Validated Polygraph Techniques[1]

Donald J. Krapohl

Abstract

There are many polygraph techniques currently used in the field. Much research has taken place over the past 30 years that has attempted to validate at least a portion of those techniques. The present article attempts to encapsulate the findings of the research. Methods that have replicated research support are identified, and mean accuracies across studies are calculated. Implications for technique development are discussed.

Introduction

One of the most-asked questions from experienced polygraph examiners is: which polygraph techniques are "validated"? Understandably, examiners want to use the most accurate techniques available and with today's more educated examiner, the focus has shifted more toward science than in years past. Because to date no list of validated methods has been published, examiners are left to employ more informal methods of selection.

How, then, do polygraph examiners make these important decisions about the technique they will practice on the public? There is more than one answer to this question. For most examiners it's pretty simple: we choose our technique the same way a duck chooses his mother – it was the first thing we saw. Not everyone is so fixated, of course. A venturesome minority may begin to use a technique they learned at a seminar or read about in a professional publication. Least often and least desirable, a hardy few become enamored with techniques they've devised themselves, methods often based on personal experiences or simple hypotheses about the mechanisms of psychophysiology or psychometrics.

None of these approaches can be called scientific, and all of them are vulnerable to a host of systematic errors. It may be acknowledged that many techniques have been used for years by perhaps many hundreds of polygraph examiners, who take this fact as proof of validity. Popularity should not be confused with validity, however (consider the lesson of astrology). Some methods appear to be effective in eliciting confessions, but neither is this a measure of validity. Nor should public endorsements or self-endorsements from individuals by themselves satisfy the requirement. Validation is a careful process, having no shortcuts, and it allows us to have a level of confidence in the methods we use.

For clarity, scientific validation of a technique will be defined here as the existence of replicated and published research which found the technique to be accurate. What is "accurate"? While the scientific threshold for validity is often set at anything with a robust effect above chance level, the threshold according to standards of the American Society for Testing and Materials (ASTM, 2005) is 90% for evidentiary polygraph techniques and 80% for investigative polygraph techniques, inconclusives excluded. Both evidentiary and investigative polygraph techniques are permitted an inconclusive rate of up to 20% of all cases. By way of definition, evidentiary polygraph examinations are those conducted specifically for courtroom purposes.

[1]The author wishes to express his appreciation to Gordon Barland Ph.D. and Stuart Senter Ph.D. for their very insightful comments and suggestions to an earlier draft. The views expressed in this article are solely those of the author, and do not necessarily represent those of the Department of Defense, the US Government or the APA. Comments and reprint requests should be sent to dkrapohl@aol.com.

Investigative polygraph examinations are used for non-judicial purposes, such as applicant testing, sex offender management, criminal investigation, and counterintelligence screening to name a few. Because no other validity standard exists in the profession, the ASTM minimums will be used here. Of the many polygraph techniques that have undergone the validation process, not all of them meet the exacting ASTM standards. The validity research is covered later in this paper.

The validation process is necessarily slow and meticulous, and does not always lead to the conclusion that a technique is valid. For example, the putative relationship between the vocal micro-tremor and deception has undergone the validation process. There are many scientific studies, and by consensus they show low or no validity for this approach to deception detection (National Research Council, 2003). This inconvenient fact hasn't deterred the marketers of voice stress devices, however, who seem to roll out new versions fairly regularly. To coin a phrase, voice stress appears to be promising.....and promising and promising. Nevertheless, under the cold eye of scientific inquiry, it has yet to make good on its promises.

Contrast voice stress analysis with some of the most commonly used polygraph techniques. Many polygraph techniques are not supported by good research, but because they share principles that have been confirmed during the research of similar techniques, they are almost certainly valid techniques (more about this later). In other words, these polygraph techniques are probably valid, but have not gone through the validation process. I bring this to the attention of the reader so that the subsequent information on validated techniques can be given context and proper weight. For the research to be included in the present summary, the following criteria, which I believe to be reasonable, had to be met:

1. The research had to be published in full.

2. The research had to be replicated.

3. The published polygraph technique had to be identified by name or reported in sufficient detail so that the correct name for the technique could be determined.

4. When multiple techniques were reported, accuracy figures had to be available for each technique.

5. The accuracy figures had to be broken out separately for truthful and deceptive cases.

6. Ground truth criteria must have been independent of the polygraph results.
7. The testing and scoring technique must have been representative of field practices.

8. Field cases must have been randomly selected, or with laboratory studies, subjects must have been randomly assigned to either deception or non-deception conditions.

9. The formulation of decisions of deception or truthfulness on individual cases could not consider the results of other examinations on the same crime.

10. For laboratory data, programmed countermeasure cases were excluded.

The list below shows the polygraph techniques meeting the criteria above. Listed first are the unweighted averages of the true negative, true positive, and inconclusive rates for these studies, followed by the supporting research citations. It should be noted that the accuracy figures are based on human decisions rather than algorithm decisions. While current evidence suggests that some algorithms perform better than the average human scorer on single-issue examinations, algorithms are not available for all techniques. When algorithm outcomes were reported in the research along with human decisions, only human decisions were used so to afford an apples-to-apples comparison among techniques.

Examiners may find that their preferred method did not meet the definition of "validated." However, they may draw comfort that this list is current only up to the date of publication, and it is certain to grow in the future. The twin goals of this article are to

inform polygraph examiners of the existing state of the field, and to encourage more research that could advance it.

Army Modified General Question Technique (MGQT)

Unweighted mean accuracy

Deceptive cases (N = 168): 97% correct without inconclusives. 7% inconclusive.

Truthful cases (N = 60): 25% correct without inconclusives. 35% inconclusive.

Overall: 61% correct without inconclusives. 21% inconclusive.

Citations

Blackwell (1998).
Krapohl & Norris (2000).
Podlesny & Truslow (1993).

Concealed Information Test (CIT, AKA Guilty Knowledge Test)

Unweighted mean accuracy

Deceptive cases (N = 843): 76% correct. Inconclusives are generally not allowed.

Truthful cases (N = 404): 83% correct. Inconclusives are generally not allowed.

Overall: 80% correct.

Citations

There are literally scores of studies using the Concealed Information Test. For an excellent review of the literature (with citations) and a meta-analysis of 50 published data sets, see MacLaren (2001).

Federal Zone Comparison Technique (AKA Army ZCT)

Unweighted mean accuracy

Deceptive cases (N = 141): 97% correct without inconclusives. 9% inconclusive.

Truthful cases (N = 110): 82% correct without inconclusives. 23% inconclusive.

Overall: 89% correct without inconclusives. 16% inconclusive.

Citations

Blackwell (1998).
[1]Krapohl (2005).
[2]Yankee, Powell, & Newland (1985).

[1] Investigative (traditional) decision rules used
[2] Experienced group data used.

Reid Technique

Unweighted mean accuracy

Deceptive cases (N = 88): 88% correct without inconclusives. 7% inconclusive.

Truthful cases (N = 88): 78% correct without inconclusives. 5% inconclusive.

Overall: 83% correct without inconclusives. 6% inconclusive.

Citations

[3]Jayne (1990).
[4]Horvath (1977).
[3]Horvath (1988).

[3] Reid method data used only.
[4] Verified cases only.

Relevant-Irrelevant (RI) Screening Test

Unweighted mean accuracy

Deceptive cases (N = 79): 90% correct without inconclusives. 0% inconclusive.

Truthful cases (N = 61): 73% correct without inconclusives. 0% inconclusive.

Overall: 83% correct without inconclusives. 0% inconclusive.

Citations

Correa & Adams (1981).
Krapohl, Senter, & Stern (2005).

Test for Espionage and Sabotage

Unweighted mean accuracy

Deceptive cases (N = 65): 83% correct without inconclusives. 0% inconclusive.*

Truthful cases (N = 119): 93% correct without inconclusives. 3% inconclusive.*

Overall: 88% correct without inconclusives. 2% inconclusive.*

* The TES protocol permits retesting when an initial series is found inconclusive.

Citations

Research Division Staff (1995a).
Research Division Staff (1995b).

Utah Zone Comparison Technique

Unweighted mean accuracy

Deceptive cases (N = 116): 92% correct without inconclusives. 12% inconclusive.

Truthful cases (N = 116): 89% correct without inconclusives. 11% inconclusive.

Overall: 91% correct without inconclusives. 12% inconclusive.

Citations

[5] Honts, Hodes, & Raskin (1985).
[5] Honts, Raskin, & Kircher (1987).
[5] Honts, Raskin, & Kircher, J.C. (1994).
[6] Kircher, & Raskin, (1988).
Raskin, & Hare, (1978).

[5] Programmed countermeasure cases excluded.
[6] Human scoring condition only.

Recall that according to ASTM standards for evidentiary and investigative polygraph techniques, examinations for evidentiary purposes require a minimum accuracy of 90% without inconclusives, and an inconclusive rate overall of 20% or less, and investigative examinations have a lower standard; 80% accuracy without inconclusives and an overall inconclusive rate of 20% or less. According to these standards, only the Utah Zone Comparison Technique is sufficiently researched and valid for evidentiary purposes (See Table 1). It is worth noting that the Federal ZCT fell below the threshold by a single percentage point.

For investigative examinations, the list of validated methods would include the Federal Zone Comparison Technique, the Reid Technique, the Concealed Information Test, the Relevant-Irrelevant Screening Test, and the Test for Espionage and Sabotage. The research on the Army Modified General Question Technique (MGQT) did not indicate an accuracy sufficiently high for either category of examination, and therefore it should not be used as a standalone technique. In other words, the MGQT could be employed in a screening-type application to guide an examiner as to where to focus attention, but it should be followed up with a technique that provides sufficient validity.

Table 1. Rank order of polygraph techniques by accuracy (excluding inconclusives).

Technique	Accuracy without Inconclusives	Inconclusive Rate
Utah ZCT	91%	12%
Federal ZCT	89%	16%
TES	88%	2%
RI	83%	0%
Reid	83%	6%
CIT	80%	0%
MGQT	61%	21%

As a closing comment, a word about validated *techniques* versus validated *principles*. In 2002, the American Polygraph Association (APA) Board of Directors undertook the task of developing a list of acceptable techniques, that is, those that the APA could support because of the research. The task turned out to be more complicated than anticipated. For example, there are

several techniques that are highly similar, but a technique by one name received the research while the others did not. Does this mean that one technique is valid while the other is not? Also, some techniques had evolved over time to forms that were not identical to that tested in the validity research. How much this may have affected validity is unknown. The literature search also turned up some research that was transparently self-serving and of questionable value. Other studies were of very poor design, and many reports used to bolster validity claims were never published. The task of cataloging valid techniques ultimately made clear what was really important about validation: valid principles.

Valid principles, it can be agreed, are the building blocks of valid techniques. Therefore, if one knows which principles are valid and which are invalid, development of valid techniques is a straightforward process. Subsequent reading of the research literature suggests that there are several important principles that can be relied upon. Here is an incomplete list:

- Single-issue testing is more accurate than multiple-issue or multiple-facet testing.
- "Successive hurdles" can increase decision accuracy in multiple-issue testing.
- Two-stage decision rules produce fewer inconclusives than one-stage decision rules. The proportions of correct and incorrect decisions are not affected.
- The total chart minutes concept is false.
- Changing cutoffs by themselves will merely affect the types of decision errors.
- Decision rules can be set to minimize the cost of errors for a particular application.
- Exclusionary comparison questions do not improve decision accuracy over non-exclusionary probable-lie questions.
- There are approximately 12 tracing features that are valid for chart interpretation.
- Some computer algorithms outperform most human scorers in blind scoring of single-issue examinations.
- Highly complex scoring rules can reduce human scoring reliability, which can, in turn, erode decision accuracy.
- On average, deceptive examinees react stronger to relevant questions than truthful examinees react to probable-lie questions. Decision rules can be adjusted to compensate for this imbalance.
- Relevant questions immediately preceded by an irrelevant questions produce significantly lower scores than relevant questions preceded by comparison questions. This is true for both truthful and deceptive examinees.
- The value of symptomatic questions to reduce inconclusives is highly questionable.
- An acquaintance test given before other testing improves decision accuracy.
- The data channels shown to contribute to decision accuracy are: respiration, electrodermal activity, blood volume (cuff), and vasomotor (plethysmograph).
- The Utah 3-to-5 chart rule reduces inconclusives. The proportions of correct and incorrect decisions are not affected.

Polygraph techniques in many shapes and sizes could be assembled from valid principles, including those above. The use of valid principles brings with it significant benefits. They could be used as benchmarks to help avert professional disagreements that involve non-critical differences between techniques, and to help identify deficient techniques. If the profession were to adopt this approach to technique development and abandon the "science-lite" methods of the

past, we could find the field moving toward higher accuracy, fewer disagreements, and more credibility.

References

American Society for Testing and Materials. (2005). *Standard Guide for the Conduct of PDD Screening Examinations.* Available online at: ASTM.org.

Blackwell, N.J. (1998). *PolyScore 3.3 and psychophysiological detection of deception examiner rates of accuracy when scoring examination from actual criminal investigations.* DTIC AD Number A355504/PAA. Department of Defense Polygraph Institute, Ft. McClellan, AL. Printed in *Polygraph, 28*(2) 149-175.

Correa, E.J., & Adams, H.E. (1981). The validity of the pre-employment polygraph examination and the effects of motivation. *Polygraph, 10*(3), 143-155.

Jayne, B.C. (1990). Contributions of physiological recordings in the polygraph technique. *Polygraph, 19*(2), 105-117.

Honts, C.R., Hodes, R.L, & Raskin, D.C. (1985). Effects of physical countermeasures on the physiological detection of deception. *Journal of Applied Psychology, 70*(1), 177-187.

Honts, C.R., Raskin, D.C., & Kircher, J.C. (1987). Effects of physical countermeasures and their electromyographic detection during polygraph tests for deception. *Journal of Psychophysiology, 1*(3), 241-247.

Honts, C.R., Raskin, D.C., & Kircher, J.C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology, 79*(2), 252-259.

Horvath, F.S. (1977). The effect of selected variables on interpretation of polygraph records. *Journal of Applied Psychology, 62*(2), 127-136.

Horvath, F.S. (1988). The utility of control questions and the effects of two control question types in field polygraph techniques. *Journal of Police Science and Administration, 16*(3), 198-209.

Kircher, J.C., & Raskin, D.C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *Journal of Applied Psychology, 73*(2), 291-302.

Krapohl, D.J. (2005). Polygraph decision rules for evidentiary and paired-testing (Marin Protocol) applications. *Polygraph, 34*(3) 184-192.

Krapohl, D.J. & Norris, W.F. (2000). An exploratory study of traditional and objective scoring systems with MGQT field cases. *Polygraph, 29*(2), 185-194.

Krapohl, D.J., Senter, S.M., & Stern, B.A. (2005). An exploration of methods for the analysis of multiple-issue relevant/irrelevant screening data. *Polygraph, 34*(1), 47-61.

MacLaren, V.V. (2001). A qualitative review of the Guilty Knowledge Test. *Journal of Applied Psychology, 86*(4), 674-683.

National Research Council. (2003). *The Polygraph and Lie Detection.* Committee to Review the Scientific Evidence on the Polygraph, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Podlesny, J.A., & Truslow, C.M. (1993). Validity of an expanded-issue (modified general question) polygraph technique in a simulated distributed-crime-roles context. *Journal of Applied Psychology, 78*(5), 788-797. Reprinted in *Polygraph, 23*(3), 195-218.

Raskin, D. C., & Hare, R. D. (1978). Psychopathy and detection of deception in prison in a prison population. *Psychophysiology, 15*, 126-136.

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 McClellan, AL. Printed in *Polygraph, 26*(2), 79-106.

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 McClellan, AL. Printed in *Polygraph, 27*(1), 68-73.

Yankee, W.J., Powell, J.M., III, & Newland, R. (1985). An investigation of the accuracy and consistency of polygraph chart interpretation by inexperienced and experienced examiners. *Polygraph, 14*(2), 108-117.