

Using Single Issues in Screening Examinations

J. Patrick O’Burke, Contributions by Dennis Westerman

Edits by Walt Goodson

*Notice: The authors want to thank the polygraph program managers who provided information for this article. Those persons were provided confidentiality for this article out of respect, and nothing herein is a criticism of their programs or policies on polygraph use. The views and opinions expressed in this article are those of the authors and do not necessarily represent those of any other person. Address correspondence about this article to Patrick O’Burke at jpoburke@the-polygraphinstitute.com

The National Research Council (2003)¹ report offers a discouraging perspective on the use of the polygraph for screening tests. Two historical problems can be inferred from this report. First is the inclusion of multiple issues within a test

format. This is most likely driven by consumers, and occasionally polygraph examiners, who erroneously perceive the polygraph as a lie detector, capable of discriminating truth or deception when mixed in the test. Second is formulating the relevant issue test question for a screening examination. The authors would suggest that examiners should embrace single-issue testing for screening to the same degree that they embrace single issues for specific issue testing. This is consistent with the bulk of the literature that the best test is a single issue. However, change is often difficult without some thoughtful consideration as to what are the best evidence-based practices we can engage in as professionals.

¹National Research Council. 2003. *The Polygraph and Lie Detection*. Washington, DC: The National Academies Press. <https://doi.org/10.17226/10420>.

The most optimal practice for formulating a relevant issue test question for a polygraph test is by articulating the behavior in question in a manner that's easily understood by the examinee. Reported allegations often illustrate the behavior that can drive the relevant question construction. Screening tests, however, are about unreported incidents to make risk, or suitability decisions about the examinee. While it may seem better to craft a more comprehensive focus for the relevant questions for risk decisions, this practice will typically result in unsatisfactory decision outcomes. The NRC report speaks to the difficulties of determining the issue criterion for test questions for espionage. As well, police reformists have recently suggested polygraphs with relevant questions to screen out applicants with maladaptive attitudes and beliefs. This use may seem soothing to the public, but it is more likely the outcomes will not meet their expectations.

Polygraph examiners annually conduct thousands of comparison question polygraph tests (CQT) over suspect behaviors from specific reported events, as well as suspect behaviors where no event has been reported. The American Polygraph Association (APA) Standards of Practice define testing where no incident or allegation has been made as a screening exam and further that screening tests may be conducted as single or multiple-issue exams.² Almost

universally, polygraph examiners would agree that it is a best practice to limit testing to one single identified event in any given test format. Conversely, most examiners choose to test over multiple issues with screening tests where no identified reports are known. This practice seems particularly odd since identified events should lead to more precise wording for suspect behavior than more broadly worded questions dictated by no known events. A likely conclusion is that examiners are motivated by convenience in testing and supporting polygraph consumer needs for adopting a multiple-issue testing format.

This inexplicable tendency may be supported, if not caused by, some of the APA Model policies, including those for Law Enforcement Pre-Employment Screening³ and Post-Conviction Sex Offender (PC-SOT) testing.⁴ Understandably, the APA acknowledges the challenges with conducting screening tests in 4.2 of the Law Enforcement Model Policy with the following:

"Screening exams may at times be narrowed to a single target issue of concern, in the absence of a known incident or known allegation. However, most screening exams include multiple issues of concern (mixed issues), in which it is conceivable that a person could be involved in one or more issues while remaining uninvolved in other issues of concern."

²American Polygraph Association. Standards of Practice. [Electronic version] Retrieved 10/25/2022 from <http://www.polygraph.org>.

³American Polygraph Association. Model Policy for Law Enforcement/Public Service Preemployment Polygraph Screening. [Electronic version] Retrieved 10/25/2022 from <http://www.polygraph.org>.

⁴American Polygraph Association. Model Policy for Post-Conviction Sex Offender Testing. [Electronic version] Retrieved 10/25/2022 from <http://www.polygraph.org>.



This language underscores the reality that truth and deception are often mixed on a screening test and not homogenous in their inclusion.

In the PCSOT Model Policy, the APA suggests that examiners “*should*” conduct Maintenance exams as a multi-issue test. Curiously, the APA acknowledges that examiners are not prohibited from using a single issue “*when such an approach will lend to more accurate or satisfactory resolution of the investigative target.*” The PCSOT Model Policy suggests that sexual history formats can be used with two to four separate issues. It may be a question to try and understand if polygraph association’s model policies provide direction for evidence-based polygraphs or support the examiners’ practices.

Perhaps examiners are over-relying on the APA Meta-Analytic Survey (2011) in selecting a format for screening without fully scrutinizing the APA’s report. Examiners typically select formats from the Investigative category for screening use. Inclusion as an investigative format requires a technique to demonstrate a mean accuracy of 80%, and not exceed 20% in inconclusive test results. It is essential to understand the difficulty of conducting a field study research project on an unidentified behavior since the ground truth would be impossible to know for the research participants. As such, most research for investigative formats involves laboratory studies that could be genera-

lized to field polygraph use. Appropriately, the ad hoc committee chose to treat field studies and laboratory studies with equal consideration. However, examiners should be mindful that challenges could be present in understanding the studies.

There have been limited laboratory studies on screening issues, but their results have underscored notable challenges worth understanding. Generally, these studies show that polygraph performs better than chance in identifying deception on the test but performs less than satisfactorily when trying to discriminate which specific issues are deceptive when truth and deception are mixed on the test. For example, Correa and Adams (1981) reported that accuracy was 100% on subjects who were deceptive, but only 68% accurate in determining which issue the participant was deceptive on.⁶

Examiners choosing to use multiple issue formats may not appreciate the potential issues most test data analysis (TDA) models impose when relying solely on Spot-Score-Totals. These could include decreased specificity in decision-making, lower mean diagnostic scores and increased inconclusive results. Two-stage grand-total scoring models, commonly used with single-issue test formats, provide a better balance of sensitivity and specificity, an increase in mean scores, and a reduction in inconclusive test results.

⁵American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. [Electronic version] Retrieved 10/25/2022 from <http://www.polygraph.org>.

⁶Correa, E., Adams, H., (1981). The Validity of The Preemployment Polygraph Examination and the Effects of Motivation. Polygraph. 1981 10(3). pp 143-155



In Barland (1981)⁷ a research project was conducted that attempted to evaluate the CIST as a directed lie comparison (DLC) format against various test evaluation methods. There were fifty-six participants, with twenty-six subjects programmed as truthful, and thirty programmed as deceptive to one of five issues about their background. The CIST was a multiple-issue counter-espionage format with two relevant issue questions bracketed by a

comparison on each side. There were five relevant issue questions on each chart worded as, “*Did you put false information about (each of 5 issues) on that form*”? Each CIST chart was repeated for three charts. The table below shows that inconclusive scores, not achieving a +/- 3 spot score total, were problematic. Excluding the inconclusive results, the test did correctly identify 76% of the truthful and 81% of the deceptive.

ACCURACY OF EXAMINER TEST DECISIONS USING ZONE METHOD					
Examiner Decisions					
		NDI	DI	INC	TOTAL
SUBJECTS	Truthful	16	5	5	26
	Deceptive	5	21	4	30
	Total	21	26	9	56

While overall test accuracy is central, it may be more important to evaluate whether examiners can correctly identify which issue the subject was deceptive. As an aside, the authors observe that the relevant test questions have structurally similar wording in this study. Despite this, there were 250 truthful responses to relevant questions by all participants, and thirty deceptive responses. (One of the five programmed issues for each of the

thirty deceptive participants.) Excluding the inconclusive results, 91% of the truthful answers were correctly identified. The disturbing issue is that only 63% of the deceptive answers were correctly identified. Barland (1981) clearly illustrates that a polygraph is not a “lie detector” and suggests that spot-score totals can be problematic with increased inconclusive results.

⁷Barland, Gordon H., (1981). Validation and Reliability Study of Counterintelligence Screening Test. *Polygraph*, 2012, 41(1). pp 1-17



ACCURACY OF DECISIONS OF INDIVIDUAL QUESTIONS USING ZONE METHOD					
Examiner Decisions					
		NDI	DI	INC	TOTAL
SUBJECTS	Truthful	193	18	40	250
	Deceptive	10	17	3	30
	Total	202	35	43	280

The issue of inconclusive results is further noted in Barland, Honts and Barger (1989), who conducted three experiments on mock espionage screening.⁸ In one portion of the study, 94% of the innocent subjects were correctly identified, yet inconclusive results had to be excluded. In another portion of the experiment, subjects were tested over three security violations. The authors had to exclude 24% of inconclusive test results to determine the experiment identified 79% of the innocent test subjects and 93% of the guilty subjects.

In 1995, the Department of Defense (DoD) sought to evaluate issues from previous screening studies. The DoD studies

were reported as part A⁹ and another as Part B.¹⁰ These studies were designated as the Test for Espionage and Sabotage (TES). These studies chose to use the DLC based on prior studies that supported its use. The studies also sought to use the format approach described by Reed (1995)¹¹ in using multiple presentations of two independent relevant issues within a single chart, which is structurally the same as the current DLST. These studies noted that with multiple issue test formats, the examinee does not always respond physiologically the greatest to the question they are deceptive on. As well, the TES studies were required to have a positive score for each relevant question total for a truthful determination creating issues with inconclusive results.

⁸ Barland, Gordon; Honts, Charles; Barger, Steven; 1989/03/24. *Studies of the Accuracy of Security Screening Polygraph Examinations*. DO - 10.13140/RG.2.1.1700.8803

⁹ Research Division Staff (1995a). *A comparison of psychophysiological detection of deception accuracy rates obtained using the counterintelligence scope Polygraph and the test for espionage and sabotage question formats*. DTIC AD Number A319333. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph*, 26(2), 79–106.

¹⁰ Research Division Staff (1995b). *Psychophysiological detection of deception accuracy rates obtained using the test for espionage and sabotage*. DTIC AD Number A330774. Department of Defense Polygraph Institute. Fort Jackson, SC. Reprinted in *Polygraph*, 27(3), 171–180. Research Division Staff (1995b).

¹¹ Reed, S.D. (1995). *Psychophysiological detection of deception--Single test interview*. Paper presented at the meeting of the American Academy of Forensic Science, Seattle, WA.

The DLST (TES) format is listed in the APA meta-analytic survey with reference to its use with independent relevant issue questions, and that accuracy may be improved with the use of successive hurdles. The APA footnote references appear to discuss the practice described in DLST training, where a “breakout” test for each relevant issue on the original test format is commonly taught. The APA meta-analytic survey reports a mean score average for the DLST using ESS scoring as 3.4 for truthful and -2.131 for deceptive subjects.

Since the introduction of the DLST, examiners continue to report sub-optimal spot-score totals and appear encouraged to use unique approaches to resolve test decisions. There are notable law enforcement agencies that report using grand total scoring for test formats that contain multiple issue independent relevant issue questions. This seems foreign to outsiders who report discomfort using grand total scoring for multiple-issue tests. Other law enforcement agencies have reported adopting other workaround solutions for sub-optimal scores by avoiding reporting an inconclusive result and reporting all subjects as either truthful or deceptive.

Both workaround solutions are concerning considering the previous literature and potential errors for screening. As a result, other agencies and examiners have chosen to use the DLST as a single-issue screening test format, the SIST. This is nothing more than conducting the breakout format as described in the original DLST testing process. In this situation, agencies have chosen to formulate the two relevant questions as dependent

questions around one single issue. This approach does seem to be intuitively better for grand total scoring.

Another lesser-discussed problem occurs with multiple issue test formats and varying base rates. The base rate is described as the prior probability of finding truth or deception to an issue that is being examined. With symmetrical populations of truthful and deceptive persons to a particular issue we can expect symmetrical error rates, an equal number of false positive and false negative results. When the anticipated base rate is high, such as omitting something from the application, we can expect to have more deceptive persons than truthful to this issue. When the base rate is low, such as being involved in a terrorist act, we can expect to see far more truthful persons. When we have significant asymmetrical populations, such as the terrorist example above, we should expect asymmetrical error rates. If multiple-issue testing is problematic, then mixing relevant issues with varying base rates is inviting disaster in trying to understand polygraph test accuracy and creates poor outcomes for consumers attempting to make informed risk-based decisions.

This issue can be important when discussing PCSOT testing with multiple issue formats. The literature agrees that re-offense rates for sex offenders are low. Nelson, Handler, and Thiel (2021) described their literature review as concluding that the recidivism rate for sex offenders who re-offend on supervision was around 5% to 10%. However, sex offenders are believed to have higher rates for simple supervision rule violations such as substance abuse or boundary violations.



As a result, mixing test issues with significantly differing base rates could, and most likely will, result in less desirable outcomes for the corrections or therapist decision-maker.

This is the place where polygraph examiners must be aware of how they are reporting test results and how their consumer understands their polygraph report. There are examiners who erroneously report truth and deception within a single test format. Other examiners report deception but include numerical scores with some relevant questions having positive scores. These examiners may or may not understand they are inferring a split test result by suggesting which relevant the subject responded to the most. As pointed out by the literature, the ability of an examiner to discriminate which issue the subject is deceptive on is problematic at best.

One of the contributors to this article is a retired polygraph supervisor who has conducted quality control on thousands of polygraph examinations with the DLST format. This contributor has also taught polygraph schools and polygraph associations about the use of the DLST. This contributor concludes from agency experience that it is extremely difficult to discern which issue the subject is deceptive on when mixed in the test session. In fact, this author has reported applicants often did not show the most significant response to the deceptive issue, but to other issues on the test. It was only through the process of subsequent breakout tests and post-test interviewing that conclusive determinations could be made as to which issue the subject had been deceptive.

In all this historical literature and experience, it is clear one crime means one test, yet this has rarely impacted how screening tests are given. The most straightforward and most apparent solutions to deal with issue discrimination, inconclusive results, and base rate problems, would be to use a single issue test. Using the DLST would be the most efficient of the validated formats.

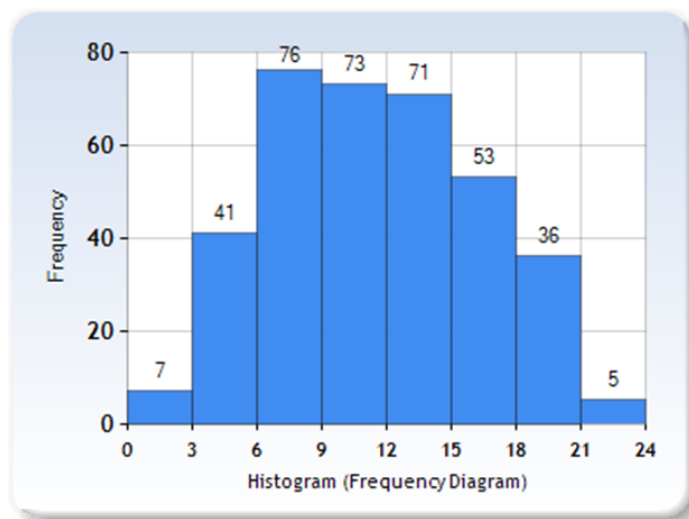
In fact, a large state police agency that has requested confidentiality evaluated their historical use of the polygraph for pre-employment screening. This agency had previously used two separate zone comparison formats, each having three independent relevant issues. A deceptive question from any of the two-zone formats would be broken out into a separate single-issue zone comparison format. This process was deemed too time-consuming for the agency's needs. This agency implemented the DLST as a multiple-issue format but had challenges meeting ESS multiple-issue scoring rules requiring a positive score for each spot score total. Time constraints by avoiding subsequent breakout formats dictated an agency policy to force a call of truth or deception, avoiding an inconclusive result.

Not satisfied, this agency decided to conduct a pilot study using the DLST as the SIST with a single issue. Each test was conducted with R1 and R2 as a single issue and repeated for three presentations of each relevant. No fourth presentation, and no breakout exams were conducted. The agency chose to conduct four test formats. The test issues were A-illegal drug use, B-Serious Crimes, C-Domestic Violence, and D-Sex Crimes. Each test had a visual mind map, and only the test

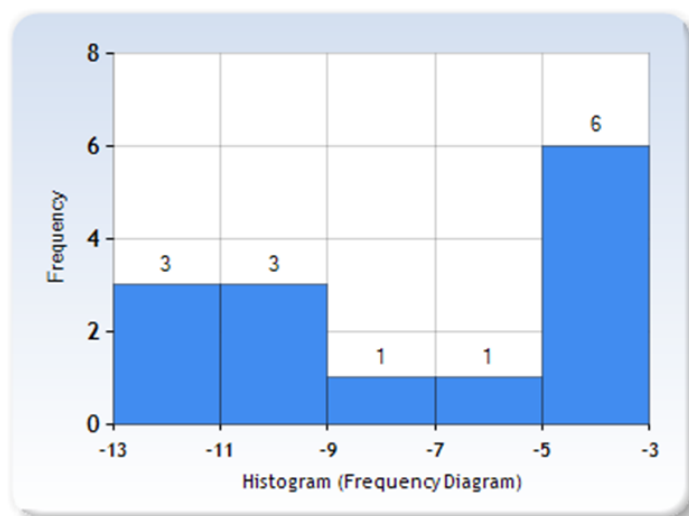


questions for that test were discussed. DLC comparisons were used as described for the original DLST. The agency chose to use ESS scoring with +2 for truthful and -4 for deceptive classifications. If a subject was found deceptive on any test, another examiner immediately conducted a quality assurance (QA) review. If the QA examiner concurred, the testing process was stopped, and a post-test interview was conducted. Any remaining tests were not conducted past this point.

This agency provided the author with the ESS scoring results from 108 applicants who submitted to a polygraph. There were 363 SIST tests with ESS hand scores meeting the truthful classification. The mean score was +10.9, with a standard deviation of 4.8. Readers wanting more information about the DLST regarding scoring, decision-making and inconclusive test results are urged to read the APA metanalytic survey.



There were 14 SIST tests with ESS hand scores meeting the deceptive classification. The mean score was -7.7 with a standard deviation of 3.4.



There were nine SIST tests with ESS hand scores that were in the inconclusive range. The mean inconclusive score was -0.7.

This agency has been very satisfied with the SIST as a single-issue screening format and has formally adopted this process as their standard for polygraph screening on applicants. The SIST has achieved higher mean scores than their use of multiple-issue test formats. The ability to discriminate per each issue has been well received and increased examiner confidence. As well, all the agency examiners have been satisfied with what seems to be fewer charts conducted and more predictable time management required per applicant.

The literature is clear that examiners wanting to conduct the most accurate polygraph test formats will select formats and questions that pertain to a single issue. While model policies include language about multiple issues and successive hurdles, in reality, this does not occur for many agencies or examiners. In crafting model policies, polygraph associations should be more aware of how model policies may lull examiners into being over-confident that they can accurately discriminate between multiple issues. As well, consumers of polygraph test results may be blinded in understanding the limitations polygraph has when examiners choose to conduct multiple-issue exams.

There may be limited utility in conducting screening exams when polygraph subjects are truthful to all the issues contained in a test format. However, if this could be predicted in advance, why would we need to administer a polygraph? Clearly, most naïve consumers do not understand what the literature says when describing that polygraph does not function as a lie detector. As well, this lack of understanding seems to drift over into the issue of what are good screening test questions.

In the interim, it would seem apparent that the polygraph profession should embrace single-issue test formats for screening purposes as the best evidence-based practice. Everyone can agree on two things, words matter and change is hard. If words matter, then words in the APA Standards of Practice and Model Policies really matter to examiners and the polygraph consumer. Refining standards and model policies to embrace single-issue testing would be the first step in that long road toward standardizing polygraph administration. However, these changes are needed to drive revised training standards and acceptance of the mitigations imposed on polygraph as a forensic science. Finally, these established changes in doctrine will provide the impetus for the profession in educating the consumers of polygraph to its limitations in testing processes and reporting.

