

EUROPEAN  
**POLYGRAPH**

PUBLISHED SEMI-ANNUALLY  
2022 VOLUME 16 NUMBER 1 (55)

DOI: 10.2478/EP-2022-0002

# Should Blind Evaluation of Polygraph Charts Be a Mandatory Procedure in Evidentiary Examinations?

Marcin Gołaszewski, PhD

Polskie Towarzystwo Badań Poligraficznych  
biuro@ptbp.pl

## Abstract

The article considers the advantages of blind interpretation of polygraph charts in the context of subjectivity in polygraph examinations. The purpose of this article is to provoke a discussion on the inclusion of blind scoring in evidentiary examinations as a standard procedure. Resorting to such a method should curb the impact of cognitive bias on interpreting test data as it has been proved empirically that information on the case facts and the examinee, provided to the examiner before the examination, may influence the subsequent interpretation of the charts.

**Key words:** polygraph, charts, subjectivity, blind evaluation, evidentiary examinations

There are three situations in which polygraph experts analyse polygraph charts when it comes to familiarity with the subject and object of the examinations. The first and basic occurs when the expert runs the examination in person and analyses the data obtained later. Such an expert knows the details of the case both from the information shared by the party commissioning the examination and materials delivered, and from the interaction with the examinee (subject of the examination). The other two cases involve commissioning examination of charts to additional experts, in which case they are either given access to the same information as the examining expert had, including information on the course of the examination itself, or else the evaluating experts are absolutely independent in doing their task and have no access to any data other than the charts being the record of physiological reactions. This last case is known as “blind interpretation”. Such a procedure eliminates all the subjective factors connected to the circumstances of the case that the expert opinion to be delivered concerns. It limits subjectivism solely to the polygraph examiner’s personal preferences concerning interpretation of data wherever a certain leeway is present.

Following what M. Orne noticed, the result of a polygraph examination may be influenced in actual conditions by other evidence gathered for the case and also by the examiner’s conviction about the guilt or innocence of the subject. For a method to be fully accurate from the scientific point of view, it is necessary to separate test results from other aspects of a given case, however this cannot be achieved in practice (Orne, 1973: 177). Orne quoted data from P. Berch’s experiment comparing diagnoses of polygraph examiners with the opinions from a panel of three legal experts working on the body of evidence presented to the court and also on additional information that could not have been transformed into formally permissible evidence in the cases. Wherever the three experts were unanimous, the assessments made by the polygraphers were aligned with them in 92.4% of cases. When one of the lawyers expressed a dissenting opinion, the alignment between the assessments of the polygraphers and the panel dropped down to 74.6% (Berch 1969: 399–403). Orne presented two hypotheses that could explain the shift. The first claims that these were still the results of polygraph examinations that were more accurate than the views of the lawyers disputing the ambiguous body of evidence, while the other presented the option that, with the full unanimity in the panel of experts, the information that the lawyers used could equally well make a significant influence on the behaviour of the polygraph examiner towards the subject, the way the subject reacted, and the examiner’s final judgement (Orne 1973: 178–179).

Experts examine the charts for the diagnostic features. Their theoretical knowledge and practical experience lets them discriminate whether a given feature is highly likely, and what its identification significance is. They also describe the degree of accuracy of the diagnosis they offer, whether in descriptive or statistical terms. The space for subjectivism in polygraph examinations exists since the changes of physiological reactions of the subject must not only be measured but also subsequently interpreted according to a set of criteria approved for the given analytical method. Depending on the method, the expert has a greater or smaller leeway in the interpretation whenever they are not bound with objective and accurate criteria.

There are factors that influence the cognitive processes of every person issuing opinions. These include the emotional condition, preconceptions concerning the subject of the analysis, and the environment in which the expert functions, especially pressure on behalf of the party commissioning the opinion. The plethora of information to be processed makes the examiner's mind apply simplified rules of cognition that may be helpful for making the decisions but loading the process with the risk of cognitive errors.

It has been proved empirically that the information on the person subjected to polygraph examinations that an expert learns before analysing the data recorded by the polygraph influences the subsequent interpretation of the charts (Elaad, Ginton, Ben-Shakhar 1994: 280-281; Elaad, Ginton, Ben-Shakhar, 1998; Shurany, Matte, Stein, 2009: 133-139; Krapohl, Dutton, 2018: 99). Independent of the professional experience they have, experts are biased by the information they received earlier that create the preconception of a guilty or innocent individual. This has impact on the numerical assessments of the records of changes in physiological reactions after relevant questions but, fortunately, only in the cases when the differences in the reactions compared are far from obvious, call for complicated analyses, and the final result of the test oscillates around the decision-making threshold. Such cases are not dominant in real conditions. However, unequivocal data from the examinations leaves no room for free interpretation to the expert polygrapher, and therefore not do they offer an option to confirm the hypothesis resulting from the earlier expectations. It must also be emphasised that the phenomenon of becoming influenced by the information obtained before making the expertise described here does not manifest itself particularly among polygraphers. It is also present in other forensic examinations, notably the ones that are generally believed to be more valid, to mention fingerprinting (Dror, Charlton, Pèron, 2006: 74-78).

A. Ginton's conclusions demonstrate that the consequences of previously made expectations in actual cases are lower than in the claims made by opponents of polygraph examinations (Ginton 2019). He further recognised that for practical purposes it is most important to know how many of such "inconclusive" tests may incorrectly turn into conclusive under the influence of the examiner's prior expectations. For that he assumed the level of inconclusive results at 20%, a level that is correct, as for tests recommended to various types of examinations it is the maximum permissible level set in the current validation requirements, while the average share of such results was estimated from a meta-analysis at 12.7% (Gougler, Nelson, Handler, Krapohl, Shaw, Bierman 2011). Ginton estimated that the error of confirmation realistically concerns around 3% of all the practically conducted examinations, and moreover not all of those result in inaccurate identification, as some of the prior expectations are aligned with the status quo, and they do not follow just one direction, whether inculpatory or exculpatory the subject from suspicions.

In an experiment conducted by Holmes in the 1950s, making polygraph examiners familiar with the cases from which the charts they assessed came resulted in a small increase of their diagnosing accuracy: by 8% (Holmes 1957: 67–70). The correlation was also experimentally tested by D. Wicklander and F. Hunter, who had six mutually independent polygraph experts analysing 20 sets of polygraph charts (Reid technique) (Reid, Inbau, 1977) at least two months apart. In the first round, they only shortly presented the issue that a given test concerned (e.g. stealing money from an office safe). Yet, they expanded the scope of information shared before the second interpretation, adding a short historical context of the event, basic background data on the subject together with their verbal and non-verbal behaviours, and the list of relevant test questions. While the average rate of correct diagnoses in the first their first analysis of data amounted to 88.33% [0.7–0.95], sharing additional information with the polygraphers made their opinions slightly more accurate, as the rate of correct verdicts reached 92.5% [0.8–1.0]. Four in six examiners improved the results, one remained at the previous level, and one returned poorer results then making a blind interpretation. Moreover, the number of inconclusive results dropped by half (Wicklander, Hunter, 1975).

Various studies on the accuracy of comparison questions tests in actual cases (that is outside an experimental laboratory) suggest that experts who conduct the examinations in person may obtain better results than the ones who only perform blind interpretation of polygraph charts. As much as both the groups identified deceptive individuals with nearly identical accuracy rates, they were more accurate in identifying truthful people (see Tab. 1 and Tab. 2).

Table 1. Results of selected practical studies concerning accuracy of CQT tests: tests assessed by independent examiners

Study	Deceptive (%)				Truthful (%)			
	n	Correct	Inaccurate	Inconclusive	n	Correct	Inaccurate	Inconclusive
Honts (1996: 309–324)	7	100	0	0	6	83	0	17
Honts & Raskin (1988: 56-61)	12	92	0	8	13	62	15	23
Patrick & Iacono (1991: 229–238)	52	92	2	6	37	30	24	46
Raskin <i>et al.</i> (1988)	37	73	0	27	26	61	8	31
<i>Mean</i>	27	89	1	10	82	59	12	29
<i>Average percentage</i>		98	2	-		83	17	-

Table 2. Results of selected practical studies concerning accuracy of CQT tests: tests assessed by the polygraphers who conducted the original examinations

Study	Deceptive (percentage correct)	Truthful (percentage correct)
Horvath (1977: 127–136)	100%	100%
Honts & Raskin (1988: 56-61)	92%	100%
Raskin <i>et al.</i> (1988)	95%	96%
Patrick & Iacono (1991: 229–238)	100%	90%
Honts (1996: 309–324)	94%	100%
<i>Average</i>	98%	97%

In the early 1970s F. Horvath and J. Reid selected 40 sets of polygraph charts recorded in Reid technique (20 verified as coming from guilty and the remaining 20 from innocent subjects) for an experiment and delivered them for evaluation to polygraphers with various levels of professional experience. Performing the selection, they rejected those charts that were so evident that even a layman would notice differences in reactions to critical and control questions, and also ones impossible to interpret even by a trained polygrapher. The evaluators were only given general information about the subjects of the individual tests and only granted a single working day to perform their task. Out of the seven polygraphers who had had at least a year's experience in practical diagnosing, accuracy was at 91.4%, and the range of correct assessments at 12.5% [0.85-97.5]. The remaining three polygraphers who were only taking the first steps on their career paths offered accuracy level of 79.1% [0.70–0.90]. The total rate of correct diagnoses was estimated at

87.5%, and the divergence between polygraphers' accurate diagnoses – at 27.5% (Horvath, Reid 1971: 276–281). More experienced individuals returned better results. Moreover, identification of innocent individuals was more accurate at 9.5% of false positive results, as compared to 15% of false negative results. What Horvath and Reid primarily proved was that experts in polygraph examinations are capable of attaining accurate and reliable results only using the recorded polygraph charts without knowledge of the details of the case and without personal interaction with the subject. However, they also believed that the experts who conduct the testing in person and are fully aware of the case are in a better position. They recognised additional behavioural hints as something that favours more accurate diagnoses when presented in combination with the recorded physiological data. It must be noted that they formed their view quite arbitrarily. Sometimes the behaviour of the subject helps in accurate assessment of veracity of their response, yet at times it may also be a hindrance (see: Othello's error).

It would also be impossible not to note that majority of studies examining accuracy of tests and consistency of the assessments made by polygraph experts conducted until the 1970s concerned Reid's control questions, peak of tension (POT), and relevant and irrelevant (R/I) techniques. Moreover, they were evaluated according to the qualitative method. In turn, the numerical method only entered the early tests phase, and would only accompany Baxter's technique (on a 7-point scale). The few who tried to apply a partially objectivised numerical method at the time included J. Kubis (Kubis 1962; quoted after: Matte 1996: 45–46) and G. Barland with D. Raskin (Barland, Raskin 1971: 275). More contemporary scientific studies concerning the consistency of polygrapher assessments were conducted to validate various techniques and, unlike the early attempts, they in fact only included numerical methods. Apart from the Empirical Assessment System, those were no studies focused on the manners of interpreting test data but standardised tests. Yet, due to the existence of multiple scientifically approved systems for evaluating polygraph charts, various studies used various systems. That is why the available data is connected to specific types of tests, with the tests of fundamental significance here being those of the ZCT type: diagnostic, of single-issue or multi-facet nature, and therefore most frequently used as evidence in criminal cases and other official procedures.

Can one make a claim that assessments of polygraph charts performed by the polygraphers who conducted the examination are always, or at least usually, more accurate than blind interpretation? There are counterarguments of theoretical na-

ture and ones resulting from other empirical studies than the ones quoted above. Falling back on the theory, it seems that the lower the number of factors making subjective impact on the polygraphers, the less distorted their assessment should be, in this way offering higher accuracy of decision, and the degree of consistency between various experts. If the reality were different, the subjective circumstances connected primarily to the direct interaction with the subject would, as a rule and beside the data recorded by the polygraph, have positive influence assessment accuracy.

In 2014–18, a research project on Criminal, Ethical, and Legal Problems in Instrumental and Non-Instrumental Methods of Detecting Deception (Polish title: *Instrumentalne i nieinstrumentalne metody detekcji nieszczerości – problemy kryminalistyczne, etyczne i prawne*) was conducted in Poland with participation of the author.<sup>7</sup> One of the subjects it tackled was subjectivism in polygraph examinations. The project invited 15 professional polygraphers to conduct experimental testing. Of that number, three examined the subjects in person, and the others were given the task to perform blind assessments of the polygraph curves with various methods. The subjects were 39 volunteers selected from among the students of the AFM Kraków University: 13 men and 26 women aged 20–43. The event staged for the experiment consisted of firing three shots at a silhouette of a young woman on a colourful poster at the university's shooting range. It was performed by 15 subjects, that is the "guilty" group, who were later given the task not to admit to perpetrating the act while being examined. To be further motivated they were given a banknote they could retain if they were identified negative (innocent) by the examiner. In turn, the "innocents" (24 people) never visited the shooting range, had no knowledge of what transpired there, and were to provide truthful answers during the examination. They also received pecuniary gratification but were supposed to return it in the case of a false positive identification. (The idea was to make sure that they follow their role in the experiment. Moreover, in real conditions, the suspected innocents also bear the consequences of a possible incorrect expert diagnosis.) (Widacki (ed.) 2018: 65).

Apart from the demonstration test (peak of tension test), the examiners used the same standardised format diagnostic test with comparison questions (the Utah Zone Comparison Test). They interpreted the data that the subjects returned during the test themselves. In turn, the 12 remaining polygraphers, none of whom was

---

\* Decision of the National Center of Sciences No. DEC-2013/11/B/H55/03856

familiar with the case, only performed a blind assessment of the curves received. To evaluate the polygrams (polygraph curves), they were divided into three subgroups, each one applying a different method: the ESS system, Utah, and the global method. Apart from the human assessments by expert polygraphers, the experiment also involved obtaining results from OSS-3 analytical software (algorithms based on Senter's rules and Raskin probability analysis).

The results obtained in the experiment demonstrate that, as far as general accuracy is concerned, the best results were returned by the blind interpretation of the test data performed according to the ESS system (0.85). On the other hand, when it comes to the highest number of correctly identified cases, and not just bare percentages, that method did not excel but was downright inferior to the other ones, as it returned a relatively highest number of inconclusive results (See: Tab. 4).

Table 3. Data concerning the accuracy of test results assessment according to different methods in the Kraków experiment (excluding inconclusive results)

Test data analysis method	Accuracy (percentage, n=39 exams)
ESS – blind scoring	85%
Utah – blind scoring	82%
selected computerised algorithms (OSS-3)	77%
global analysis – blind scoring	74%
ESS – original examiners	74%

Table 4. Share of inconclusive results in the Kraków experiment

Test data analysis method	Inconclusives (each of n=39)	Percentage
ESS – original examiners	5	13%
global analysis – blind scoring	5	13%
ESS – blind scoring	12	31%
Utah – blind scoring	6	15%
selected computerised algorithms (OSS-3)	4	10%

Critics of presentation of the data on the specificity and sensitivity of testing methods with exclusion of inconclusive results argue that such a practice may artificially inflate these parameters. It is, nonetheless, highly justified for practical purposes, as inconclusive results contribute nothing to the case, as they trigger no decisions, as they are not interpreted in favour or against the subject. This



group of results affects the usefulness of the method and not its accuracy. That is why they are as a rule reported separately in the case of polygraph examinations. While investigating the results above, attention is drawn by the approximately three times higher percentage of inconclusive results in the case of experts using the ESS system for blind interpretation of the curves (31%) as compared to the polygraphers who conducted the examinations themselves using the same system (13%), and also to the average number of such results while conducting ZCT tests (9.8%) determined through meta-analysis (Widacki (ed.) 2018: 65). This proves that the “blind scorers” were either overly cautious in their evaluations or they were the ones who analysed the curves obtained from the experiment more correctly, while those who performed the examinations were more motivated to take decisive decisions, in some cases, consciously or not, stretching the results. The calculations presented in Table 3 suggest that the latter hypothesis is more likely, as the experts performing blind interpretation returned a higher percentage of accurate opinions. Their only task was to analyse the curves, and they were not familiar with the context of the test. In turn, the examiners tried to assign specific individuals to the role (guilty or innocent) they played in the experiment.

What seems most important in the case of polygraph examinations is the reduction of incorrect identifications to the minimum, even if they were to mean a slightly higher share of results considered useless due to the lack of conclusive indications. The tests interpreted blindly according to the ESS system also had the highest sensitivity (0.78) and also negative predictive value (that is probability that the subject is truly not the perpetrator of the deed in question as the test result demonstrates; 0.89). In turn, the highest specificity was achieved in the case of computer algorithms (0.91), and the positive predictive value (the probability that the subject is guilty when the results of the test show so) was the highest (0.82) in blind interpretation performed according to the Utah system (see: Tab. 5).

Table 5. Classification of the best methods for analysing test data in Kraków experiment for test sensitivity, specificity, NPV and PPV.

Indicator	Test data analysis method	Value
highest sensitivity	ESS – blind scoring	0.78
highest specificity	selected computerised algorithms (OSS-3)	0.91
highest NPV	ESS – blind scoring	0.89
highest PPV	Utah – blind scoring	0.82

The experts performing blind interpretation according to numerical methods (ESS and Utah) were more accurate than the ones who conducted the examinations themselves (0.85 and 0.82 respectively, compared to 0.74). In each case the results remained slightly below the average accuracy of Utah test as reported by the APA from the meta-analysis of results of various published studies (92.1%–93%) (Gougler *et al.* 2011), a result that should rather be linked to the specific circumstances of this particular experimental setup, as the skills and qualifications of the polygraphers involved did not diverge from the global standards.

The reasons why, unlike in the experiments described earlier, the polygraphers performing the examinations in person were less efficient in delivering accurate diagnosis than those assessing the charts blindly can be various, and start with the relatively small sample, as involving a larger group of polygraphers was unrealistic in Polish conditions. Moreover, the first group felt the pressure of time and expectations to provide categorical decisions. It cannot be ruled out that they followed subjective factors resulting from the direct interaction with the test subjects. Those analysing the curves blindly were detached from all such concerns. Furthermore, it cannot be ruled out that the examiners cope better when their skills of interrogation and detection of deception based on non-verbal hints surpass the average. However, in case of lack of such talents (and the capacity of an average human being to detect deception does not exceed chance (see reviews and meta-analysis by: Bond, DePaulo, 2006; Hartwig, C.F. Bond, 2011; Vrij, 2008) and moreover no human is perfectly resilient to bias) subjective factors may make a negative impact on interpretation of the charts.

Elimination of the subjective circumstances that primarily result from the direct interaction with the subject and the impact of information about that person and the subject of the examination obtained earlier seem therefore beneficial for the analytical process. At the same time it provides arguments in favour of teamwork. The conclusions of G. Barland demonstrate the advantages of such organisation of polygrapher work. While studying accuracy and validity of the tests performed in Baxter technique, Barland realised that, when totalled, the assessments of the evaluators of polygraph curves returned a higher accuracy than the average results for an individual polygraphers (86% and 81.7% correspondingly) (Barland 1972).

Blind interpretation of the charts is useful not only for mutual consultations between polygraphers but is likewise the fundamental element of formalised quality control procedures. For example, T. Shurany *et al.* believe that quality control of polygraph examinations should be conducted in three stages: blind assessment of

the charts followed by learning the details of the case to check whether the test questions were phrased correctly. Then the third stage consisting of the audio and video recording analysis assures that the polygraphers did not infringe standards of running such examinations (Shurany *et al.* 2009: 138).

Examining the delivery of biased opinions, Kassir *et al.* believe the procedure of introducing blind assessment necessary. Moreover, such a procedure must be sufficiently rigorous for the evaluator not to know whose material they have received, and the expert conducting the examination must have no influence on the selection of the controller. It is also necessary to provide training in fundamental psychology encompassing the questions of perception, judgements, and decision making (Kassir, Dror, Kukucka, 2013: 49–50).

Resorting to blind assessment should curb the impact of preconceptions and earlier expectations on interpreting test data. This is currently a routine procedure in the United States and Israel, yet it has not always been so. When asked whether the curves should be made available after an examination in 1950, C.M. Wilson, at the time the chair of the International Society for the Detection of Deception (ISDD), argued that it would only introduce unnecessary confusion, especially if they were presented to an untrained individual. He believed that the curves tell nothing to one who did not conduct the examination himself or herself. That is the reason why he never showed anyone his curves (Wilson 1950; Ansley 1999: 28). However, at the time, the diagnostic criteria were highly imprecise and applied inconsistently. With Wilson's effectively used, how could reliability of polygraph tests be discussed at all? How to counteract professional malpractice and ordinary abuse if no one had an opportunity to control the data the polygraphers used to issue a specific opinion?

It goes without saying that control of the content should be exerted by individuals possessing profound knowledge on polygraph examinations. This causes no problem when there are official quality control procedures and professional bodies nominated to exert such control. Difficulties may set in when such substantive control remains, on the power of law, a competence of bodies that lack such knowledge. For example, in the case of jurisdictional procedures, the task of substantial control of expert witness's opinions resides with the court. However, the reason why the court involves an expert witness is the court's lack of particular knowledge necessary to assess the specific problem, yet at the same time it is the court that must subsequently perform the substantial control of the evidence it has received from the expert witness.

Following upon the above considerations on the advantages of blind interpretation and the risks ensuing from subjectivism in polygraph examinations, it is worthwhile to consider making that blind interpretation an obligatory element of the process of delivering opinions in case of examinations conducted to provide evidence before the court. This would be a burden of practical nature, especially that, as much as in the case of polygrapher teams employed in various institutions, additional evaluators can fairly easily be appointed ad hoc, in the case of providing opinions for the court, it would be necessary to appoint not one but two independent expert witnesses, one of whom would conduct the examination and the other would be given the task of conducting a fully blind interpretation of the data. In the case of diverging opinions, the right to deliver the final decision would remain with the first, being the leading expert responsible for all the procedures conducted. Alternatively, the examination would need repeating or else a team of expert witnesses, members of a recognised specialist institution, could be involved to make the final opinion. I leave the problem open, thus encouraging a creative discussion among polygrapher and legal circles.

## References

- Ansley N. (1999), Development of Deception Criteria Prior to 1950. *Polygraph*, 28 (1).
- Barland G.H. (1972), The Reliability of Polygraph Chart Evaluations. *Polygraph*, 1 (4).
- Barland G.H., Raskin D.C. (1971), *An Experimental Study of Field Techniques in "Lie Detection"*, presentation delivered to Society for Psychophysiological Research, St. Louis, 24.10.1971; reprinted in: *Psychophysiology*, 9, 1972.
- Bersh P. (1969), A Validation Study of Polygraph Examiner Judgments. *Journal of Applied Psychology*, 53.
- Bond C.F., DePaulo B.M. (2006), Accuracy of deception judgments. *Personality and Social Psychology Review*, 10.
- Dror I.E., Charlton D., Pèron A.E. (2006), Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Science International*, 156.

Elaad E., Ginton A., Ben-Shakhar G. (1994), The Effects of Prior Expectations and Outcome Knowledge on Polygraph Examiner's Decisions. *Journal of Behavioral Decision Making*, 7.

Elaad E., Ginton A., Ben-Shakhar G. (1998), The Role of Prior Expectations in Polygraph Examiners Decisions. *Psychology, Crime & Law*, 4.

Ginton A. (2019), *Basic vs. Applied Psychology perspectives lead to different implications from the same data; reevaluating the impact of prior expectations on polygraph outcomes*, "Social Sciences & Humanities Open", 1, [https://www.researchgate.net/publication/337417831\\_Basic\\_vs\\_Applied\\_Psychology\\_perspectives\\_lead\\_to\\_different\\_implications\\_from\\_the\\_same\\_data\\_reevaluating\\_the\\_impact\\_of\\_prior\\_expectations\\_on\\_polygraph\\_outcomes](https://www.researchgate.net/publication/337417831_Basic_vs_Applied_Psychology_perspectives_lead_to_different_implications_from_the_same_data_reevaluating_the_impact_of_prior_expectations_on_polygraph_outcomes) (accessed: 30.05.2022).

Gougler M., Nelson R., Handler M., Krapohl D.J., Shaw P., Bierman L. (2011), Meta-Analytic Survey of Criterion Accuracy of Validated Polygraph Techniques. Report Prepared for the American Polygraph Association Board of Directors. *Polygraph* Special Edition, 40 (4).

Hartwig M., Bond C.F. (2011), Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, 13.

Holmes W.D. (1957), The degree of objectivity in chart interpretation, In: V.A. Leonard, *Academy Lectures on Lie-detection*, vol. II, Ch. Thomas, Springfield.

Honts C.R. (1996), Criterion Development and Validity of the Control Question Test in Field Application. *Journal of General Psychology*, 123.

Honts C.R., Raskin D.C. (1988), A Field Study of the Directed Lie Control Question. *Journal of Police Science and Administration*, 16 (1).

Horvath F. (1977), The Effects of Selected Variables on Interpretation of Polygraph Records. *Journal of Applied Psychology*, 62 (2).

Horvath F., Reid J. (1971), The reliability of polygraph examiner diagnosis of truth and deception. *Journal of Criminal Law, Criminology and Police Science*, 62 (1).

Kassin S.M., Dror I.E., Kukucka J. (2013), The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, 2 (1).

- Krapohl D.J., Dutton D.W. (2018), Believing is Seeing: The Influence of Expectations on Blind Scoring of Polygraph Data. *Polygraph*, 47 (2).
- Kubis J.F. (1962), *Studies in lie-detection computer feasibility considerations*, Fordham University, New York 1962 (RADC-TR-62-205, Project No. 5534, AF 30 (602)–2270, prepared for Rome Air Development Center, Air Force Systems Command, USAF Griffiss AFB, New York.
- Matte J.A. (1996), *Forensic Psychophysiology. Using the Polygraph: Scientific Truth Verification – Lie Detection*, J.A.M. Publications, New York.
- Orne M.T. (1973), Implications of Laboratory Research for the Detection of Deception. *Polygraph*, 2 (3).
- Patrick C.J., Iacono W.G. (1991), Validity of the Control Question Polygraph Test. The Problem of Sampling Bias. *Journal of Applied Psychology*, 76 (2).
- Raskin D.C., Kircher J.C., Honts C.R., Horowitz S.W. (1988), *A Study of the Validity of Polygraph Examinations in Criminal Investigations. Final Report to the National Institute of Justice*, Salt Lake City.
- Reid J., Inbau F. (1977), *Truth and Deception. The polygraph (lie-detector) technique*, William & Wilkins, Baltimore.
- Shurany T., Matte J.A., Stein E. (2009), Influence of Case Facts on Blind Scorers of Polygraph Tests. *European Polygraph*, Vol. 3, No. 3–4 (9–10).
- Vrij A. (2008), *Detecting lies and deceit: Pitfalls and opportunities*, Second Edition, Wiley.
- Wicklander D.E., Hunter F.L. (1975), The influence of auxiliary sources of information in polygraph diagnosis. *Journal of Police Science and Administration*, 3 (4).
- Wilson C.M. (1950), Should graphs be released or shown after test? *ISDD Bulletin*, 3 (3),
- Widacki J. (ed., 2018), *Kierunki rozwoju instrumentalnej i nieinstrumentalnej detekcji kłamstwa. Problemy kryminalistyczne, etyczne i prawne*, Oficyna Wydawnicza AFM, Kraków.