



Practical Polygraph: How to Parse Categorical Results for Test Questions of Diagnostic and Screening Polygraphs

By Raymond Nelson , Ben Blalock and Mark Handler

Polygraph test results, like all scientific test results, are fundamentally probabilistic – including when reduced to categorical conclusions. Categorical results of diagnostic polygraph exam are commonly expressed using the categorical terms *deception indicated (DI)* and *no deception indicated (NDI)*. These terms are a thoughtful alternative to the arcane process of reporting polygraph test results as either *deceptive* or *truthful* – terms which tend to encourage misinformed perceptions that the test has actually detected deception or truth in a physical sense, along with misguided expectations of test infallibility. Polygraph screening exams have been expressed using the terms *significant reactions (SR)* and *no significant reactions (NSR)* in an attempt to clarify that results from screening exams can be considered informative of a need for further inquiry, but are not intended to be diagnostic. As a practical matter, DI and SR are the contextual allegory for the more general scientific term

positive, and are therefore synonymous. Similarly, NDI and NSR are an allegory for the general scientific term *negative*, and are also synonymous. Although convenient, reliance on categorical test results introduces three inherent hazards to the day-to-day repertoire of discussion and thought within a professional culture.

The first of these hazards is the tendency to neglect or forgo awareness that all test results are fundamentally probabilistic. This problem is mitigated by ensuring the adequacy of professional training in probabilistic thinking, probabilistic calculations, and the relationship between probabilistic and categorical test result. A second hazard is a tendency to regard test results materialistically – as if the test result is, of itself, a physical thing. Test results are, in reality, merely a description of the strength of evidence in support of a conclusion about an unknown parameter of interest that cannot be subject to pre-



cise physical measurement or perfect deterministic observation. A third and final hazard surrounding the use of categorical test results is the potential for confusion and misunderstanding around the relationship of the overall test result and the results of the individual questions. This potential problem is largely reduced through the use of carefully structured decision rules – used to parse the categorical test result from the numerical and probabilistic data – and through a correct understanding of the differences between event-specific diagnostic polygraphs and multiple-issue screening polygraph.

There are a number of different polygraph decision rules. Among the most commonly used are the grand-total rule (GTR), sub-total-score rule (SSR), and two-stage rules (TSR). Also, the traditional Federal Zone Rule (FZR) appears to be still widely used by examiners employed by government agencies. [Refer to Nelson (2018) for more information on polygraph decision rules.] The GTR, TSR and FZR are used for diagnostic polygraph exams. Use of the SSR for diagnostic exams – intended for use as a basis of information for decision and action in response to a known incident or allegation – is difficult to justify when considering the basis of published evidence demonstrating the reduced classification accuracy for this rule with diagnostic exams. For screening polygraphs, the SSR is the most commonly used decision rule due to the increased screening sensitivity that this rule can achieve in multiple-issue

screening polygraphs. Some screening polygraphs are developed around a single behavioral target issue, and these may rely on the GTR, TSR or FZR.

All polygraph decision rules serve to clarify and specify the logic and procedure used to determine the categorical test result from the numerical and probabilistic test data. Unfortunately, although useful, polygraph decision rules provide little information about what can be said about the results of individual test questions. This paper is an attempt to describe the logic and procedure for how results of individual questions can be parsed and reported in the context of the overall categorical test result.

Event-specific Diagnostic Exams

Event-specific diagnostic polygraph exams are those exams that are conducted in response to a known incident or known allegation. The purpose of these exams is to serve as a basis of information to inform and improve decision and action in response to the known problem. Whether formulated with uniformly primary-relevant questions – pertaining to direct involvement in the behavioral issue under investigation – or with a combination of primary and secondary-relevant questions – that may describe the examinee's indirect involvement, behavioral role or level of involvement, or knowledge of material or factual evidence pertaining to a



behavioral issue – relevant questions for diagnostic examinations are non-independent. That is, all relevant questions will have a shared source of response variance – whether the examinee was, or was not, involved in the behavioral issue that is under investigation. Factors that may induce an examinee to respond to one question can conceivably have some influence on responses to other questions. Differences in action verbs are insufficient as a basis for an assumption of independence. Independence, from the perspective of scientific and statistical data analysis, requires that different test items have no shared source of variance – whatever influences each item can have no possible effect on other test items.

For diagnostic exams all relevant questions inherit the result from the examination. That is, the results of each relevant question can be reported as positive if the test result is positive. Similarly, results for each relevant question can be reported as negative if the test result is negative. Results for all relevant questions can be reported as inconclusive if the test result is inconclusive. Results should not be parsed differentially for the individual relevant questions, as this would compound the potential for error, and introduces the potential for both false-negative and false-positive results within the same exam. Table 1 shows a number of examples for the results event-specific diagnostic exams.

Table 1: event-specific diagnostic polygraph examples using GTR (ESS-M with $\alpha=.05/.05$ and cutscores +3/-3)						
Test Format	Grand Total Score	Overall Test Result	R1	R2	R3	R4
You-Phase	-3	DI	-2 (DI)	-1 (DI)	-	-
Zone	-5	DI	-4 (DI)	+1 (DI)	-2 (DI)	-
Raskin	-12	DI	-3 (DI)	+2 (DI)	-6 (DI)	-5 (DI)
You-Phase	+4	NDI	+4 (NDI)	0 (NDI)	-	-
Zone	+7	NDI	+2 (NDI)	+1 (NDI)	+4 (NDI)	-
Raskin	+8	NDI	+4 (NDI)	-1 (NDI)	+2 (NDI)	+3 (NDI)

Multiple-issue Screening Exams

Screening polygraphs are those exams that are conducted in the absence of

any known incident or known allegation. Screening exams are sometimes conducted around a single behavioral target issue. However, because the purpose of



a screening exams is to investigate the possible existence of unknown problems, screening polygraphs are often formulated with multiple behavioral target issues. Multiple-issue screening polygraphs are often formulated using primary relevant questions (i.e., questions that directly describe the examinee's involvement in the behavioral issues under investigation).

However, relevant questions for screening polygraphs, because they are intended to investigate the possible existence of unknown problems, will commonly involve a broader time-of-reference compared to the relevant questions of diagnostic exams (which tend to involve a narrowly delimited time period pertaining to an incident or allegation). Equally important, relevant questions of multiple-issue screening polygraphs are formulated with an assumption of independent criterion variance (e.g. sex, drugs, rock-and-roll). That is, it is conceivable that a person may be involved in one or more behavioral target issues while uninvolved in one or more other behavioral target issues.

Independence of test questions is the primary consideration for multiple-issue screening polygraphs. And herein exists an important complication: although the target issues of multiple-issue screening exams are assumed to have independent criterion variance the response variance for these questions is non-independent. In other words, despite the fact that the behavioral targets may be independent

or distinct all questions continue to have some shared source of response variance during testing – responses to each individual question could conceivably influence responses to the other questions. The most obvious source of shared response variance is the (limited) attention of the examinee. When a person's attention is loaded on one or more questions there may be a corresponding reduction of response to other questions.

Most importantly, attempts to differentially parse positive and negative results for individual questions within a single examination have resulted in reduced test accuracy and increased decision errors – including the potential for false-positive and false-negative errors within a single exam. It is for this reason that the practical heuristic for the SSR can be thought of as *all or any* – meaning that all relevant questions must produce negative results for a test result to be classified as negative, while a positive result for any relevant question requires that the overall test result is classified as positive. To reduce the potential for decision error – and prevent the occurrence of both false-positive and false-negative errors within a single exam – field practice standards prohibit examiners from offering both positive and negative results within a single exam.

For multiple-issue screening exams, using the SSR, the overall test result is inherited from the results of the individual relevant



questions using the worst-case-scenario. The test result is classified positive if any (one or more) relevant question has produced a positive result. If all the results of relevant questions are negative, then the overall test result is classified as negative. In practical terms the worst-case-scenario is expressed by the relevant question with the lowest numerical score – meaning the test result is classified as deceptive if the lowest subtotal score can be classified as deceptive, and the test result can be classified as truthful if that if the lowest subtotal score can be classified as truth-

ful. Because positive and negative results are not permitted within a single examination, whenever one or more questions has produced a positive result, the categorical results is meaningless and un-interpretable for all relevant questions that have not produced a positive result. In other words, a positive result for any relevant question necessitates a positive classification for the test results, and the categorical results of individual questions are inconclusive if they are not positive. Table 2 shows a number of examples for the results of multiple-issue screening polygraphs.

Table 2: multiple-issue screening polygraph examples using SSR (ESS-M with $\alpha=.05/.05$ and cutscores $(-3/+1)$					
Test Format	Test Result	R1	R2	R3	R4
AFMGQTV2 - 2RQ	SR	-3 (SR)	+3 (INC/NO)	-	-
AFMGQTV2 - 3RQ	SR	-5 (SR)	+2 (INC/NO)	+1 (INC/NO)	-
AFMGQTV2 - 4RQ	SR	-4 (SR)	0 (INC/NO)	+6 (INC/NO)	+3 (INC/NO)
AFMGQTV2 - 2RQ	NSR	+2 (NSR)	+3 (NSR)	-	-
AFMGQTV2 - 3RQ	NSR	+1 (NSR)	+5 (NSR)	+3 (NSR)	-
AFMGQTV2 - 4RQ	NSR	+2 (NSR)	+1 (NSR)	+4 (NSR)	+2 (NSR)
AFMGQTV2 - 2RQ	INC/NO	0 (INC/NO)	+1 (NSR)	-	-
AFMGQTV2 - 3RQ	INC/NO	+2 (NSR)	-1 (INC/NO)	+4 (NSR)	-
AFMGQTV2 - 4RQ	INC/NO	+1 (NSR)	+4 (NSR)	+1 (NSR)	-1 (INC/NO)

Summary

For agencies with policies that prohibit the reporting of test results for individual relevant questions there is little or no need

for concern about the results of individual test questions. Polygraph field practitioners who work in the absence of such restrictive policies may find themselves



fraught with phone calls, emails, and other inquiries about the results of the individual relevant questions – and this can lead to erroneous or chaotic communication among referring professionals and other consumers of polygraph test results. Most importantly, referring professionals may be vulnerable to misunderstanding or manipulation of the examination result if the report does not contain explicit and correct information about what may be reasonably said of the examinees' deception or truthfulness in response to each of the relevant test questions.

For example, an overly-optimistic professional might interpret the absence of explicit question-level documentation as an indication of that a test does not indicate deception the individual relevant questions – that an examinee may have been truthful to these questions – even though the results of an event-specific diagnostic polygraph is reported as DI. Or, in the case of a multiple-issue polygraph, a naively optimistic reader might attempt to conclude that an examinee has *passed* any of the relevant questions that are not explicitly reported as SR.

Parsing the correct interpretation for individual relevant questions will serve to affirmatively document and assertively communicate the relationship between the overall test result and the results for individual relevant questions. Reporting test results in this manner will reduce the potential for misunderstanding and manipulation of the results of event-specific diagnostic exams, and allows for documentation and reporting of which relevant questions have produced the greatest loading of changes in physiological activity – and which questions have been used to parse the categorical test result – among relevant questions of multiple-issue screening polygraphs. In short: all individual questions of event-specific diagnostic polygraphs should inherit the same result as the examination, while the examination result for multiple-issue screening polygraphs are parsed from the results of the individual relevant questions using the any-or-all heuristic and the caveat that positive and negative conclusions are not permitted within a single examination.

