A Discussion on Updating the Backster Zone Comparison Technique

by J. Patrick O'Burke and Clifford "Chip" Morgan

©FotoliaLLC/zimmytws

Summary

The history of the evolution of lie detection extends back several hundred years. This included the development of the requisite technology and the development of psychology, criminology and sociology sciences as they related to lie detection. The use of polygraph for improved credibility assessment resulted from the effort of many people to eliminate harsh and abusive methods for interrogation that have been pervasive for hundreds, if not thousands, of years. Not to diminish the many contributions made in this evolution, Cleve Backster should be recognized for assimilating previous scientific developments into his Zone Comparison formats and numerical scoring system. The introduction of Backster's methodology was an important event in helping standardize the training offered to the polygraph profession in test administration and test evaluation. However, in the sixty years since its introduction there have been only minor changes in the methods originally taught by the Backster School of Lie Detection. Scientific research and professional polygraph standard practices are now indicating that it may be time to consider updating the Backster You Phase technique. We suggest that it is time to update the original Backster format to three relevant questions and incorporate the use of better research validated scoring methods.

Discussion

The evolution of psychology, criminology and sociology are very recent in terms of human history. Each of these early efforts to study human behavior and detecting deception are important in the incremental development for advancing what was to become polygraph. August Vollmer's request in the 1920's to bring crime fighting technology to the Berkeley Police Department was a catalyst for John Larson and Leonard Keeler to synthesize some of these early research findings. Their work helped develop the earliest versions of a polygraph instrument. Following the introduction of an instrument, numerous people worked to develop question formats and methods for test data evaluation. Some of these efforts produced things such as analyzing behavior cues and the use of question types that were later found to be unsupported.

Drawing from the work of many who were involved in polygraph research, Cleve Backster developed a family of formats and his own method for numerical evaluation of his formats. Nelson (2013) more fully described Backster's efforts as the first to create a family of question formats in the Exploratory, S-K-Y and You Phase formats, along with the use of a more formalized system of rules, compared to what was previously widely available, for numerical scoring analysis. Regardless of what question technique one believes is now best, Backster's intuition for understanding and solving polygraph problems was a substantial influence that shaped the way forward during the mid-20th century.

In 1960, Backster, using his own "note pack" system, introduced the You Phase format as a single issue, single facet test with three relevant questions, #33 and #35, and #37. However, the use of this third relevant question was subsequently abandoned in school training and field practice. The You Phase format also included three bracketing comparison questions, a sacrifice relevant question, neutral questions and two symptomatic questions. The comparison questions were initially devised as a probable lie question to contrast with the relevant issue question. This contrast was measured by assigning a numerical score at the relevant question for each polygraph component. Numerical scores in the You Phase were aggregated into a grand total score with an assigned cut score for a final decision of truth or deception.

The nature of the role of school director, as well as that of its primary school instructors, requires that staff should continuously review what is being

taught in school. Sixty years later, this would include whether the Backster Zone Comparison formats are still optimally structured and if the original numerical analysis method is adequately supported by scientific evidence as to its ability to provide probabilistic estimates in polygraph decision making. This need to continually evaluate the training and school curriculum is required in the American Polygraph Association school accreditation requirements. It is also fundamental to good science and the advancement of the profession.

The Standards of Practice for the American Polygraph Association (APA) define polygraph as an evidence-based test of the margin of uncertainty or level of confidence surrounding a categorical conclusion of deception regarding a test issue. While the test is often called a "lie detector", the APA definition informs us that this term is incorrect. History also shows that some have, on occasion, had issues with reporting on the accuracy of polygraph results. In 2010 and 2011, the APA conducted a meta-analytic survey of the available peer reviewed or published research on polygraph accuracy. That report clarified what appeared to have been often misunderstood in the past, particularly where research was lacking or self-sponsored. The report showed there was little evidence to suggest any polygraph test format was superior to others. Instead, the report suggested that differences in scoring methods, as well as understanding the issue of question independence, were deserving of more attention.

The APA defined a polygraph technique as comprised of both a format and an analysis method. The meta-analytic survey showed that some scoring methods were more likely than others to produce unacceptably higher rates of inconclusive results, or problems with inter-rater reliability. The authors observed the Backster scoring method is unique among other methods in requiring the examiner to make decisions about each of the relevant questions before determining which comparison question will be selected. Other scoring methods such as the ESS and the Federal 7 position method differ in that they make no conclusion about the relevant question until after the comparison question is selected. Whereas the Backster scoring method has traditionally selected the comparison question with the lesser change in physiology, except under special circumstances, these other methods require the selection of the comparison question with the greater change in physiological activity. Other scoring methods, such as the Utah method, generally dictate scoring each relevant question against the preceding

comparison question.

In developing his numerical scoring method, Backster defined his desired cut scores for examiners to make decisions for truth or deception in a highly subjective manner. His idea was that cut scores should be based on the number of charts collected and the number of relevant questions. Backster's original cut scores were symmetrical and developed more through intuition than through analysis. Backster held that cut scores should increase in a linear manner with the number of test charts. When the test score did not equal or exceed the cut scores additional charts were required to avoid an inconclusive test result. An unintended consequence of Backster's increasing cut scores was that conducting more charts tended to make it more it difficult to achieve a conclusive test result. Backster later adjusted the cut scores because of reports from the field that they were not optimal.

Although Backster's suggested cut score were reasonably effective when a decision was achieved, there does not appear to be evidence that the original cut scores, or the later adjusted scores, were ever optimized for best test performance and decision making through statistical analysis. Instead, it appears that cut scores were selected through Backster's own heuristics. The effects of this absence of an analytic approach can be observed in the meta-analytic survey, where the Backster technique, though decision accuracy is similar to other techniques, may suffer from higher inclusive rates and weaker test specificity than other methods.

The more significant issue impacting examiner test administration is that the cut scores used in the Backster method are cumulative, increasing with each chart and relevant question presentation. The original formula designed by Backster assigned two points per relevant question spot and summed the total, then adding one point to a final total to arrive at the desired cut score. For example, two charts (which we now know is unacceptable for a comparison question test) with relevant questions #33 and #35 have four spots. Multiply the four spots by two, and then add one point, for a final cut score of +9 for NDI and -9 for DI. With three You Phase charts, examiners had six spots, so they multiplied by two, and then added one point, establishing a cut score of +/- 13. This formula increases the cut scores for each additional chart that was administered. Later, in response to reports from examiners in the field, Backster adjusted the cut score the requirement for truthful subjects to one

point per relevant spot. This required a numerical score of +7 for two relevant questions and three charts. Backster may have used some analytic or optimization process to obtain his recommended cut scores, but we have not seen it.

Despite Backster's cut score revision, using a third relevant question at #37 was not adopted as common practice by field examiners - possibly because it's use tended to further escalate the required cut score, leading to increased inconclusive results when more data was obtained. Because the required cut scores continually escalated, examiners in the field were not able to enjoy the normally expected mathematical advantage that occurs from increasing the amount of data used to make a decision.

Although the Backster technique was shown to perform similar to other methods in the meta-analytic survey (American Polygraph Association, 2011), there has been some concern among field examiners and trainers regarding higher inconclusive rates and weaker test specificity for the Backster technique compared to other methods. However, Hedges, Dietchman and Samra (2013) studied the issue of whether selecting the weaker comparison for evaluation would bias test decisions against the truthful subject – an issue that needs more study. Our greatest concern is that creating a prior initial decision point at the relevant question for comparison selection mathematically increases the potential for issues in inter-rater reliability with Backster scoring rules. More possible choices mean greater deviation in inter-rater reliability.

The APA meta-analytic survey supported the use of the Backster You Phase format combined with Backster scoring as providing sufficient criterion validity for paired testing, though not at the level required for evidentiary testing. The authors are forced to consider evidence suggesting that the Backster scoring method may have contributed to those findings. We support our concerns on scoring by observing that the Federal You Phase, and the three question Federal ZCT, meet evidentiary standards and are structurally similar – indeed virtually identical for scientific or analytic purposes - to the Backster You Phase format. However, the Federal You Phase and Federal ZCT formats met the requirements for evidentiary testing only when incorporating the evidence-based improvements of the empirical scoring system (ESS). Since the Federal formats are structurally similar to the Backster format it is reasonable to conclude that the difference in performance can be largely attributed to the scoring meth-

od.

Nelson and Handler (2015) published tables for reference distributions for a variety of test question formats and analysis methods, including both Backster numerical scoring system and the ESS with event specific Zone Comparison formats. They reported that when using Backster scoring with the two-question You Phase format, the mean grand total scores were -12 for the guilty, and +6 for the innocent. These reported mean (Backster) scores are near the lower margins of the Backster established cut scores, which provides some limited insight as to why inconclusive rates were observed as greater for the Backster technique than for other methods.

Conclusion

We conclude that a Backster event specific format would benefit from being conducted as a single issue three relevant question format using ESS scoring. The Backster School will refer to this format going forward as the Backster ZCT. The expected benefits include additional data collection with increased confidence in decision making with ESS scoring. The authors believe this change is consistent with the APA meta-analytic survey, pointing towards the use of three-question formats as a more optimal solution to two-question versions.

The reader should not interpret that this decision as inferring or suggesting that the original two question Backster You Phase is invalid. Simply, collecting more data is often advantageous for test data analysis. We can have greater confidence with three relevant question formats compared to two relevant question formats. Some readers may suggest the Backster School should redesign the Backster scoring method. The currently validated ESS scoring method is already a synthesis of research efforts learned from studies of the Backster, Utah and Federal seven and three position scoring methods. At this point we believe the wise and responsible choice, when seeking evidence-based solutions for school training and field practice, is to make use knowledge that is already available. For this reason, we conclude that the Backster three question ZCT (Zone Comparison Test) using ESS scoring is an optimal solution for teaching in the Backster School of Lie Detection for use with event specific issues.

Following the decision to go forward an emphasis on using the Backster ZCT with three relevant questions, it is then appropriate to discuss how this will be addressed in the Backster School curriculum. The school will use a simple teaching method that eases

examiner concerns about developing question wording for three relevant questions as single issue, single facet solutions. The school solution is to teach students a "what-when-where" approach to relevant questions that consistently that can easily be adapted to a wide variety of situations. For example, with case facts from an employee theft from a convenience store cash register, the following wording can be taught:

(What) R33-Did you steal any of that money from the store cash register?

(When) R35-Did you steal any of that money reported missing last Friday?

Where) R37-Did you steal any of that money from the 7-11 on First Avenue?

In addition, going forward the Backster school will teach that the use of multi-facet event specific relevant questions is also permissible within the Backster ZCT three question format. We concluded that this is well supported in the meta-analytic survey and consistent with the published evidence on the Federal ZCT and Utah formats. A variety of solutions are possible in a multi-facet approach. As a general practice, the Backster school will teach students to utilize two primary issue questions and a third relevant as an evidence connecting, or secondary involvement question. For example, with case facts from a robbery at the 7-11 store, the following would be suggested:

(What) R33-Did you steal any of the money taken in that 7-11 robbery?

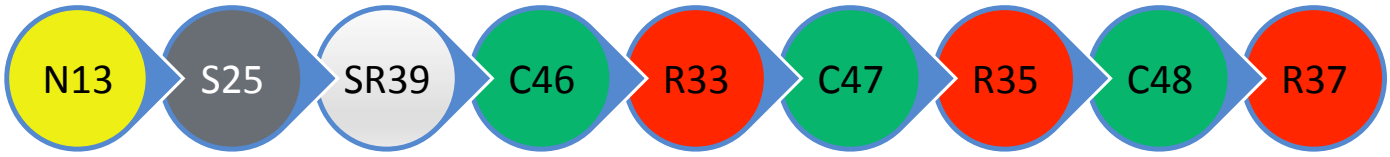
(When) R35-Did you rob that 7-11 store last Friday?

(Where) R37-At the very time that 7-11 was robbed on First Avenue, were you there?

Additionally, we intend to incorporate other evidence-based changes into the updated curriculum taught at the Backster School of Lie Detection. First, the use of Directed Lie Comparison questions, as well as Probable Lie Comparisons, will be taught as acceptable. Blalock, Nelson, Handler, and Shaw (2011) fully discussed that Directed Lie Comparison questions perform as well as Probable Lie Comparison questions. Arguments against the use of Directed Lie Comparison questions are not based on empirical data, and there is nothing to be gained from restricting field examiners from making effective use of evidence-based solutions. The Backster technique has already been discussed in the courtroom with ESS scores and DLC questions,

The Backster School will also teach

that the second symptomatic question #26 should be removed from the ZCT format. Research suggests that the symptomatic question adds no value to the ZCT format beyond informing the test subject of what is to be asked. There is sufficient evidence to suggest that the symptomatic does not identify outside issues, does not fix inconclusive tests and does not explain inconclusive rates. The updated Backster ZCT format will appear as below:



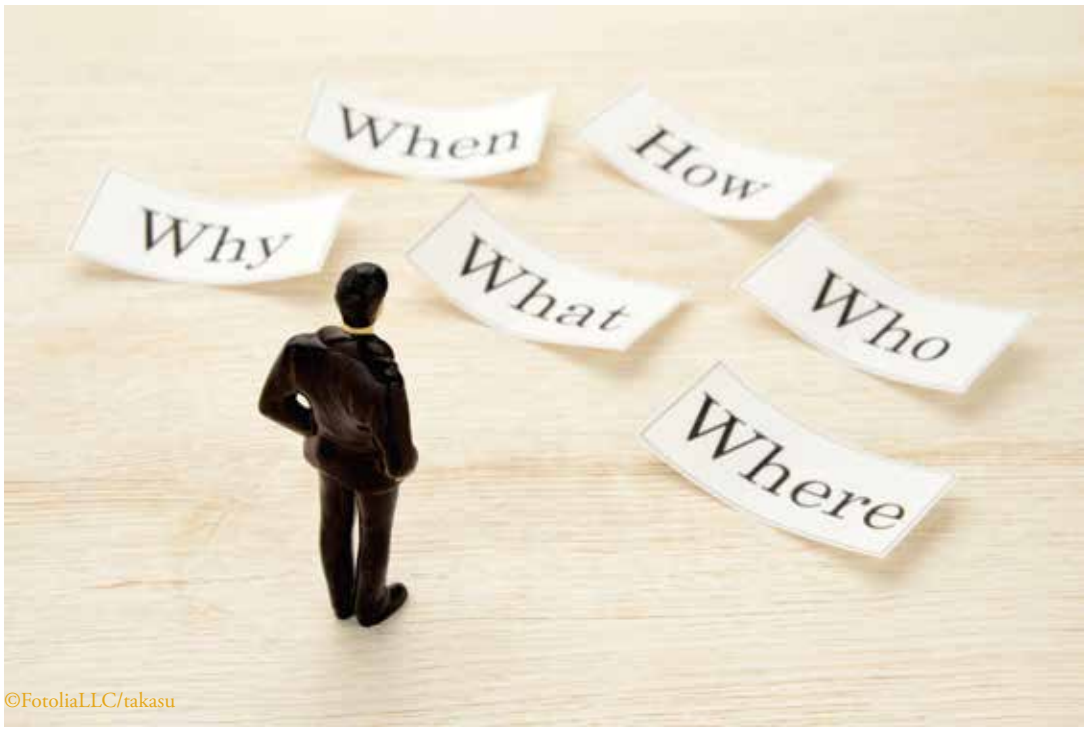
We argue that this updated format for the Backster ZCT is substantially similar to the Federal three question ZCT, as well as the Utah format. Visualize the removal of the #8 Symptomatic question from the Federal ZCT and the two formats are virtually identical. It is our position that there is no basis of evidence to argue against the generalization of extant knowledge regarding effect sizes of ESS scores for event-specific exams with three relevant questions, and no basis of evidence to argue against the notion that this format can already achieve the level of accuracy required for APA evidentiary standards.

To reiterate the Backster School is not suggesting the earlier You Phase format is invalid, nor are we suggesting that examiners are compelled to abandon its use. The Backster School is merely striving to provide a more optimal test format for examiner use into the future. It is our sincerest desire to respect the contributions that Cleve Backster provided to the polygraph community and our decisions reflect our desire to be good stewards of his legacy.

Being good stewards of the Backster school, includes being responsible to our profession, and this means that we must accept the fact that everything that we would like to know is not currently known. Nor was everything known fifty years ago. The Backster School will continue to remain open to new research that explores what is still yet unknown. Being on the cutting edge of training means we choose to honor the legacy of those who came

before us and we commit to strive to improve on what we have been given. We welcome your comments.

The authors sincerely thank the following for their suggested edits for this paper; Raymond Nelson, Don Krapohl, Barry Cushman, and Skip Webb.



References

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. Polygraph, 40(4), 196-305. [Electronic version] Retrieved Sept. 24, 2018 from http://www.polygraph.org/section/research-standards-apa-publications.

Backster, C. (1963). Standardized polygraph notepack and technique guide: Backster zone comparison technique. Cleve Backster: New York.

Hedges, K., Deitchman G., & Samra D. (2013). A Comparison of the Backster Scoring and Traditional Federal Scoring Using Computerized Analysis. Polygraph. 42(1), 18-28. Tagged XML BibTex Google Scholar Download: Backster and Federal Compared (127.12 KB)

Nelson, R., & Handler M. (2015). Statistical Reference Distributions for Comparison Question Polygraphs. Polygraph. 44(1), 91-114. Tagged XML BibTex Google Scholar Download: 44(1)-statistical reference distributions.pdf

Nelson, R. (2013). Backster Polygraph Techniques and the APA Standards of Practice. Polygraph. 42(1), 29-34. Tagged XML BibTex Google Scholar Download: backster_techniques.pdf (46.3 KB)