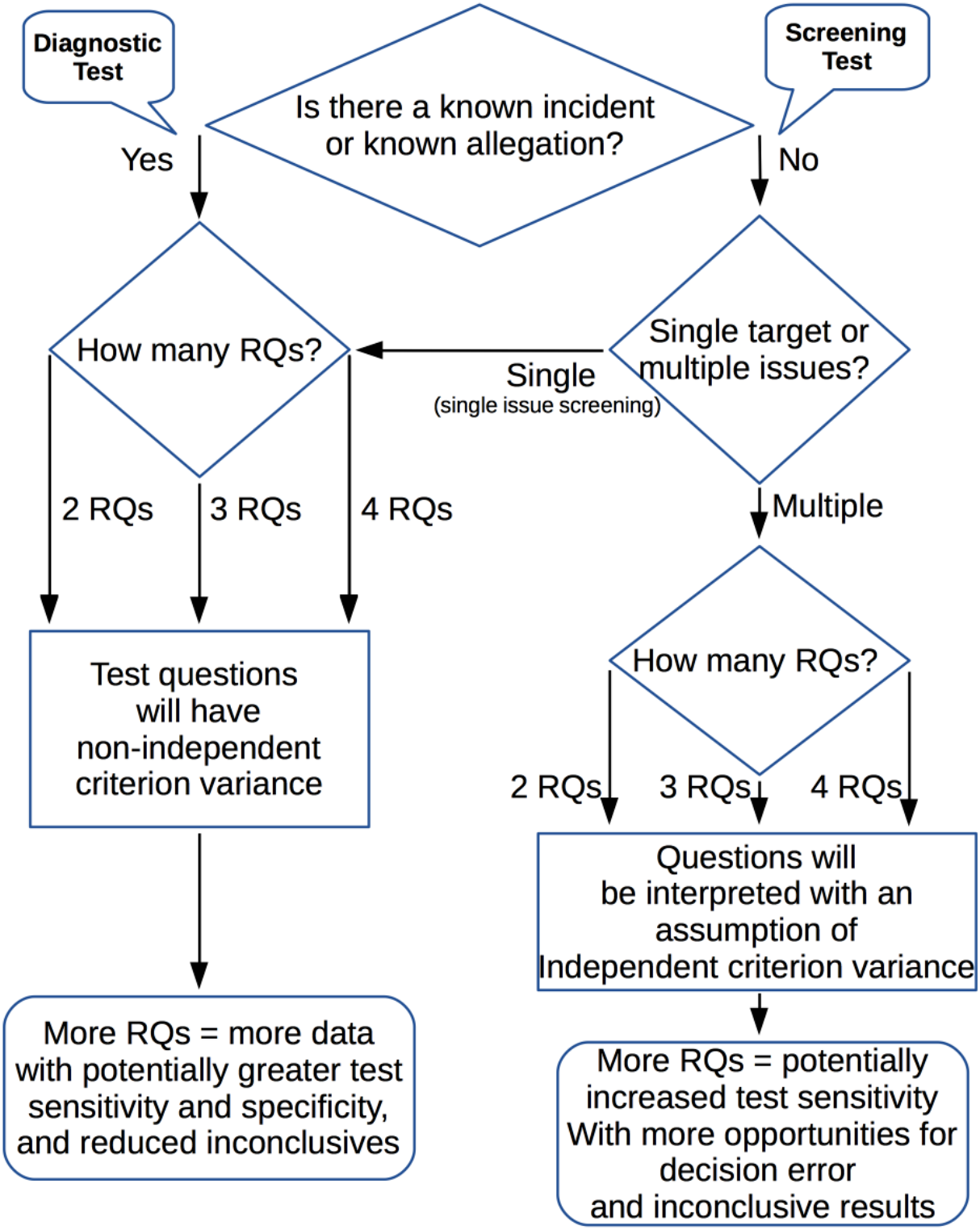
Practical Polygraph: How to Select a Polygraph Test Format

Raymond Nelson[1,3] and Mark Handler[2,3]



Selection of a polygraph test format should be a rational process based on an evaluation of the case referral information, the needs of the referring professional, and our scientific knowledge about polygraph techniques. In decades past, the traditional way of selecting a polygraph format may have been to use the technique that carried the name of the founder of the polygraph school at which an examiner received one's training. Use of named polygraph techniques served the profession well for a time – especially inasmuch as it allowed for the recognition and discussion of standardized procedures that that could be studied for the fundamental principles that influence their effectiveness. But field practitioners for a time seemed to harbor an unrealistic, almost mythological, expectation that each different named polygraph technique were sacred, as if they were based on fundamentally different scientific principles, along with an irrational attitudes about the potential value of field practices and scientific knowledge gleaned from professionals trained at other polygraph schools. This amounted to a form of indoctrination or professional imprinting in which the test format was selected as a matter of memorization or habit, with little thought or awareness about the fundamental principles that govern the effectiveness of the test question format.

Although many of the several polygraph techniques in use today have begun to appear more similar than different, some older polygraph techniques that remain in use today still retain aspects and elements of their original design – even though some of the hypothesized design advantages have been inconsistent with scientific evidence. Polygraph examiners who make use of these techniques are required to pay diligent attention to the these original design elements, regardless of their inconsistency with scientific evidence, or risk criticism and the suggestion that they have done something "invalid."

At the present time, over-reliance on named polygraph technique has begun to contribute more to stasis that to progress and professionalism, as polygraph field practitioners have experienced sometimes great difficulty engaging in rational discussion

---

1 Raymond Nelson is a psychotherapist and polygraph examiner who works as a polygraph research specialist for Lafayette Instrument Company (LIC) . Mr. Nelson is a past-president and currently serves as an elected member of the Board of Directors for the American Polygraph Association (APA). Mr. Nelson is one of the developers of the OSS-3 and ESS models for the analysis of polygraph test data and has published numerous papers on many aspects of the polygraph test.

2 Mark Handler is police polygraph examiner who has published numerous studies on many aspects of the polygraph test. Mr. Handler teaches regularly at several polygraph training programs, and is employed by Converus, Inc., a company that markets credibility assessment technology. Mr. Handler currently serves as the editor-in-chief for the APA.

3 There are no financial or proprietary interests associated with this publication. The views and opinions expressed herein are those of the authors, and not necessarily those of LIC, APA, or Converus Inc.

of polygraph techniques in terms of basic principles of science and scientific testing – not because of the absence of scientific foundations for the polygraph test, but due instead to the continued emphasis on traditional concepts that have outlived their usefulness. In recent years the polygraph profession's continued over-reliance on named polygraph techniques has served only to increase the difficulty of coherent discussion and shared learning among polygraph professionals and between polygraph professionals and other areas of applied science and forensics.

The antidote to professional stasis is simple: a) avoid reliance on dogmatic notions that are not based on replicable and reproducible scientific evidence, b) make use of all available generalizable scientific information, and c) interact synergistically with other areas of basic and applied science. At a practical level, polygraph professionals will be able to engage in more rational discussion among themselves and with professionals from other areas of science and forensics by moving beyond the dogmatic tradition of named polygraph techniques, and by emphasizing the basic principles of polygraph scientific testing. This can only occur with the recognition that many different named polygraph techniques are in fact constructed of similar scientific principles and are in fact not fundamental different from each other – despite the differences in name. At this time the formal name of the polygraph technique adds only confusion, not clarity, to our thinking and to our discussions about test validity.

Although the topic of scientific validity is a deep and complex discussion, we can simplify it here by emphasizing two important points: 1) construct validity, which refers to the validity of the underlying mechanisms and principles that determine the test effectiveness, and 2) criterion validity, referring to the degree to which whether the test results and conclusions correspond to the external criterion of interest to the test (i.e., the actual deceptive or truthful state of the examinee's answers). Construct validity is intertwined with the scientific theories that govern whether the test functions as expected, and also help us to understand the potential applications limitations of a test. Criterion validity has obvious practical value. Polygraph field practitioners will be primarily concerned with the validity of the applied or analytic theory of the polygraph test, which is focused on the way in which we expect the reported physiological data to conform to statistical reference models that can enable us to make practical and effective probabilistic classifications of deception and truth-telling. The analytic theory of the polygraph test holds that greater changes in physiological activity are loaded at different types of test stimuli as a function of decep-

tion or truth-telling in response to the investigation target stimuli (American Polygraph Association, 2011; Honts & Peterson, 1997; National Research Council, 2003; Nelson, 2016; Office of Technology Assessment, 1983; Senter et al., 2010).

The tradition of named polygraph techniques may have served the polygraph profession well in its early years. Named polygraph techniques enabled the recognition and discussion of ideas and let to the formulation of standardized practices that could be studied for the validity of their underlying hypotheses. At the time of the Meta-Analytic Survey of Validated Polygraph Techniques (APA, 2011), the authors of that report began to advise that the tradition of named polygraph techniques has outlived its usefulness. In other words, the validity of a polygraph technique is not fundamentally determined by or associated with its name. This became even clearer when members of the APA Research Committee (Nelson, Handler, Oelrich & Cushman, 2014) responded to inquiry about the use of a traditionally multi-issue test format as a single-issue event-specific diagnostic polygraph format without symptomatic questions, and concluded that the present scientific knowledge base and presently available normative reference models would be sufficient to enable us to evaluate the data and interpret the results.

That rational and useful discussion could be engaged without the use of named techniques was further demonstrated by Nelson and Handler (2015), who published statistical reference tables for commonly used polygraph formats and commonly used analysis methods without the use or emphasis on named techniques. Instead, statistical reference tables were described in terms of whether the results would be interpreted with or without an assumption of independent criterion variance, together with information about the number of target questions and the method of analysis.

The premise of this paper is that for both practical and scientific purposes, the selection of a polygraph technique should be based on the known advantages and disadvantages associated with the scientific principles that encompass the structure of a polygraph test format. What then are the fundamental issues that define a polygraph test format?

Polygraph test formats can be easily understood with two questions:

1) Does the case referral information indicate a need for a screening or diagnostic test, and

2) How many relevant questions will be used?

The first question is simply a matter of the presence or absence of a known

incident or allegation. Diagnostic tests, also known as event-specific[1] tests, are done in response to a known incident or allegation. The purpose of a diagnostic test is to provide a basis of information to improve decisions and actions that will be taken in response to a known problem. Screening tests are those tests that are done in the absence of a known incident or allegation.

If there is a known incident or allegation then a diagnostic test is needed. Because our decisions and actions are expected to have some effect on the future of another person, there is an ethical obligation to select a diagnostic polygraph format that will reduce the level of uncertainty and improve our decision making to the greatest extent possible. As shown in a meta-analytic survey of validated polygraph techniques (APA, 2011) polygraph formats that make use of the grand-total score have consistently outperformed test formats that emphasize the use of subtotal scores as the basis for decision making. Polygraph test formats that use the grand-total scores were shown to provide equal or greater test sensitivity and specificity levels, along with lower false-positive and false-negative rates, and fewer inconclusive results. Use of the grand-total score is premised on an assumption that the criterion variance does not vary independently for different questions within a diagnostic test. In other words, all target questions of diagnostic tests will address the examinee's behavioral involvement in a single known incident or allegation.

Screening tests are those test that are conducted in the absence of any known incident or allegation [2,3,4,5,6,7,8,9,10]. These tests may involve a single

---

1 These have also been referred to as specific-issue and single-issue tests, though these terms are ambiguous as to whether they refer to a single issue diagnostic test or single issue screening test. Hence the term event-specific is thought to more clearly refer to the diagnostic investigation of a known allegation or known incident.

2 ASTM (2012) E2035 -12, section 3.1. "A PDD examination in which the relevant issues are not related to a known event, and during which more than one issue can be addressed within the same test series... Examples include applicant and counterintelligence screening as well as some forms of PCSOT."

3 APA (2015) Standards of Practice section 1.1.6. "A polygraph examination conducted in the absence of a reported incident or allegation. Screening exams may be conducted as single issue or multiple issue exams."

4 APA (2011) Meta Analytic Review, footnote 5, page 205. "Screening tests are any tests conducted in the absence of a known problem, and are intended to search for possible problems. Screening tests, because of the absence of any known problems, and because of interest in several types of possible problems, are often constructed around multiple issues. The terms multi-issue and mixed-issue are used interchangeably. It is not the number of issues that defines the distinction between diagnostic and screening tests, but the presence or absence of a known problem."

target issue, though they usually involve multiple issues of concern. Multiple issue screening tests are thought to have the advantage of potentially increased test sensitivity to a wider range of topics of concern to the screening context. The trade-off or disadvantage for these exams is that there is some increased risk of testing error and inconclusive results due to the effects of multiplicity[11] when decision making is based on probability scores for multiple target issues.

The number of relevant target issues is important to both diagnostic and screening polygraph testing contexts, though for different reasons. Use of more relevant target questions in the diagnostic context will mean more data upon which to base a single probabilistic test result that can become the basis for the categorical test result. More available data will lead to generally smaller errors of measurement and to generally greater test precision. This can be observed in the

---

5 Krapohl, Handler & Sturm (2012). Terminology reference 3rd edition, page 77. "A polygraph examination conducted in the absence of a reported issue or allegation to investigate whether an examinee has withheld information regarding engagement in behaviors encompassed by the relevant questions that cover specified periods of time... Screening examinations may be designed to investigate both multiple and single types of behavior."

6 National Research Council (2003), page 1. For (employee) screening, there is no specific event being investigated and the questions must be generic.

7 Handler, Honts, Krapohl, Nelson & Griffin (2009), page 72. "Unlike diagnostic tests, which are used for criminal investigation polygraphs, screening examinations are conducted in the absence of any known incident or allegation."

8 Meijer, E., Verschuere, B., Merckelbach, H., & Crombez, G. (2008), page 8. "...specific incident polygraph tests used in a known incident. In screening the examiner does not know whether an incident took place."

9 Iacono, W. (2007). p. 688. "Screening procedures are typically employed by the government and private agencies to detect security risks. For these procedures, it is not known whether a particular incident has taken place…"

10 Crewson, P. E. (2003). page 60. "Screening applications involve the use of an assessment tool on a general population in which there is no specific evidence of disease. Diagnostic correlates with the polygraph specific-issue test and is reserved for studies where there is prior evidence a condition exists, such as when a test is ordered after a clinical examination of a patient suggests an abnormality.

11 Multiplicity refers to the compounded probability of error when making multiple simultaneous or repeated statistical decisions. These effects can be reduces, though not completely eliminated, by the use of statistical corrections.

published studies on polygraph criterion accuracy (APA, 2011) for which diagnostic polygraph formats with more relevant target issues are observed to have significantly lower inconclusive rates compared to formats with fewer relevant target questions.

In the polygraph screening context, use of more relevant target issues is associated with increased risk for testing error and inconclusive results due to a well-known statistical phenomena referred to as multiplicity. Increased risk for testing error does not preclude the use of multiple issue tests in screening situations (i.e., where there is no known incident or allegation for which some action is required). Some testing contexts may determine that the increase in test sensitivity from using multiple test target issues will serve their operational objectives and mission priorities. Results of screening tests are commonly evaluated along with other information before proceeding with decisions and actions in the screening context. Circumstances that warrant greater testing precision may benefit from a reduction of the number of target issues in the test question format. It is a matter of both science and ethics that multiple issue tests are not used in diagnostic polygraph contexts. It is difficult to imagine some ethical justification for the selection of a sub-optimal testing format in a diagnostic context – where the purpose of the test is to inform and improve decision making in response to a known allegation or incident.

Summary

Together with the selection of the method for test data analysis, answers to these two questions will predictably influence the effectiveness of a polygraph test format. These questions will also influence the selection of the statistical reference model that will be used to calculate the probability results that can become a basis for a categorical test result.

To summarize:

- **Diagnostic tests are those tests that are conducted in response to a known incident or allegation.**
- **Relevant target questions for diagnostic tests should be non-independent, making sensible use of the grand-total score as a basis for conclusions.**
- **Use of more relevant questions is associated with potentially increased test sensitivity and specificity and reduced inconclusive results for diagnostic exams.**
- **Screening tests are those tests that are conducted in the absence of any known incident or allegation.**

- **Screening tests can address single issue or multiple issues of concern to the screening context.**

- **Use of subtotal scores as a bases for decision making for multiple issue exams is premised on an assumption of that the criterion variance of the relevant target questions may vary independently[12].**

- **Due to the effects of multiplicity, use of more relevant target questions in multiple issue screening exams is associated with potentially greater opportunity for decision error and inconclusive results.**

We recommend that polygraph examiners and polygraph training programs begin to emphasize the notion of polygraph test formats in terms of the basic scientific principles that govern the test effectiveness. We further recommend that the polygraph profession begin to outgrow the tradition of named techniques, as continued reliance on named techniques serves to distract attention away from important scientific principles. Continued use of named techniques will place the polygraph profession at risk for becoming stuck in "traditions" that may represent the state of the science of polygraph at the time of their origin but may be inconsistent with scientific evidence today. Emphasis and attention to basic scientific principles associated with polygraph test formats will help polygraph examiners to avoid becoming limited to the scientific status of polygraph testing at the time of their original training, and will enable them proceed into the future while making use of generalizable scientific knowledge as it applies to the polygraph test as a method for probabilistic quantification and decision making under uncertain conditions involving deception and truth-telling.

12 Recorded physiological responses and test data for multiple issue exams is not independent because these responses have a shared source of variance in the form of the examinee. For this reason, standardized polygraph field practices preclude the occurrence of both positive and negative results within an exam, and instead emphasize a single categorical test result that is the composite of the results of the multiple issue target questions.

References

American Polygraph Association (2011). Meta-analytic survey of criterion accuracy of validated polygraph techniques. Polygraph, 40(4), 196-305. [Electronic version] Retrieved August 20, 2012, from http://www.polygraph. org/section/research-standards-apa-publications.

APA. (2015). Standards of Practice. American Polygraph Association. Chattanooga, TN.

American Society for Testing and Materials. (2012). Standard Terminology Reference for Forensic Psychophysiology. (E2013-12). West Conshohocken, PA. ASTM International.

Crewson, P. E. (2001). A Comparative Analysis of Polygraph With Other Screening and Diagnostic Tools. Research Support Service. Report No. DoDPI01-R-0003. Reprinted in Polygraph (32) 57-85.

Handler, M., Honts, C., Krapohl, D., Nelson, R. & Griffen, S. (2009). Integration of Pre-employment Polygraph Screening into the Police Selection Process. Journal of Police and Criminal Psychology, 24, 69-86.

Honts, C. R. & Peterson, C.F. (1997). Brief of the Committee of Concerned Social Scientists as Amicus Curiae United States v Scheffer. Available from the author.

Iacono, W. (2007). Detection of Deception. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), Handbook of Psychophysiology. Cambridge: Cambridge University Press,

Krapohl, D., Handler, M., & Sturm, S. (2012). Terminology Reference for the Science of Psychophysiological Detection of Deception. 3rd Edition. American Polygraph Association. Chattanooga, TN.

Meijer, E., Verschuere, B., Merckelbach, H., & Crombez, G. (2008) Sex offender management using the polygraph: a critical review. International Journal of Law and Psychology, 31(5), 423-429.

National Research Council (2003). The Polygraph and Lie Detection. National Academy of Sciences.

Nelson, R. (2016). Scientific (analytic) theory of polygraph testing. APA Magazine, 49(5), 69-82.

Nelson, R. & Handler, M. (2015). Statistical reference distributions for comparison question polygraphs. Polygraph, 44(1), 91-114.

Nelson, R., Handler, M., Oelrich, M. & Cushman, B. (2014). APA research committee report: proposed usage for an event-specific AFMGQT test format. Polygraph, 43(4), 155-167.

Office of Technology Assessment (1983). The validity of polygraph testing: A research review and evaluation. [Re-printed in Polygraph, 12, 198-319.].

Senter, S., Weatherman, D., Krapohl, D., & Horvath, F. (2010). Psychological set or differential salience: a proposal for reconciling theory and terminology in polygraph testing. Polygraph, 39(2), 109-117.