection criteria produced a matched random sample (N = 22) of You-Phase examinations, previously described by Nelson, Handler, Blalock and Cushman (In press), selected from the confirmed case archive at the United States Department of Defense. Eleven confirmed truthful examinations met the selection criteria of physically healthy adult criminal suspects, psychotropic reportedly not taking medications, whose examinations consisted of three test charts. Examinations consisted of two relevant questions regarding a single issue, three comparison questions, and other procedural questions as required by the You-Phase technique. Eleven matching confirmed examinations deceptive were randomly selected.

Examinee data were completely anonymous, and selection into the study sample had no effect on the criminal investigation or case outcome. The sample cases were confirmed as deceptive via a combination of extrapolygraphic evidence and confession, or extrapolygraphic evidence or confession which inculpated an alternative suspect, exonerating the examinee. Examination results from the original examiners were not 100 percent accurate.

All examinations were conducted by US Federal and local law enforcement agencies, according the procedures to described by the Department of Defense, (2006). Although the test question sequence and test administration procedures are largely similar for the Federal and Backster You-Phase techniques, one important difference in test administration procedures deserves further explanation. Department of Defense (2006) procedures include the insertion of neutral questions in response to artifacted or unstable segments. Backster has no such procedure, and proscribes the comparison of RQs to CQs that are not immediately temporally adjacent to, either before or after, the RQ. Backster scoring protocols require the selection of the single temporally adjacent CQ for comparison with the RQ, thereby precluding the use of the Backster Either-Or-Rule when neutral questions are inserted into the test question sequence. Nine of the 22 You-Phase exams in the confirmed case sample included inserted neutral questions. The sample data and data collection plan were reviewed by the third and fourth authors (GA and CB) prior to the initiation of scoring activities. After the data collection and prior analysis Adams (personal to the communication, July 1, 2011) affirmed that the confirmed field sample, examiners, and sufficiently representative, scores were inserted neutral questions notwithstanding, to proceed with the data analysis and present the results as a cautious portrayal of the criterion accuracy of Backster numerical scores of You-Phase exams.

Analysis

Two-way ANOVAs were used to investigate confounding variables such as age, gender, number of event-specific examinations completed, years of experience in law enforcement polygraph units, membership in professional associations, and attendance at Backster work conferences.

Bootstrap methods were used to calculate the pairwise correlation between the

numerical scores of the seven survey participants. Monte Carlo methods were used to calculate variance estimates and statistical confidence intervals for a dimensional profile of criterion accuracy.

Results

All statistical results were evaluated with a level of significance set at alpha = .05.

There were no significant differences in unweighted decision accuracy or inconclusive rates as a function of age or gender. Neither were differences significant as a function of the number of event specific examinations completed, education, years of experience in law enforcement polygraph units, membership in a professional association, or attendance at Backster work conferences. Ethnicity was not evaluated because six of the seven participants listed the same ethnicity, and any conclusion based on a single individual would be suspect.

Reliability

The mean deceptive total score for three charts was -19.649 (SD = 6.482), and the mean truthful total score was 3.612 (SD = 10.010). A pairwise bootstrap of 1000 iterations of the pairwise correlation between numerical scores produced a Pearson product moment correlation of .567 (95% CI = .406 to .729). This is interpreted as a moderate to high correlation of the numerical results produced by the seven participants.

Criterion validity

Categorical decisions. based on numerical scores were made using three-chart grand total scores. Cutscores recommended by the Backster School of Lie Detection $(2011)^7$ are +7 or greater for truthful classifications, and -13 or lower for deceptive classifications based on three presentations of the test questions (i.e., test charts). Twelve cases were classified as deceptive using a rule unique to the Backster technique taught at the Backster School of Lie Detection (2011), which allows for a decision using two of three test charts if the three chart grand total is inconclusive. Five cases were also classified as truthful using two charts when the threechart grand total was inconclusive. The recommended two-chart grand total cutscores are +5 or greater for truthful classifications, and -9 or lower for deceptive classifications. Application of the two-chart cutscores to all cases resulted in no changes in categorical decisions.

Table 2 shows the mean, standard error and statistical confidence intervals of the dimensional profile of criterion accuracy, including: test sensitivity to deception, test specificity to truth-telling, false-negative and false-positive error rates, inconclusive rates for deceptive and truthful cases, positive predictive value (PPV), negative predictive value (NPV), the proportions of correct decisions, without inconclusive results, for deceptive and truthful cases, and the unweighted average of the proportions of correct decisions and inconclusive results for the deceptive and truthful cases. Unweighted decision accuracy, excluding inconclusive results was significantly greater than chance (p < .001).

Discussion

Results from this study support the validity of the hypothesis that Backster numerical scores of confirmed You-Phase examinations can differentiate deception from truth-telling at rates that are significantly greater than chance.

The absence of any significant effect demographic variables both for is uninteresting and interesting. There are no compelling arguments why most demographic variables would interact with PDD test accuracy. Education, experience and continuing training are exceptions. However, the study participant with the least amount of experience produced the highest inconclusive rates, but the highest level of decision accuracy, excluding inconclusives. The participant with the greatest level of experience produced the lowest inclusive rates, but next to the lowest level of decision

⁷ Based on changes made by Cleve Backster in 1983.

	Mean	St. Err.	95% CI
Unweighted Accuracy	.828	.079	.674 to .982
Unweighted Inconclusives	.115	.070	.001 to .254
Sensitivity	.943	.073	.800 to .999
Specificity	.543	.148	.254 to .833
FN Error	.009	.028	.001 to .063
FP Error	.274	.128	.022 to .525
D Inc	.048	.067	.001 to .180
T Inc	.183	.121	.001 to .421
PPV	.775	.110	.559 to .991
NPV	.986	.044	.899 to .999
D Correct	.991	.029	.933 to .999
T Correct	.665	.154	.364 to .966

Table 2. Criterion Accuracy for Backster numerical scores of You-Phase exams

accuracy. The practical meaning of this is simply that less experienced examiners are able to make good use of their training to reach accurate conclusions. These results also suggest that experience may be more closely correlated with confidence than accuracy. Demographic and training variables should continue be addressed in future research.

A limitation of the present study is that it does not attempt to evaluate the potential use of statistical norms, decision theory, optimal cutscores, or subtotal scores to increase criterion accuracy or decrease inconclusives. It is clear from the results of this study that the presently recommended cutscores may be suboptimal for decisions based on three charts. Satisfactory criterion accuracy was obtained only through the use of cutscores recommended for decisions based on two charts (+5/-9). Future research on Backster You-Phase examinations should investigate the use of normative data to develop statistically optimal cutscores that can make use of all obtained data and more effectively prioritize objectives regarding test specificity, sensitivity, error rates, and inconclusive results. Decision rules, involving the potential use of both grand total and subtotal scores, should also be the focus of future research.

Another acknowledged limitation of the present study is that sample cases used in this study involve the use of Federal You-Phase exams in a study of the Backster numerical scoring system. Although it would be preferable to study the Backster numerical system with examinations conducted using the Backster version of the You-Phase test question sequence, no suitably random and representative sample was available. Careful consideration was given to the differences between the Federal and Backster You-Phase techniques. The second and third authors reviewed and approved the sample data prior to data collection.

It is the opinion of the primary author that the evidence is absent, and arguments are unconvincing regarding any advantageous or detrimental impact on criterion accuracy from subtle and esoteric differences, between Backster and Federal You-Phase the techniques. Neither linguistics nor the order of technical questions are factored into the numerical scoring protocols. The major difference that could affect the numerical scores and criterion accuracy is the insertion of neutral questions in response to unstable artifacted data segments. Federal or procedures include, and indeed require, the insertion of neutral questions in response to artifacted or unstable segments. Backster

makes no assumption that the insertion of neutral stimuli will correct deficiencies in the data, and proscribes the use of the Either-or-Rule when two CQs are not temporally adjacent to an RQ in the recorded question sequence. Backster recommends temporary extension of question pacing for unstable or artifacted data segments, which precludes the upgrading of a numerical score to the maximum scores of +3 or -3, using the Question Pacing Upgrading Rule shown in Appendix A. In the case of inserted neutral questions, Backster scoring protocols require the selection of the single temporally adjacent CQ for comparison with the RQ. Conceptual arguments support a variety of possibilities resulting from this difference. It is conceivable that forgoing the use of the Either-Or-Rule could move the location, shape, and variance of either or both distributions of scores (i.e., CQ and RQ) in a positive or negative direction. It is also possible that this difference could have no effect on decision outcomes. The exact effect of inserted neutral questions on the criterion accuracy of Backster numerical scores of You-Phase examinations is a hypothesis that requires further research and which cannot be resolved by opinion or speculation. This should be the focus of future research.

Study participants registered notes in 13 examinations and on 19 test charts. Notes pertained to the insertion of neutral questions into nine of the 22 sample cases. Pneumograph data were reported as distorted and uninterpretable for one examination. Three examinations were reported as suspect for use of countermeasures. With the exception of the insertion of neutral questions into the test question sequence of Backster You-Phase exams, these conditions can be expected to be encountered in field settings.

Another limitation to the present study is the sample data which consist of confirmed field investigation cases. It is unknown how the cases were selected for inclusion in the confirmed case archive at the Department of Defense. However, it is known that samples constructed from confirmed field investigation cases are non-random in that cases are selected based on the availability of confirmation data. Concerns are unavoidable regarding assumptions about how well the sample cases represent the population of both confirmable and un-confirmable cases in field settings. As always, there is an unknown possibility that the exclusion of inconclusives, false-positives. and false-negative errors without confirmation data could lead to an overestimation of criterion accuracy rates. Although the scientific concern stemming sampling bias will be that of from overestimation of criterion accuracy, some readers may also be inclined to interpret the present results as an underestimation of the accuracy of the Backster You-Phase in field settings. These are questions that should be addressed by comparing the present results with those of future studies.

No single study can serve as a answer regarding definitive matters of empirical validity, and these results are presented by the authors as simply All study results should be informative. compared to the results of other studies in order to develop a more comprehensive understanding of the issues under investigation. Regardless of the acknowledged limitations of the present study, the study participants achieved a high level of criterion accuracy that was significantly greater than chance. Further research on the Backster You-Phase technique is recommended.

References

- Backster School of Lie Detection (2011). Basic polygraph examiner's course chart interpretation notebook. Backster School of Lie Detection: San Diego.
- Backster, C. (1963). Standardized polygraph notepack and technique guide: Backster zone comparison technique. Cleve Backster: New York.
- Bell, B. G., Raskin, D. C., Honts, C. R. & Kircher, J. C. (1999). The Utah numerical scoring system. *Polygraph*, 28(1), 1-9.
- Department of Defense (2006). Federal psychophysiological detection of deception examiner handbook. Reprinted in *Polygraph*, 40(1), 2-66.
- Gordon, N. J., Mohamed, F. B., Faro, S. H., Platek, S. M., Ahmad, H. & Williams, J. M. (2005). Integrated zone comparison polygraph technique accuracy with scoring algorithms. *Physiology & Behavior*, 87(2), 251-254.
- Honts, C.R., & Hodes, R.L. (1983). The detection of physical countermeasures. *Polygraph* 12(1):7-17.
- Honts, C.R., Hodes, R.L., & Raskin, D.C. (1985). Effects of physical countermeasures on the physiological detection of deception. Journal of Applied Psychology, Vol 70(1), Feb 1985, 177-187.
- Kircher, J. C., Kristjiansson, S. D., Gardner, M. K. & Webb, A. (2005). Human and computer decision-making in the psychophysiological detection of deception. University of Utah.
- Matte, J. A. (1978). Polygraph Quadri-Zone Comparison Technique. Polygraph, 7(4), 266-280.
- Matte, J. A. (1996). Forensic Psychophysiology Using the Polygraph. Williamsville, NY: J.A.M. Publications: New York.
- Matte, J. A. & Reuss, R. M. (1989). A field validation study of the Quadri-Zone Comparison Technique. *Polygraph*, 18, 187-202.
- Meiron, E., Krapohl, D. J. & Ashkenazi, T. (2008). An assessment of the Backster "Either-Or" Rule in polygraph scoring. *Polygraph*, 37, 240-249.
- Podlesny, J. A. & Raskin, D. C. (1978). Effectiveness of techniques and physiological measures in the detection of deception. *Psychophysiology*, 15, 344-359.
- Raskin, D. C. & Hare, R. D. (1978). Psychopathy and detection of deception in a prison population. *Psychophysiology*, 15, 126-136.
- Raskin, D., Kircher, J. C., Honts, C. R. & Horowitz, S. W. (1988). A study of the validity of polygraph examinations in criminal investigations. Final Report, National Institute of Justice, Grant No. 85-IJ-CX-0040.
- Reid, J. E. (1947). A revised questioning technique in lie detection tests. *Journal of Criminal Law* and Criminology, 37, 542-547.
- Weaver, R. S. (1980). The numerical evaluation of polygraph charts: Evolution and comparison of three major systems. *Polygraph*, 9, 94-108.

Appendix A

Backster Chart Analysis Rules

Primary Rules

- 1. Either-Or Rule
- 2. Non-reinforcement Rule
- 3. Green Zone 'Yes' Answer Penalty Rule
- 4. Change in Amplifier Sensitivity Rule
- 5. Timely Reaction Rule
- 6. Anticipatory Reaction Rule
- 7. Lack of Reaction Via Deduction Rule
- 8. Delayed Cardio Reaction Recovery Rule
- 9. Minimum Lack-of-Reaction Rule
- 10. Plunging GSR Baseline Rule

Secondary Rules

- 1. Green Zone Abuse Rule
- 2. Tracing Average Trend Change Rule
- 3. Presence of Reaction Via Deduction Rule
- 4. Single Cycle Trend Conformance Rule

Upgrading Rules

- 1. Question Pacing Upgrading Rule
- 2. Tracing Purity Upgrading Rule
- 3. Reaction Intensity Upgrading Rule

Tracing Oddity Rules

- 1. Listening Reaction vs. Listening Distortion Rule
- 2. Answer Reaction vs. Answer Distortion Rule
- 3. Stabilized Blood Pressure Trend Rule
- 4. Extra-Systole Cluster Rule