ELSEVIER

Physiology & Behavior xx (2005) xxx – xxx

PHYSIOLOGY & BEHAVIOR

Integrated zone comparison polygraph technique accuracy with scoring algorithms

Nathan J. Gordon [a,*], Feroze B. Mohamed [b], Scott H. Faro [b], Steven M. Platek [c], Harris Ahmad [b], J. Michael Williams [c]

[a] *Academy for Scientific and Investigative Training, Philadelphia, PA 19103, United States*
[b] *Department of Radiology, Functional Brain Imaging Center, Temple University Hospital and School of Medicine, Philadelphia, PA 19140, United States*
[c] *Department of Psychology, Evolutionary Cognitive Neuroscience Laboratory, Drexel University, Philadelphia, PA 19104, United States*

Received 1 July 2005; received in revised form 16 August 2005; accepted 25 August 2005

Abstract

The Integrated Zone Comparison Technique (IZCT) was utilized with computerized polygraph instrumentation as part of a blind study in the detection of deception. Three scoring algorithms: ASIT Poly Suite (Academy for Scientific Investigative Training's Horizontal Scoring and Algorithm for Chart Interpretation), PolyScore® 5.5, and the Objective Scoring System (OSS) were assessed in the interpretation of the charts generated. Where "Inconclusives" were excluded, accuracy for the IZCT with all three algorithms was 100%. When "Inconclusives" were counted as errors, overall accuracy for the IZCT with ASIT Poly Suite was 90% and accuracy with PolyScore and the Objective Scoring System was 72%.

Keywords: Integrated zone comparison technique; Polygraph; Scoring algorithms; PolyScore; Objective scoring system; ASIT Poly Suite; Horizontal scoring system; fMRI

1. Method

A blind study to assess the accuracy of the fMRI and polygraph in the detection of deception was set up using procedures developed in the polygraph profession. Twelve volunteer medical students were solicited to participate in the study. Using random selection half of the group was told that someone had fired a gun in the hospital and a video in the area showed someone resembling them may have done it, the other half was instructed to actually fire a gun.

The mock crime scenario of shooting a gun was chosen to maximize the severity of the mock crime and maximize the sensory experience of the "shooters," who would experience holding the gun, firing it, the sound of the gun and the smell of the burnt gun powder. It was felt that this sensory intensity would create a more "real life" experience than traditional analog studies.

Both groups were given an initial monetary payment for participating in the study, and informed that they would receive an additional monetary bonus if they were determined to be truthful when denying that they fired the gun. Thus, both the truthful and deceptive suspects were motivated toward a truthful outcome, which is consistent with real life examinations; both truthful and deceptive suspects desire to come out truthful.

Half of the suspects were first tested utilizing a Lafayette LX-4000 computerized instrument, which monitored thoracic and abdominal breathing, electro-dermal activity, and cardiovascular activity using a standard blood pressure cuff, and then were tested using the fMRI. The other half were first tested in the fMRI, and then tested with the polygraph.

Overall results of the fMRI indicated unique areas of the brain are involved in truth telling and deception [1]. In summary, during deception there were fourteen regions that were significantly active, and during truth telling there were seven areas of significant activation. Further analysis is being conducted to determine whether these group differences can be turned into differences in individual assessments which will allow for accurate determinations of truth or deception.

All suspects were first interviewed using the Forensic Assessment Interview Technique (FAINT) [2]. This is one of

* Corresponding author. Tel.: +1 215 7632 3349.
E-mail address: truthdoctor@polygraph-training.com (N.J. Gordon).

0031-9384/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.physbeh.2005.08.046

many variations actually used in the field for pre-test interviews of suspects about to undertake the polygraph examination.

The polygraph technique used was the Integrated Zone Comparison Technique (IZCT) [3].

The question format was:

1.	Irrelevant	Is your first name _____?
2.	Outside Issue	Do you understand that I will only ask the questions I reviewed?
3.	Weak relevant	Do you intend to deliberately lie to any test question?
4.	Irrelevant	Were you born in ___?
5.	Comparison	During the first (−2 years from current age), did you ever lie to make yourself look better?
6.	Relevant	Today, did you fire that gun?
7.	Irrelevant	Is your last name ___?
8.	Comparison	In your entire life, did you ever cheat?
9.	Relevant	Regarding that gun, did you fire it today?
10.	Irrelevant	Is today Sunday?
11.	Comparison	During the (−2 years from current age), did you ever take credit for something you did not do?
12.	Relevant	Did you lie about whether you fired that gun today?
13.	Countermeasure	Did you deliberately do anything to try and beat this test?

The IZCT uses a zone format first introduced by Cleve Backster in the 1960's [4]. This format has been shown in studies to be the most accurate specific issue testing technique [5]. The actual examinations conducted were single-issue examinations. All relevant questions dealt with the single act of shooting the gun, as compared to a multi-issue format which would deal with secondary issues, such as knowledge or presence of the issue under inquiry.

All suspects were first given a "known number" demonstration or acquaintance test. Each examinee was asked to pick a number between 2 and 5, and reveal their choice. The examiner then administered a single chart instructing the examinee to answer "no" to every question from 1 to 6, including the number they actually selected. The examiner then explained and showed the examinee that the "lie" to the number they actually picked was identifiable and the polygraph procedure was able to detect it.

This was followed by three charts of the IZCT. The first chart was administered as a Silent Answer Test with the comparison–relevant question order as: C5–R6–C8–R9 –C11–R12, the second chart the examinee answered each question out loud in a mixed sequence by rotation of relevant questions (C5–R12–C8–R6–C11–R9), and in the third chart administered the examinee answered out loud with another rotation of relevant questions where the relevant questions now preceded the comparison questions (R9–C5–R12–C8–R6–C11).

The Silent Answer Test (SAT), utilized in the IZCT first chart was originated by Horvath and Reid [6]. They used the SAT only in the latter part of their testing process if the examiner was having problems making a decision. Horvath and Reid reported among other advantages that the SAT produced enhanced electro-dermal reactions, and, "even if the subject failed to react significantly on the SAT, it tends to induce greater responses on the later tests." The Utah technique used by Raskin et al. also utilized the SAT on the fourth crime chart if the first three charts were not conclusive.

The SAT was introduced to the examinee in the following manner: "In this first test I am going to ask you the questions I just reviewed with you. During the test I don't want you to answer out loud. I just want you to listen to the questions one more time, get used to being attached to the instrument, and having me ask you questions. I want you to make sure you have understood all of the questions, feel comfortable with them, and most importantly, that you have answered every question truthfully. If you remember anything you haven't told me about, you can tell me as soon as the test is over, but don't say anything out loud during this first test: just listen."

The purpose of using the SAT in the first chart is that most examinees do not consciously perceive the chart as a threat, since they are not answering out loud, and lying. Thus, they rarely attempt any type of mental or physical countermeasures. This not only results in an excellent chart of physiological tracings, but also excellent reactions to the appropriate zone of questions where deception will be attempted. It is not the utterance of "yes" or "no" that creates sympathetic nervous system arousal. It is the examinee's cognitive recognition of the threat the questions pose to his or her well being in a format to which they will attempt deception.

Golden, in the Listen-Answer Technique, hypothesized that maximum psychophysiological stress would be generated during the presentation of an incriminating question, when the person was instructed just to listen, and that vocalization to the same question would actually generate psychophysiological relief. He made the analogy that for the deceptive person, not to be able to utter their lie was like a person stubbing their toe and not being able to yell out in pain, thus resulting in greater psychophysiological reactions to occur.

In Chart 2, the examinees were instructed to answer each question truthfully out loud. The examiner further instructed that lying to any question, regardless of which question it was, could result in the examinee failing the test. This verbal stimulation further helps self-set examinees to the questions that pose the greatest threat to them.

As previously stated, during this chart the relevant question positions are rotated to allow each relevant question to be next to a different comparison question. This will ultimately pair each of the relevant questions with each of the comparison questions once during the three-chart examination. Mixing of the question order is also done as a safeguard against habituation and anticipation.

In Chart 3, the sequence of the chart is reversed from a green–red format (comparison–relevant), to a red–green format (relevant–comparison). The author believes that a green–red (comparison followed by relevant question sequencing) testing format biases the test toward a truthful outcome, and that a red–green testing format (relevant followed by comparison question sequencing) biases a test toward a deceptive outcome. In the IZCT, the first two charts are biased

toward a truthful outcome, and the third chart is biased toward a deceptive outcome. In the IZCT this reversal of the question order in chart 3 safeguards against both false positives and false negatives, giving the overall process a more accurate and balanced conclusion.

Prior to conducting the polygraph examinations there was a concern about the possibility of an increase in false-negative outcomes (deceptive suspects appearing truthful) since the threat of being detected of shooting the gun was minimal, resulting in the theoretical loss of a small monetary gain, however, reactions to the Comparison Questions would reveal actual indiscretions in the suspect's past now being revealed to their professors.

2. Results

One of the deceptive suspects confessed during the FAINT interview and was excluded, leaving five (5) deceptive and six (6) truthful suspects. All charts were interpreted using three different systems: ASIT Poly Suite (Academy for Scientific Investigative Training's Horizontal Scoring System and Algorithm for Chart Interpretation) [7], PolyScore 5.5® [8], and the Objective Scoring System (OSS) [9].

Utilizing the "Horizontal Scoring Technique and Academy's Algorithm for Chart Interpretation" each physiological parameter monitored by the examiner was placed in a rank order hierarchy from greatest reaction receiving a "6" to the least reaction receiving a "1." Criteria for determining the greatest reaction in the pneumo was based on suppression, apnea and duration of reaction (lack of air). The electro-dermal reactions were ranked based on the height of the reaction squared multiplied by the duration of the reaction. The greatest number received the highest rank. The cardio reactions were ranked based on diastolic blood volume increases.

Since both thoracic and abdominal breathing were ranked, these two parameters scores for each question were then averaged, to maintain a final decision which would be based evenly (1/3) for each of the three parameters monitored. This component input differs from that of the other two algorithms which make decisions based on a weighted system heavily favoring electro-dermal activity.

Once each question's parameters were ranked, and the pneumo reactions were averaged, the sum of the three scores (average of the pneumos, plus the electro-dermal and cardio) represented the question score. Comparison question scores were given positive numbers and relevant question scores received negative numbers. The sum of all of the comparison and relevant scores represented the examination score. If the examination score was a +13 or higher the examinee was determined to be truthful when they denied shooting the gun. If the score was a −13 or lower, they were determined to be deceptive. Any score between the ±13 was determined to be inconclusive.

PolyScore® was introduced by Dale Olsen and John Harris of Johns Hopkins University's Applied Physics Laboratory in 1989 through a series of contracts with Department of Defense agencies. The features used in the algorithm were evaluated by testing them in a logistic regression-produced decision rule which produces a score from various linear weight combinations of the features by "logit" conversion, which calculates a probability of deception. If a score is 0.95, the correct interpretation based on the data in the PolyScore® data base is that 95% of the time when similar features are present, deception is indicated. Scores that are 95% or above are therefore interpreted as deceptive and scores that are .05 or less are interpreted as truthful. Scores in between those criteria are interpreted as inconclusive. The algorithm devotes 54% of its average decisions based upon electro-dermal activity, 24% to blood volume, 8% to pulse rate and 14% to breathing.

The Objective Scoring System (OSS) introduced by Donald Krapohl and Barry McManus in 1999, utilizes measurements of criteria established by Raskin et al. in 1988: Timm line length of the pneumo (lack of air equals reaction) which is measured for 10 s after question onset, electro-dermal amplitude and blood volume increases. The physiological parameters are weighted similar to PolyScore®, with 50% of the decision generated from electro-dermal activity, 25% from respiration and 25% from blood volume. This system uses a traditional 7 point scale, where scores of a +6 or higher are interpreted as truthful, −6 and lower are deceptive, and scores in between are deemed inconclusive.

The accuracy of polygraph results are looked at in two ways: inconclusive results considered as errors, and inconclusive results just being excluded. Obviously, opponents to the polygraph procedure prefer the former view, and proponents the latter.

If "Inconclusives" are not considered errors, then all three systems had 100% accuracy using the IZCT polygraph technique. There were no false-positives or false-negatives.

Of the six (6) deceptive suspects PolyScore and OSS had one (1) Inconclusive each, and there were no "Inconclusives" by ASIT Poly Suite. Accuracy for ASIT Poly Suite was 100%, and accuracy for PolyScore and OSS was 83%, when "Inconclusives" are considered errors.

Of the five (5) truthful suspects PolyScore and OSS had two "Inconclusive" findings and ASIT Poly Suite had one. Accuracy for PolyScore and OSS was 60% and ASIT Poly Suite was 80% when "Inconclusives" are considered errors.

When "Inconclusives" are considered errors, overall accuracy for PolyScore and OSS was 72% (83% for DI and 60% for NDI), and overall accuracy for ASIT Poly Suite was 90% (100% for DI and 80% for NDI).

Comparison of the three algorithms

Name	Ground zero	PolyScore 5.5	OSS	ASIT Poly Suite
P D	DI	DI probability >.99	DI probability >.99	DI −39
K S	DI	DI probability >.99	DI probability >.99	DI −40
J R	NDI	NDI probability >.99	NDI probability >.99	NDI +26
J B	NDI	INC probability DI .78	INC +8	INC +5
L H	NDI	NDI probability >.99	NDI probability >.99	NDI +36
M C	DI	DI probability >.99	INC 0	DI −24
N M	DI	DI probability >.99	DI probability >.99	DI −58
S M	DI	INC probability >.64	DI probability >.98	DI −23
B B	DI	DI probability >.99	DI probability >.99	DI −32
S P	NDI	NDI probability >.99	NDI probability >.99	NDI +54
S K	NDI	INC probability DI .94	INC +3	NDI +16

Code:

DI: deception indicated; NDI: no deception indicated; INC: inconclusive.

3. Discussion

When "Inconclusives" were not viewed as errors all three algorithms had 100% accuracy. There were no false-positives or false-negatives. The first question, then becomes, should "Inconclusives" be considered errors?

If a patient seeks the assistance of a physician for a medical problem, and after examining the patient the doctor requests the patient to go elsewhere for further medical tests prior to making a diagnosis, has that physician erred? We do not believe this would be viewed as an error, nor do we believe this physician could be sued for malpractice.

Likewise, if a polygraphist concludes an examinee's data is insufficient to make an accurate determination of truth or deception, and therefore requests additional data be collected prior to making a decision, we do not see how this could be considered an error.

The more accurate query is whether the "Inconclusive" rate is so high that it, in itself, invalidates the use of the procedure. In this study the "Inconclusive" rate for the ASIT Poly Suite for DI suspects was 0% and 20% for NDI suspects. Overall there was a 9% "Inconclusive" rate. Both PolyScore and OSS had a 17% "Inconclusive" rate for DI suspects, and a 40% "Inconclusive" rate for NDI suspects. Overall these two systems had a 28% "Inconclusive" rate.

Inconclusive rates

	DI suspects	NDI suspects	Overall
ASIT Poly Suite	0%	20%	10%
PolyScore 5.5	17%	40%	28%
OSS	17%	40%	28%

In this initial study the number of subjects was small (6 DI and 5 NDI), and it will be interesting in future experiments to increase the suspect size and observe its affect, if any, on the various algorithm results.

References

[1] Mohamed FB, Faro SH, Gordon NJ, Platek SM, Ahmad H, Williams JM. fMRI study of truth and deception using classical question technique. 11th Scientific Meeting of the International Society of Magnetic Resonance in Medicine, p. 1920.

[2] Gordon NJ, Fleisher WL. Effective interviewing and interrogation techniques. New York: Academic Press; 2002.

[3] Gordon NJ, Fleisher WL, Morsie H, Habib W, Salah K. A field validity study of the integrated zone comparison technique. Polygraph 2000;29(3):220–5.

[4] Backster C. Technique fundamentals of the tri zone polygraph test. New York: Backster Research Foundation; 1969.

[5] Raskin DC, Kirscher JC, Honts CR, Horowitz SW. Validity and reliability of detection of deception. Washington, D.C.: National Institute of Law Enforcement and Criminal Justice; 1978.

[6] Horvath H, Reid JE. The polygraph silent answer test. J Crim Law Criminol Police Sci 1972;63(2).

[7] Gordon NJ. The Academy for Scientific Investigative Training's Horizontal Scoring System and Examiner Algorithm for Chart Interpretation. Polygraph 1999;28(1):56–64.

[8] Mattè JA. Forensic psychophysiology; using the polygraph. Williamsville, NY: J.A.M. Publications; 1996.

[9] Krapohl D, Mc Manus B. An objective method for manually scoring polygraph data. Polygraph 1999;28(1):209–22.