EUROPEAN
POLYGRAPH

Volume 4 • 2010 • Number 2 (12)

Marek Leśniak*
Department of Criminalistics
Faculty of Law and Administration
University of Silesia in Katowice
POLAND
Magdalena Zubańska
Higher Police School in Szczytno
POLAND

A Comparison of Polygraph Examination Accuracy Rates Obtained Using the Seven-position Numerical Analysis Scale and the Objective Scoring System (A Study on the Polish Population)

Background

Different sorts of polygraph examinations have been conducted in Poland over the last sixty years (Krzyścin, 226, 227). It is curious that there are few empirical studies which concern the accuracy of such examinations in relation to the population of Poland. Hundreds of well-documented scientific studies on the accuracy of polygraph examination have been published all over the world. However, most concentrate on the American or Israeli population, so the problem of national differences should be taken into consideration, too. The nature of reactions to polygraph test questions usually justifies the transfer of

* marek.lesniak@us.edu.pl

the results of foreign studies onto Polish people. However, not having significant differences in this area should be confirmed by some empirical research. The current paradigm of empirical science demands replications of studies concerning the same variables – the most certain tool of scientific cognition is meta-analysis (Shaughnessy, Zechmeister, Zechmeister 2002, 271–273).

The objective of the present study is a comparison of psychophysiological detection of deception accuracy rates obtained by using the seven-position numerical analysis scale and the Objective Scoring System (this algorithm is sold together with Lafayette's polygraph software).

Both scoring systems rely on comparing the responses to the relevant question to responses to the appropriate comparison question (by individual physiological recorded parameters). The OSS was introduced by Donald Krapohl and Barry McManus in 1999 (Gordon, Mohamed, Faro, Platek, Ahmad, Williams, 253). Fifty per cent of decisions are generated from electro-dermal activity tracing, 25% from respiratory tracing and 25% from cardiovascular tracing. In contrast with a global evaluation (based on a form of a general impression) the numeral manual score allows adequately trained interpreters to reach extremely high reliability (Raskin, Honts, 18–19). Similarly to OSS features, during manually numerical scoring, changes in electro-dermal activity are more important than changes in respiration and cardiovascular activities. The reason for this is that the electro-dermal physiological parameter is only present where a ratio method is used solely in scoring the responses (Swinford, 18). Any expert's diagnosis based on numerical scoring may be checked easily by another expert. The polygraph technique used was the Utah Directed-Lie Test. This choice was driven both by the fact that this technique is practised in Poland and to avoid ethical (related to the characteristics of examinees) and technical difficulties in establishing effective comparison questions (Raskin, Honts, 24).

Method

Ultimately the research includes 43 participants (six polygraph outcomes were rejected on account of their very poor quality) – 23 men and 20 women. Individuals were recruited from students of the University of Silesia in Katowice. Some of them (called the "guilty subjects" group) had taken a 10-zloty note out of the cabinet in room number 3.53. Others (called the "innocent subjects" group) did not open the cabinet and did not see the note. Members of the

"guilty" group had been instructed to keep the note in their pocket until the time of the polygraph examination and during this examination too. The polygraph examinations were carried out by eight experts; they were certified as having completed skills training in this area. The experimenters assumed that the experts had adequate training requirements. The participants were tested with a Lafayette LX-4000 polygraph that monitored thoracic and abdominal breathing, electro-dermal activity and cardiovascular activity (only a standard blood pressure cuff). The experimenters did not conduct the polygraph examination; they measured the size of particular reactions on charts and interpreted the results. The interpreters got to know the actual role of each participant only when they had finished interpreting the polygraph outcomes. Polygraph examinations were conducted according to the rules that were described by David C. Raskin and Charles R. Honts (22–24). During the manual numerical evaluations, the interpreters followed Jimmie Swinford's instructions (1999) precisely. The OSS scoring system was utilized in version number 2.

The test questions for participants were as follows:

1. Are you aware of the fact that I will ask only the questions we have discussed?
2. Do you intend to answer untruthfully all of the questions about the taking of a ten-zloty note?
3. Is it Wednesday today?
4. Prior to 2009, did you ever take something that did not belong to you?
5. Did you take the 10-zloty note out of the cabinet from room number 3.53?
6. Did you see the 10-zloty note inside the cabinet in room number 3.53?
7. Prior to 2009, did you ever do anything that was dishonest or illegal?
8. Did you open the cabinet which the 10-zloty note was put in?
9. Are you holding the 10-zloty note in your pocket now?
10. Before the age of 18, did you ever lie to get out of trouble?
11. Are we in Katowice now?

Results

The results of the accuracy estimation are presented below in two ways: the first shows inconclusive results as errors, and according to the second inconclusive results are excluded (Gordon, Mohamed, Faro, Platek, Ahmad, Williams, 253).

Concerning the OSS interpretations within the "guilty subjects" group the percentage of participants correctly classified as "guilty" was 26% altogether (38% – did not include 32% IC diagnoses). Within the "innocent subjects" group the percentage of participants correctly classified as "innocent" was 83% altogether (95 – did not include 3% IC diagnoses).

Concerning the manually numerical score within the "guilty subjects" group the percentage of participants correctly classified as "guilty" was 58% altogether (100% – did not include 42% IC diagnoses). Within the "innocent subjects" group the percentage of participants correctly classified as "innocent" was 67% altogether (80% – did not include 25% IC diagnoses).

As far as all participants are concerned the percentage of participants correctly classified as "guilty" or "innocent" by means of computer scoring was 58% altogether (78% – did not include 21% IC conclusions).

As far as all participants are concerned the percentage of participants correctly classified as "guilty" or "innocent" by means of the manually numerical score was 61% altogether (93% – did not include 35% IC conclusions).
The scoresheets shown below reveal a range of divergence between computer scoring and manually scoring (the figures presented in the scoresheets reflect the number of conclusions).

Within the "guilty" subjects group:

	DI computer scoring diagnoses	NDI computer scoring diagnoses	IC computer scoring diagnoses
DI manually numerical scoring diagnoses	3	4	3
NDI manually numerical scoring diagnoses	0	0	0
IC manually numerical scoring diagnoses	2	4	3

Within the "innocent" subjects group:

	DI computer scoring diagnoses	NDI computer scoring diagnoses	IC computer scoring diagnoses
DI manually numerical scoring diagnoses	0	0	2
NDI manually numerical scoring diagnoses	0	16	0
IC manually numerical scoring diagnoses	1	4	1

Discussion

According to Jerzy Konieczny's recommendation (61, 62), a criterion for acceptability of a polygraph examination technique accuracy was taken from the American Society for Testing and Materials. Among other criteria, the method of forensic identification is accepted if its accuracy reaches 90 per cent (when its outcomes may be used as evidence in court) or 80 per cent (for investigation purposes). Simultaneously, inconclusive decisions may not exceed the 20-per cent limit. The results of the present research do not meet this criterion. Presumably, the primary cause for this is the low level of participant motivation to deceive the polygrapher. As a source of reaction the experimenters could depend only on the subject's feeling of discomfort (when lying is found as something against his or her internalized moral standards) or a participant's joy in misinforming an expert. The research thus faced a typical laboratory study problem.

It is necessary to mention that the quality of recording physiological changes did not belong to the highest. Six subject's charts were rejected on the account of their poor quality. Amongst the accepted charts, improper polygrapher activity sometimes disrupted the course of recording. During a manual evaluation such spots received the number "0". One of the many merits of the numerical scoring systems is the requirement for experts to base on the highest quality record.

An explanation for the divergences received between results of manual numerical evaluations and computer algorithm diagnosis can be found in the characteristics of applied estimation systems. Both put special emphasis on electro-dermal changes. But the diagnostic importance of other recorded changes (concerning the respiration and cardiovascular activity) is more sig-

nificant in the manual numerical evaluation. The electro-dermal activity component is the most responsive of all the parameters recorded by the polygraph (Swinford, 17, 18). At the same time it is more susceptible to coincidental disturbances. It is significant that the mentioned divergences are serious in case of relevant participants. In this instance, some subjects' motivation and the related reactivity to relevant questions were not strong enough.

The accuracy rates received were lower than the results of published reports about laboratory studies into Directed-Lie Polygraph Tests (Raskin, Honts, 25; Kircher, Packardt, Bell, Bernhardt, 35). The explanation can be found in the above-mentioned circumstances: problems with participants' motivation and a lack of skills on the part of the expert.

The results of the present study confirm the statement that a computer diagnosis should not be a single basis for a polygrapher's interpretation decision, but may play a major part in this kind of decision-making (Konieczny, 178–181).

References

Gordon N. J., Mohammed F. B., Faro S. H., Platek S. M., Ahmad H., Williams J. M. (2006), *Integrated Zone Comparison Polygraph Technique Accuracy with Scoring Algorithms*, Physiology and Behaviour, 87 (2).

Kircher J. C., Packard T., Bell B. G., Bernhardt P.C. (2010), *Effects of Prior Demonstrations of Polygraph Accuracy on Outcomes of Probably-Lie and Directed-Lie Polygraph Tests*, Polygraph, 39 (1).

Konieczny J. (2009), *Badania poligraficzne. Podręcznik dla zawodowców*, Wydawnictwa Akademickie i Profesjonalne, Warszawa.

Krzyścin A. (2000), *The Debate over Polygraph in Poland*, Polygraph, 29 (3).

Raskin D. C., Honts Ch. R. (2002), *The Comparison Question test*, [in:] *Handbook of Polygraph Testing*, ed. by M. Kleiner, Academic Press, San Diego.

Shaghnessy J. J., Zechmeister E. B., Zechmeister J. S. (2002), *Metody badawcze w psychologii*, Gdańskie Wydawnictwo Pscyhologiczne, Gdańsk.

Swinford J. (1999), *Manually Scoring Polygraph Charts Utilizing the Seven-position Numerical Analysis Scale at the Department of Defense Polygraph Institute*, Polygraph, 28 (1).