



**METODY INTERPRETACJI
DANYCH W BADANIACH
POLIGRAFICZNYCH**

PROBLEM SUBIEKTYWIZMU EKSPERTÓW PRZY OPINIOWANIU

MARCIN GOŁASZEWSKI

MARCIN GOŁASZEWSKI

**METODY INTERPRETACJI
DANYCH W BADANIACH
POLIGRAFICZNYCH**

PROBLEM SUBIEKTYWIZMU EKSPERTÓW PRZY OPINIOWANIU

KRAKÓW 2023

Rada Wydawnicza Krakowskiej Akademii im. Andrzeja Frycza Modrzewskiego:
Klemens Budzowski, Maria Kapiszewska, Zbigniew Maciąg, Jacek M. Majchrowski

Recenzja:
prof. dr hab. Tadeusz Tomaszewski

Redaktor prowadzący: Halina Baszak Jaroń

Redakcja: Filip Rekucki-Szczurek

Korekta: Carmen Stachowicz

Projekt okładki: Marcin Michałowski

Skład: Oleg Aleksejczuk

ISBN 978-83-67491-26-6
e-ISBN 978-83-67491-23-5

DOI: 10.48269/978-83-67491-23-5

Copyright© by Marcin Gołaszewski & Krakowska Akademia im. Andrzeja Frycza Modrzewskiego
Kraków 2023

Żadna część tej publikacji nie może być powielana ani magazynowana
w sposób umożliwiający ponowne wykorzystanie,
ani też rozpowszechniana w jakiegokolwiek formie
za pomocą środków elektronicznych, mechanicznych, kopiujących,
nagrywających i innych, bez uprzedniej pisemnej zgody właściciela praw autorskich

Na zlecenie



Krakowskiej Akademii
im. Andrzeja Frycza Modrzewskiego
www.ka.edu.pl

Wydawca: Oficyna Wydawnicza KAAFM, Kraków 2023

Druk sumptem Marcina Gołaszewskiego

Spis treści

Przedmowa	9
Wprowadzenie	11
I. Czynniki rzutujące na poprawność i zgodność ocen poligraferów	33
1. Obszary subiektywizmu i potencjalne elementy pozadyskursywne przy ekspertyzach z zakresu badań poligraficznych	33
1.1. Ogólne czynniki ujemnie wpływające na miarodajność testów poligraficznych, w tym subiektywizm badającego	36
1.2. Uprozczone reguły myślenia, błędy postrzegania	43
1.3. Presja otoczenia, wpływ informacji przekazywanych biegłemu na późniejsze opiniowanie	50
2. Niedokładność i błędy pomiaru poligrafu	70
2.1. Istota pomiaru	70
2.2. Warunki w środowisku badawczym	72
2.3. Oprzyrządowanie wykorzystywane w badaniach poligraficznych	74
2.4. Pomiar poligrafu, przetwarzanie sygnału i rejestrowanie danych	76
3. Zakres niezbędnych wiadomości specjalnych biegłego z zakresu badań poligraficznych	83
3.1. Weryfikacja wiarygodności biegłego z zakresu badań poligraficznych jako źródła dowodowego	83
3.2. Zdefiniowanie kompetencji poligrafera w standardach profesjonalnych, standardy kształcenia	87
4. Znaczenie osobistych właściwości poligrafera	92

4.1. Umiejętność odpowiedniego zaprezentowania wyników badań	92
4.2. Psychologiczne predyspozycje kandydata na poligrafera i rozwój zawodowy	94
II. Dokładność identyfikacji i zgodność między poligraferami przy różnych metodach interpretacji zapisów na poligramach	109
1. Właściwości testów poligraficznych i ewolucja metod analitycznych	109
2. Główne metody interpretacji danych w badaniach poligraficznych	113
2.1. Metoda jakościowa	115
2.2. Metody numeryczne (ilościowo-jakościowe)	121
2.2.1. Klasyfikowanie zmian w reakcjach metodą przypisywania rang (HSS i ROSS)	126
2.2.2. System Lykkena	130
2.2.3. System analizy według Rządu Federalnego USA (tzw. federalny)	133
2.2.4. System UTAH	137
2.2.5. Empiryczny System Oceniania (ESS) i jego zaktualizowana wersja (ESS-M)	140
2.2.6. Znaczenie konsekwencji, spójności w sposobie reagowania	155
3. Ewaluacje badających vs. ślepa interpretacja zapisów na poligramach	158
4. Algorytmy komputerowe	162
4.1. Algorytmiczna całościowa analiza danych testowych	162
4.2. Narzędzia wspomagające manualną interpretację niektórych danych	177
5. Badania empiryczne dotyczące zgodności ocen i diagnoz biegłych z zakresu badań poligraficznych	181
5.1. Zgodność ocen poligraferów – przegląd badań w piśmiennictwie	181
5.2. Własne badania dotyczące dokładności i zgodności diagnoz poligraferów posługujących się różnymi metodami analitycznymi	192

III. Sposoby ograniczenia marginesu subiektywizmu w badaniach poligraficznych	201
1. Walidacja i standaryzacja testów, ujednoczenie procedur badań poligraficznych	201
1.1. Trafność i rzetelność testów	201
1.2. Walidacja metod badań poligraficznych	205
2. Doprecyzowanie kryteriów diagnostycznych	209
2.1. Naukowe potwierdzenie cech diagnostycznych zapisów na poligramach – rys historyczny	209
2.2. Przesłanki diagnozowania w systemie Backstera	211
2.3. Cechy diagnostyczne wyodrębnione przez szkołę USAMPS, system Rządu Federalnego USA	216
2.4. Wytyczne w systemie Utah	225
2.5. Kryteria diagnostyczne w ESS i ESS-M	226
3. Zapewnienie jakości badań poligraficznych, znaczenie pracy zespołowej	228
3.1. Kontrola sądowa opinii biegłego	228
3.2. Instytucjonalne procedury kontroli jakości badań poligraficznych	232
 Podsumowanie	 241
Bibliografia	253

Przedmowa

Niniejsza książka powstała na podstawie rozprawy doktorskiej pod kierunkiem prof. dra hab. Jana Widackiego, pomyślnie obronionej dnia 1 grudnia 2021 r. na Wydziale Prawa, Administracji i Stosunków Międzynarodowych Krakowskiej Akademii im. Andrzeja Frycza Modrzewskiego. Korzystając z okazji, w tym miejscu serdecznie dziękuję Promotorowi za inspirację w wyborze tematu rozprawy i motywowanie do dokończenia dzieła. Wdzięczność wyrażam również wobec Recenzentów – prof. dra hab. Tadeusza Tomaszewskiego z Uniwersytetu Warszawskiego i dra hab. Marka Leśniaka – profesora Uniwersytetu Śląskiego. Przekazane uwagi przyjąłem z pokorą i szczerze je doceniam.

Dziękuję również Koleżankom i Kolegom – poligraferom, którzy wzięli udział w badaniach empirycznych, bez których rozprawa nie mogłaby powstać. Dziękuję najbliższej Rodzinie – za pomoc w tworzeniu warunków do pracy naukowej i wiarę w moje możliwości, Przyjaciołom – za wspierającą obecność w moim życiu. Prosto z serca dedykuję tę książkę niezwykle ważnej dla mnie Osobie, która sprawiła, że spojrzałem na świat przez różowe okulary.

Podjęta przeze mnie tematyka jest wyrazem ciekawości naukowej – z jednej strony, z drugiej zaś – konsekwencją własnych doświadczeń z praktyki zawodowej badań psychofizjologicznych z wykorzystaniem poligrafu. Jeżeli badam, to po to, aby uzyskać jakąś wiedzę samą w sobie, albo zmierzam do ustalenia, które ma wesprzeć określoną decyzję czy działanie. Odpowiednia interpretacja danych zebranych przez badającego jest kluczowym zadaniem w ramach całego procesu badawczego. Wszechstronna znajomość metod interpretacji danych testowych jest ważna dla poligraferów ze względu na potrzebę dokładnego i rzetelnego diagnozowania wiarygodności. Świadomość istnienia przy tym subiektywizmu (i z czego to może wynikać) ma

znaczenie dla praktyków, dla zlecających badania poligraficzne oraz stron zainteresowanych treścią opinii poligrafera w kontekście prawidłowej oceny dowodu.

Sądzę, że dla osób szkolących się w zakresie badań poligraficznych ta książka będzie lekturą obowiązkową. Doświadczonym poligraferom również może przydać się w codziennej praktyce jako uporządkowane kompendium wiedzy o metodach analizy danych testowych. Odbiorcy opinii mogą odnaleźć tu wskazówki, o co i jak można by potem biegłego fachowo zapytać. Liczę ponadto na udział w dalszej popularyzacji badań poligraficznych oraz inspirację poligraferów do refleksji w kierunku dążenia do maksymalnego obiektywizmu w procedurach opiniowania, a także do zaangażowania się w kolejne badania naukowe, które jeszcze bardziej wzmocnią metodologiczne fundamenty przedmiotowej specjalności.

Wprowadzenie

Badania poligraficzne (psychofizjologiczne przy użyciu poligrafu) należą do czynności z obszaru kryminalistyki. Zostały zaprojektowane przede wszystkim na potrzeby śledcze, choć z czasem znalazły również inne zastosowania – głównie w procedurach kadrowych (przedzatrudnieniowych i w ramach kontroli wewnętrznej w danej instytucji), a także w pracy operacyjnej służb (przy weryfikacji wiarygodności osobowych źródeł informacji) – oraz na rozmaite potrzeby prywatne (np. przy próbach sprawdzenia lojalności partnera czy ustaleniach w sprawie okoliczności jakichś strat materialnych w przedsiębiorstwie). Przydatność badań poligraficznych wykazano ponadto w nadzorze i terapii osób skazanych za przestępstwa przeciwko wolności seksualnej (obecnie w USA i Wielkiej Brytanii – jako jeden z warunków przedterminowego zwolnienia)¹.

W niniejszej pracy badania poligraficzne autor rozpatrywał przede wszystkim z punktu widzenia celów procesu karnego. Wykorzystane jako dowód, zgodnie art. 2 k.p.k.², mają więc przyczynić się do tego, by

¹ Por. m.in. M. Gołaszewski, *Użyteczność badań poligraficznych w ograniczaniu recydywy wśród sprawców przestępstw przeciwko wolności seksualnej*, „Prokuratura i Prawo”, 12, 2016, s. 5–22; A. Leszczyńska, *Użyteczność badań poligraficznych w terapii i kontroli osób skazanych za przestępstwa przeciwko wolności seksualnej*, „Seksuologia Polska”, 11(2), 2013, s. 48–55; D.T. Wilcox, *The Use of the Polygraph in Assessing, Treating and Supervising Sex Offenders. A Practitioner’s Guide*, Chichester 2009; D.T. Wilcox, R. Gray, *The Use of the Polygraph with Sex Offenders in the UK*, „European Polygraph”, 6, 1(19), 2012, s. 55–67; E.H. Meijer, B. Verschuere, H.L.G. Merckelbach, G. Crombez, *Sex Offender Management Using the Polygraph. A Critical Review*, „International Journal of Law and Psychiatry”, 31(5), 2008, s. 423–429; D. Grubin, M. Kamenskov, R. Dwyer, T. Stephenson, *Post-conviction Polygraph Testing of Sex Offenders*, „International Review of Psychiatry”, 31(2), 2019, s. 141–148.

² Ustawa z dnia 6 czerwca 1997 r. – Kodeks postępowania karnego (Dz.U. z 1997 r., nr 89, poz. 555).

sprawca przestępstwa został wykryty i pociągnięty do odpowiedzialności karnej, a osoba niewinna nie poniosła tej odpowiedzialności. Mogą ponadto wydatnie wspomóc organy procesowe w doprowadzeniu do rozstrzygnięcia sprawy w rozsądnym terminie poprzez eliminację nietrafnych wersji śledczych i w konsekwencji – zmniejszenie liczby czynności procesowych do wykonania. W polskiej procedurze karnej badania poligraficzne wykonywane są w postaci ekspertyzy – dla celów eliminacyjnych i dowodowych (art. 192a § 2 k.p.k.) bądź tylko dowodowych (art. 199a k.p.k.). Zbadanie za pomocą poligrafu reakcji na krytyczne bodźce testowe, ocena znaczenia tych reakcji oraz wnioskowanie na tej podstawie o związku osoby badanej z określonym zdarzeniem i wiedzy o szczegółach tego zdarzenia wymagają niewątpliwie wiadomości specjalnych, a wówczas zasięga się opinii biegłego (art. 193 k.p.k.). Biegły może być za takowego uznany, jeśli posiada specjalistyczną wiedzę, wykszolenie i wystarczające doświadczenie w danym obszarze. Zagadnienie kwalifikacji biegłego z zakresu badań poligraficznych – jako jeden z podstawowych czynników rzutuujących na poprawność i zgodność ocen biegłych-poligraferów (dalej będą używał tych nazw zamiennie) – nieco szerzej został omówiony w części I monografii.

Każde badanie wykonane przez biegłego w ramach ekspertyzy prowadzi do uzyskania dowodu pośredniego – czyli takiego, który prowadzi do ustaleń, na podstawie których można w drodze rozumowania redukcyjnego wyciągać wnioski na temat faktu głównego. Nie inaczej jest w przypadku opinii z badań poligraficznych:

Ustalenie dokonane w badaniu poligraficznym wymaga dopiero logicznego powiązania z faktem głównym, w szczególności wykazania poprzez inne dowody, że bardziej prawdopodobne jest, iż odnotowane reakcje badanego na pytania krytyczne są wynikiem tego, że na pytania krytyczne odpowiadał on nieszczerze, niż że są jakimś artefaktem o nieznamnej przyczynie. Może to wbrew pozorom, ale identycznie postępujemy z dowodem z badań DNA. Ujawnienie śladu DNA na odzieży podejrzanego wymaga stwierdzenia, że bardziej prawdopodobne jest, iż znalazł się on tam w związku z zabójstwem niż w innych okolicznościach³.

Dowód może przybrać charakter zarówno odciążający, jak i obciążający. Zgodnie z art. 7 k.p.k. dowody podlegają swobodnej ocenie

³ J. Widacki, *Opinia z badań poligraficznych w procesie karnym*, „Palestra”, 7–8, 2017, s. 10.

przez organ postępowania, z uwzględnieniem zasad prawidłowego rozumowania oraz wskazań wiedzy i doświadczenia życiowego. Regulacje ustawy procesowej nie zawierają żadnych dyrektyw, które nakazywałyby określone ustosunkowanie się do konkretnych dowodów, jak również nie wprowadzają różnic co do wartości poszczególnych dowodów⁴. Natomiast – mając na uwadze art. 170 k.p.k. – okoliczność, która ma być udowodniona, musi mieć znaczenie dla rozstrzygnięcia sprawy, a dowód być przydatny do stwierdzenia tej okoliczności.

Ekspertyzę poligraficzną można zlecić na dowolnym etapie postępowania karnego – wobec podejrzanego, świadka, oskarżonego, a nawet – wyjątkowo – osoby o nieustalonym jeszcze statusie procesowym (por. art. 192a k.p.k.)⁵. Po stronie ograniczeń pozostają: konieczność uzyskania zgody osoby badanej oraz zakaz dowodowy łączenia badania poligraficznego z czynnością procesową przesłuchania, wynikający z art. 171 § 5 ust. 2 k.p.k.⁶

Sąd Najwyższy sprecyzował, że zakaz ten dotyczy nie tylko samej czynności procesowej przesłuchania, lecz także czynności pozostających w bezpośredniej relacji z przesłuchaniem. Za niedopuszczalne uznał przeprowadzanie badania bezpośrednio przed lub bezpośrednio po przesłuchaniu, kiedy mogłoby ono wpływać na swobodę wypowiedzi osoby przesłuchiwanej, stanowić rodzaj nacisku lub zagrożenia z jej punktu widzenia⁷. W praktyce amerykańskiej, po zakończeniu fazy testów, płynnie przechodzi się do fazy przesłuchania. Natomiast w modelu europejskim badanie poligraficzne ma charakter bardziej

⁴ A. Sakowicz (red.), *Kodeks postępowania karnego. Komentarz*, wyd. 7, Warszawa 2016, s. 46.

⁵ Por. J. Widacki, *Opinia z badań poligraficznych...*, *op. cit.*, s. 5–10; M. Gołaszewski, *Badania poligraficzne w polskim procesie karnym po Postanowieniu Sądu Najwyższego z dnia 29 stycznia 2015 r.*, „Przegląd Bezpieczeństwa Wewnętrznego”, 13(7), 2015, s. 139–150.

⁶ Niezależnie od dwóch przepisów Kodeksu postępowania karnego (192a § 2 i 199a), w których wskazuje się na „środki techniczne mające na celu kontrolę nieświadomych reakcji organizmu”, podstawą powołania biegłego z zakresu badań poligraficznych w procedurze karnej jest przede wszystkim art. 193 k.p.k. Analogicznie – art. 278 k.p.c. oraz art. 84 k.p.a. – gdy tylko organ postępowania uzna, że w sprawie wymagane są wiadomości specjalne (w tym z zakresu badań poligraficznych). Nie istnieją przepisy zakazujące użycia poligrafu w jakimkolwiek postępowaniu (z wyjątkiem ograniczenia wynikającego z art. 171 § 5 pkt. 2 k.p.k.).

⁷ Postanowienie SN z dnia 29 stycznia 2015 r.

kliniczny (diagnostyczny). Nie zawiera zwykle elementów konfrontowania z faktami. Po testach, zamiast przejścia w fazę przesłuchania, umożliwia się osobie badanej złożenie uzupełniających oświadczeń, doprecyzowanie nieścisłości czy odniesienie się do przyczyn zaistnienia ewentualnych znaczących zmian reakcji fizjologicznych na określone bodźce, przy zachowaniu pełnej swobody wypowiedzi.

Badania poligraficzne przeprowadza się z wykorzystaniem specjalistycznej aparatury w postaci poligrafu, który mierzy zmiany różnych reakcji fizjologicznych oraz aktywność motoryczną, mięśniową osoby badanej. Istotą tych pomiarów jest to, że owe reakcje (poza ruchami ciała) występują niezależnie od woli człowieka i normalnie pozostają nieuświadomiane, ponieważ odpowiada za nie autonomiczny układ nerwowy⁸. Poligraf należy więc zaliczyć do grupy określanej przez Kodeks postępowania karnego jako *środki techniczne mające na celu kontrolę nieświadomych reakcji organizmu*.

Wyłącznie w Polsce znaczna część środowiska prawniczego (i nie tylko) posługuje się nazwą *wariograf*, wprowadzoną przez prof. Pawła Horoszowskiego na początku lat 60. Nazwa ta (o etymologii łacińsko-greckiej) – w odróżnieniu od słowa *poligraf* (o etymologii w całości greckiej), niesłusznie uznawanego za kalkę z języka angielskiego – miała pozwolić na uniknięcie konotacji związanych z przemysłem drukarskim. Choć obie nazwy są synonimiczne (oznaczają różnorodność, wielość zapisów) – należy zważyć, że już w XIX wieku N. Cybulski mianem poligrafu określa urządzenie rejestrujące więcej niż jedną funkcję organizmu. Poligraf Mackenziego, skonstruowany w pierwszej dekadzie XX wieku, nie był więc pierwszym urządzeniem o tej nazwie. Krytycy nazwy wariograf podnoszą ponadto, że w krajach, gdzie mówi się innymi językami niż polski – i funkcjonuje tam wyłącznie nazwa poligraf – wieloznaczność tego wyrazu, odnoszącego się m.in. również do drukarstwa, nie stanowi problemu⁹. Ponadto pomysłodawca określenia wariograf nie był ani konstruktorem, ani producentem takiego sprzętu. Nie miał też szczególnych zasług we wprowadzaniu badań poligraficznych

⁸ O tym, jak autonomiczny układ nerwowy (włącznie z kluczowym elementem jego części przywspółczulnej – nerwem błędnym) reguluje procesy psychofizjologiczne, przejrzyście mówi zapoczątkowana w 1994 r. przez S. Porgesa – teoria poliwałgalna – zob. *idem, Teoria poliwałgalna*, Kraków 2020.

⁹ Por. J. Widacki, *Poligraf czy wariograf*, „Problemy Kryminalistyki”, 294(4), 2016, s. 35–39.

do polskiej praktyki śledczej (wprawdzie wykonał pierwsze takie ekspertyzy, ale nie ustrzegł się przy tym poważnych błędów w sztuce)¹⁰.

Czytelnicy mogą zwrócić uwagę na niejednorodną terminologię w odniesieniu do celu badania poligraficznego (np. „detekcja nieuczciwości”, „ocena wiarygodności”, „ustalenie czy przy pytaniach krytycznych zaistniały znaczące zmiany w reakcjach fizjologicznych”). Ściśle rzecz biorąc – celem bezpośrednim badania poligraficznego jest ustalenie sposobu reagowania na określone bodźce testowe na podstawie niezależnych od woli człowieka zmian reakcji fizjologicznych organizmu, za które odpowiada autonomiczny układ nerwowy. Natomiast cel pośredni wyznacza kontekst, w którego ramach prowadzone jest badanie, a ten bywa różny. W sprawach karnych np. chodzi o wiedzę i związek z określonym zdarzeniem. W postępowaniach rekrutacyjnych badanie poligraficzne służy poszerzeniu informacji na temat kandydata do pracy lub służby i wprost ocenie jego wiarygodności.

¹⁰ Autor niniejszej monografii chciałby wyjaśnić w odniesieniu do wspomnianego prof. P. Horoszowskiego, że nie kwestionuje jego zasług dla kryminalistyki czy kryminologii: jego prace należały do pierwszych polskich nowoczesnych podręczników do tych przedmiotów. Autor docenia także książkę *Od zbrodni do kary*, która spopularyzowała współczesną kryminalistykę i przybliżyła ją całemu pokoleniu prawników, a to, że Profesor zmuszono do opuszczenia Polski w 1968 r. – było ewidentnym wyrządzeniem mu krzywdy. Podobnie jak ośrodek krakowski autor ma jednak nieco inne zdanie na temat zasług P. Horoszowskiego w zakresie badań poligraficznych. Prawdą jest, że P. Horoszowski, który jeszcze w podręczniku nader krytycznie i ideologicznie podchodził do poligrafu, po stypendium w USA radykalnie zmienił zdanie, zakupił poligraf i przywiózł go do Polski. Środowisko poligraferów, które i autor niniejszej monografii reprezentuje, dumnie obchodziło 50-lecie badań poligraficznych w polskim procesie karnym i to właśnie P. Horoszowski jako pierwszy takie badanie przeprowadził. Natomiast prawdą jest również, że uczynił to bez specjalistycznego przeszkolenia, a tym bardziej doświadczenia. Nie wykonał do tego czasu żadnych własnych badań eksperymentalnych, a mimo to przystąpił od razu do badań w sprawach karnych. W efekcie popełnił szkolne błędy, narażając oskarżonych na pomyłkę sądową – stąd krytyczny stosunek autora, lecz bez wyolbrzymiania wspomnianych zarzutów, wszak P. Horoszowski uczciwie nazywał zastosowaną przez siebie metodę jako „eksperymentalno-testową”. Warto jednak zwrócić uwagę, że zwyczajowo prawo do wprowadzania własnego nazewnictwa przysługiwało wynalazcom aparatury i technik badań poligraficznych. Dziś spór o to, czy mamy do czynienia z poligrafem czy wariografem nie wydaje się zresztą najistotniejszy, a obie nazwy funkcjonują w Polsce (choć tylko w Polsce) równolegle.

Współczesny poligraf jest wyposażony w: a) dwa pneumografy, mierzące ruchy górnej i dolnej części klatki piersiowej, odzwierciedlające zmiany w cyklu oddechowym; b) mankiet do pomiaru ciśnienia krwi i tętna; c) elektrody mierzące zmiany w przewodności elektrycznej skóry (odruch skórno-galwaniczny – GSR/EDA); d) czujnik aktywności motorycznej (obligatoryjnie – na siedzisko, a fakultatywnie również pod stopy, przedramiona i czujnik aktywności mięśni żuchwy); e) fotopletysmograf (PPG/PLE), który mierzy reakcje naczynioruchowe (zmiany w amplitudzie pulsu).

Przedmiotem zainteresowania badającego jest sposób, w jaki osoba poddawana testom reaguje na zaprezentowane bodźce – zwykle werbalnie – jakimi są zadawane pytania albo powtarzane wyrażenia (opisujące jakieś przedmioty, miejsca czy ludzi), a rzadziej również bodźce wizualne. Sygnały odbierane równocześnie przez poszczególne czujniki poligrafu podlegają analogowemu lub cyfrowemu przetworzeniu i zobrazowaniu w postaci zapisów graficznych na poligramach. Procedura badania poligraficznego nie sprowadza się jednak wyłącznie do prostych pomiarów testowych. Wymaga szerszego zaangażowania poligrafera, który nie tylko wystąpi w roli operatora poligrafu, ale uprzednio opracuje pytania i taktykę czynności z udziałem badanego, przeprowadzi wywiad bezpośredni, a po testach dokona, określoną metodą, interpretacji zarejestrowanych danych w odniesieniu do tego, jak zachowywał się badany, jak odpowiadał na pytania.

W badaniach poligraficznych mamy do czynienia z identyfikacją śladów pamięciowych i emocjonalnych dotyczących określonego zdarczenia oraz związanego z tymi śladami (lub ich brakiem) kognitywnego zaangażowania osoby badanej przy bodźcach testowych (zwykle werbalnie zadawanych pytaniach) różnych kategorii (krytycznych i kontrolnych).

Zgodnie z aktualnie upowszechnioną i akceptowaną koncepcją *zróżnicowanego znaczenia* (z j. ang. *differential salience*) – poszczególne bodźce (pytania) zaprezentowane badanemu w teście poligraficznym są w różnym stopniu istotne dla badanego, mają owo zróżnicowane znaczenie¹¹. Obserwowana fizjologicznie *istotność* jest funkcją

¹¹ Zob. S. Senter, D. Weatherman, D. Krapohl, F. Horvath, *Psychological Set or Differential Salience. A Proposal for Reconciling Theory and Terminology in Polygraph Testing*, „Polygraph”, 39(2), 2010, s. 109–117; M. Gołaszewski, *Psychologiczne podstawy efektywności badań poligraficznych – nowe podejście w XXI w.*, „Studia Prawnicze. Rozprawy i Materiały”, 2(11), 2012, s. 101–112.

podstawowych procesów psychologicznych związanych z emocjami, aktywnością poznawczą (selektywną uwagą, pamięcią, wysiłkiem mentalnym) oraz uwarunkowaniem behawioralnym odnoszącym się do danego bodźca testowego¹².

Najważniejsza z punktu widzenia badań poligraficznych pamięć autobiograficzna obejmuje historię życia człowieka, dotyczy zdarzeń umiejscowionych w konkretnym miejscu i czasie. Uwaga jest filtrem, który dokonuje wstępnej oceny napływających informacji oraz umożliwia koncentrację na tym, co jest subiektywnie postrzegane za najistotniejsze przez odbiorcę.

Emocje wywoływane są u badanego w związku z cechami bodźca testowego (zazwyczaj treść i kontekst zadanego pytania). Mogą być one zróżnicowane i nie ma znaczenia czy będzie to strach, zniesmaczenie czy radość z oszukiwania. Ponadto pamięć przechowuje przeżycia emocjonalne towarzyszące określonemu zdarzeniu z przeszłości – zarówno te, które człowiek sam jest w stanie świadomie odczuwać i opisać (zakodowane w pamięci deklaratywnej), jak również pewien ogólny stan emocjonalny, którego nie da się scharakteryzować słowami.

Uwarunkowanie behawioralne sprawia natomiast, że w czasie określonego czynu powstaje pobudzenie emocjonalne jako reakcja bezwarunkowa związana z tym czynem i jego okolicznościami. Krytyczne pytanie testowe jest bodźcem warunkowym, który wywołuje pobudzenie emocjonalne – z jednej strony potencjalnie związane z czynnością kłamania, ponieważ relacja między kłamstwem a pobudzeniem emocjonalnym jest uwarunkowana w czasie socjalizacji. Z drugiej zaś strony – z powodu przywołania w pamięci podobnych stanów emocjonalnych i reakcji fizjologicznych, jakie wystąpiły w owym czasie¹³.

Pośród różnych technik badań poligraficznych dwie mają znaczenie podstawowe: *techniki pytań porównawczych* (CQT – z j. ang. *comparison questions technique*) – współcześnie najczęściej stosowane oraz *techniki wiedzy o czynie* (GKT – z j. ang. *guilty knowledge test* lub CIT – z j. ang. *concealed information test*).

¹² Zob. J. Kahn, R. Nelson, M. Handler, *An Exploration of Emotion and Cognition during Polygraph Testing*, „Polygraph”, 38(3), 2009, s. 184–197.

¹³ Szerzej na ten temat: J. Wojciechowski, *Podstawy teoretyczne powstawania reakcji w czasie badań poligraficznych*, „Problemy Kryminalistyki”, 276, 2012, s. 19–31.

Te pierwsze mają wyższą dokładność, bardziej wszechstronną użyteczność i małą podatność na upływ czasu od weryfikowanych okoliczności. Natomiast techniki wiedzy o czynie są możliwe do wykorzystania rzadziej – przede wszystkim ze względu na trudność w wytypowaniu odpowiedniej liczby szczegółów, które uczestnik danego zdarzenia (sprawca przestępstwa, świadek naoczny) mógł i powinien zapamiętać, a jednocześnie obiektywnie nie może ich znać osoba bez faktycznego związku z tym zdarzeniem. Do tego, z biegiem czasu, może następować zapominanie tych szczegółów przez osobę winną, w skrajnym wypadku nawet wyparcie ich ze świadomości. Z kolei osoby niewinne mogą pozyskiwać wiedzę na temat okoliczności sprawy z komunikatów medialnych, plotek czy podczas kolejnych czynności procesowych z ich udziałem.

W technikach CQT zwraca się uwagę na różnice w reakcjach badanego na dwa typy pytań. Pierwszy – to pytania związane ze sprawą, która jest przedmiotem badania, zwane pytaniami *krytycznymi, istotnymi, relewanтными* (z j. ang. *relevant*). Drugi typ – to tzw. pytania *porównawcze* (z j. ang. *comparison*), dawniej nazywane *kontrolnymi*, występujące w dwóch wersjach: z *prawdopodobnym kłamstwem* (PLC – z j. ang. *probable lie comparison*) i *nakierowanym* czy inaczej – *poleconym, wyreżyserowanym kłamstwem* (DLC – z j. ang. *directed lie comparison*). Pytanie PLC jest tak zaprezentowane i w takim kontekście, by badany prawdopodobnie nie udzielił do końca szczerzej odpowiedzi. Ewentualnie zakres pytania powinien być na tyle szeroki, by badany nie miał całkowitej pewności, że niczego nie pominął, a jeśli nawet odpowie szczerze – takie pytanie będzie wymagało większego kognitywnego zaangażowania niż przy pytaniach relewantnych. Z kolei pytanie DLC polega na tym, że badany otrzymuje instrukcję udzielania odpowiedzi odwrotnych w stosunku do postrzeganego przez siebie stanu faktycznego, w konsekwencji wywołując stan dysonansu poznawczego i angażując swoje zasoby poznawcze.

Osoba nieszczerza typowo reaguje w bardziej znaczący sposób na pytania krytyczne, przy których ma świadomość wprowadzania w błąd, niż na relatywnie mniej dla niej ważne pytania porównawcze. U osób odpowiadających na pytania relewantne zgodnie z prawdą zwykle jest odwrotnie – istotniejsze reakcje fizjologiczne występują przy pytaniach porównawczych, co do których są niepewni pełnej rzetelności udzielonej odpowiedzi, albo wprost nieszczerzy. W przypadku

badanych odpowiadających w przeciwstawnym sposób – tj. nieszczerych oraz prawdomównych – występują odmienne poziomy istotności (najczęściej ze względu na odczuwane zagrożenie) i kognitywnego zaangażowania, wynikające z treści różnych kategorii pytań wykorzystywanych w testach z pytaniami porównawczymi, dlatego test typu CQT sprawdza się jako skuteczna metoda identyfikacji.

Pomocniczo, uzupełniająco stosuje się także wszelkie techniki, które mają potwierdzenie naukowe i dokładność przewyższającą statystyczną szansę. Nie mogą samodzielnie (bez powiązania z testem diagnostycznym, skonkretyzowanym i znacznie dokładniejszym) służyć za podstawę do wydania określonej opinii na temat reakcji odnoszących się do szczerości badanego, ale bywają użyteczne z innych powodów. W tej grupie technik najczęściej pojawia się test *szczytowego napięcia* (POT – z j. ang. *peak of tension*), wykorzystywany jako test w celach wykrywczych, w *wydobywaniu* śladów pamięciowych osoby badanej, dotyczących nieznanymi śledczym okoliczności danego zdarzenia, naprowadzaniu na nowe wersje śledcze (np. na współsprawców danego przestępstwa czy miejsce ukrycia zwłok), dlatego nazywany jest testem *poszukującym, pogłębiającym czy rozbitcia*.

Skuteczność tzw. *testów rozpoznania* (CIT, GKT, POT) jest konsekwencją różnicy w znaczeniu kluczowego bodźca względem pozostałych (neutralnych) – z punktu widzenia osoby, która zna – *rozpoznaje* właśnie – umieszczony w sekwencji bodźców (w postaci pytań, opisów przedmiotów, wykazu osób, fotografii czy dźwięków) szczegół danego zdarzenia, chociaż temu przeczy.

Finalnie biegły ma za zadanie przedstawić zleceniodawcy zaistniałe w toku badania spostrzeżenia – przy czym nie wolno mu formułować konkluzji dotyczącej winy ani ocen prawnych, pozostających wyłączną domeną sądu. Biegły danej specjalności nie może również wykroczyć poza swoje kompetencje i wypowiadać się w zakresie wiadomości specjalnych przynależnych innej specjalności, a w przypadku badań poligraficznych dość często mamy do czynienia z sytuacją, kiedy poligrafer ma wykształcenie psychologiczne. Istnieje wówczas pokusa, aby uwzględnić w opinii twierdzenia odnoszące się np. do osobowości osoby badanej.

Jak wynika z art. 201 k.p.k., opinia biegłego w postępowaniu karnym (a w gruncie rzeczy – również w każdym innym) powinna być pełna, jasna i niesprzeczna wewnątrznie. Na przykładzie opinii

z zakresu badań poligraficznych można powiedzieć, że byłaby ona *niepełna*, gdyby biegły np. nie przedstawił odpowiedzi na pytania wyrażone w postanowieniu o dopuszczeniu dowodu (ale takie, na które mógł i powinien był odpowiedzieć). Zarzut niepełności opinii powstałby ponadto wtedy, gdyby biegły zawarł w niej wyłącznie konkluzje, bez uzasadnienia, opisu metod badawczych i sprawozdania z przebiegu badania, w tym wszystkich zadanych pytań testowych. Ważne jest, aby biegły miał zdolność przedstawienia powodów, dla których wybrał określoną teorię, techniki badawcze i sposób interpretacji danych pozyskanych w trakcie swoich czynności oraz jak to się ma do innych uznanych podejść. Każde odejście od standardów (np. modyfikacja sekwencji pytań w danym teście, ilości powtórzeń) czy odrzucenie jakiegoś wyniku powinno być należycie uzasadnione i – w miarę możliwości – poparte wynikami badań empirycznych.

Opinia *niejasna* to po prostu taka, której nie można zrozumieć – np. dlatego, że jej tezy są wieloznaczne, biegły posłużył się nielogicznymi argumentami albo sformułował wnioski bez oparcia w rezultatach przeprowadzonego badania. Nie należy tego mylić z taką niejasnością, kiedy skomplikowana materia – w połączeniu z brakiem wcześniejszego zaznajomienia uczestników postępowania choćby z podstawami danej metody – sprawia, że pomimo poprawnej metodologicznie opinii w dalszym ciągu niektórym może wydawać się niezrozumiała. Z kolei *sprzeczność wewnętrzna* w opinii występuje wtedy, gdy zawarte w niej tezy wzajemnie się wykluczają lub są rozbieżne z wynikami testów.

Na świecie, wraz z postępem naukowo-technologicznym, obserwuje się ogólną tendencję do wzrostu znaczenia dowodów z opinii biegłych. Biegli mają wspomagać organy postępowania w podejmowaniu trafniejszych decyzji, ale trzeba też pamiętać, że niektóre z dowodów – jak w przypadku badań DNA – są postrzegane za tak mocne, że ich nieprawidłowe przeprowadzenie może prostą drogą prowadzić do skutków sprzecznych z celami postępowania karnego – do błędnych wyroków. Dlatego tak doniosłe znaczenie mają zagadnienia naukowej walidacji, standaryzacji testów i kontroli jakości, które poruszam w części III niniejszej monografii.

Problematyczna wydaje się też właściwa ocena dowodu z opinii biegłego przez uczestników postępowania, a przede wszystkim sąd. Z jednej strony rolą sądu jest dokonanie formalnej i merytorycznej

oceny opinii przedstawionej przez biegłego, ale z drugiej – sąd nie posiada przecież wiadomości specjalnych, z którego to powodu biegły został powołany. Trudność bywa jeszcze większa, gdy w sprawie pojawią się opinie sprzeczne. Nawet pomimo istnienia kodeksowych narzędzi w postaci przesłuchania biegłego na rozprawie czy konfrontacji biegłych dokonanie właściwego wyboru jest w takiej sytuacji nie lada wyzwaniem, tym bardziej gdy jakiś obszar rzeczywistości czy typ badań kryminalistycznych jest w środowisku prawniczym mało znany – tak jak badania poligraficzne, które w procesie kształcenia przyszłych sędziów, prokuratorów czy adwokatów są wciąż poruszane zdawkowo. Wzięcie biegłego w krzyżowy ogień pytań przez strony postępowania (w szczególności przez obronę) wymaga zapoznania się z istotą przeprowadzonej ekspertyzy. Aby było to skuteczne, może zachodzić konieczność posiłkowania się wiedzą innego biegłego w danej dziedzinie.

Opinie biegłych z zakresu badań poligraficznych należą do kategorii tzw. *dowodów naukowych*. Tak się składa, że warunki określające dopuszczalność dowodu naukowego w procesie karnym kształtowały się historycznie właśnie w związku z badaniami poligraficznymi – w kraju, gdzie miały miejsce narodziny poligrafu, czyli w USA. W 1923 r. obrońca oskarżonego o zabójstwo J. Frye’a wnosił w Federalnym Sądzie Apelacyjnym w Waszyngtonie o poddanie swojego klienta testowi szczerości opierającemu się na zmianach ciśnienia krwi mierzonego sfigmomanometrem¹⁴. Test miał przeprowadzić W. Marston. Wypełniona po brzegi sala rozpraw czekała na debiut urzędnika do wykrywania kłamstw¹⁵. Sąd sformułował wówczas kryterium wskazujące, że nowa metoda badawcza powinna mieć odpowiednie podstawy naukowe i uzyskać powszechną akceptację pośród przedstawicieli danej dziedziny. Psychofizjologiczna detekcja nieszczerości dopiero zaczynała się rozwijać i postawione przez sąd warunki nie mogły zostać spełnione, a dowodu ostatecznie nie dopuszczono. Po raz pierwszy aprobatą testu szczerości (z wykorzystaniem psychogalwanometru)

¹⁴ 293 F. 1013 (D.C. Cir. 1923), casetext.com/case/frye-v-united-states-7 [dostęp: 20.05.2020].

¹⁵ Szerzej na temat okoliczności sprawy i mitów na ten temat (jedne wersje mówią o winie Frye’a, inne o jego niewinności, podawane są też różne powody, dla których skazany ostatecznie uniknął kary śmierci): B.A. Stern, D.J. Krapohl, *The Infamous James Alphonso Frye*, „Polygraph”, 32(3), 2003, s. 188–199.

w amerykańskim sądzie nastąpiła w hrabstwie Queens (stan Nowy Jork) w 1938 r. w sprawie *People v. Kenny*¹⁶. Badanie przeprowadził W.G. Summers, dziekan wydziału psychologii na Uniwersytecie Fordham. Sąd wziął pod uwagę przedstawione przez biegłego wyniki eksperymentów laboratoryjnych, które ukazywały wręcz zadziwiająco wysoką dokładność nowej metody diagnozowania wiarygodności (blisko 100%), a biegły jeszcze podkreślał, że reakcje emocjonalne w sprawach rzeczywistych są bardziej intensywne od tych eksperymentalnych, a przez to jego ocena jest bezbłędna. Prokurator okręgowy docenił wówczas wartość naukową i użyteczność badań przy użyciu psychogalwanometru, ale sprzeciwiał się dopuszczeniu dowodu z opinii biegłego, powołując się na to, że w dalszym ciągu trwają jeszcze eksperymenty i za wcześnie, by mówić o powszechnej akceptacji dla tego typu praktyk. Dopuszczając dowód, sąd zauważył, że biegli z innych uznawanych dowodowo specjalności niejednokrotnie różnią się w swoich interpretacjach, a to ława przysięgłych decyduje w konkretnym przypadku, czy zeznania biegłego zasługują na wiarę. Jeśli więc sądy akceptują opiniowanie psychiatryczne, czy pismoznawcze – to powinny również dopuszczać dowód z badań psychogalwanometrem, kiedy będą ku temu odpowiednie podstawy. W komentarzu do tego postanowienia, który ukazał się w tym samym roku na łamach „Journal of Criminal Law and Criminology”, autor (H. Jordan) poszukiwał kompromisowego rozwiązania, które polegałoby na dopuszczeniu takiego dowodu tylko wtedy, gdy oskarżenie i obrona wyrażą wzajemną zgodę na konkretny test oraz jego wyniki¹⁷.

Poligraf tego typu, jaki współcześnie znamy (a więc składający się z czujników mierzących co najmniej zmiany w cyklu oddechowym, ciśnieniu krwi i tętnie oraz reakcje elektrodermalne), powstał najprawdopodobniej w latach 30. Prawdopodobnie, a nie na pewno – ponieważ w piśmiennictwie obecne są różne daty (od 1935 do nawet 1949 r.), wskazujące na to, kiedy L. Keeler uzupełnił poligraf dwukanałowy o psychogalwanometr¹⁸. Wcześniej, w 1925 r., opatentował sposób rejestracji reakcji (w kanałach: oddechowym i kardio) na rolce

¹⁶ *People v. Kenny*, 167 Misc. 51, 3 N.Y.S.2d 348 (1938), casetext.com/case/people-v-kenny-19 [dostęp: 20.05.2020].

¹⁷ H.W. Jordan, *Evidence – Admissibility of Deception („Lie Detector”) Tests*, „The Journal of Criminal Law and Criminology”, 29, 1938, s. 291.

¹⁸ Szerzej na ten temat: J. Widacki, *Historia badań poligraficznych*, Kraków 2017.

papieru, przesuwającego się ze stałą prędkością, a także technikę zapisu reakcji kardiograficznej z wykorzystaniem mankietu i specjalnie skonstruowanego mieszka. Natomiast historycznie twórcą pierwszego poligrafu dwukanałowego (charakteryzującego się ciągłym zapisem parametrów fizjologicznych) był J. Larson. Sposób pomiarów oraz zadawania pytań na przykładzie jednej ze spraw kryminalnych, do której został zaangażowany, opisał w artykule zatytułowanym: *Modification of the Marston Deception Test*, opublikowanym w „Journal of Criminal Law and Criminology”¹⁹. Skorygował i poszerzył koncepcję innego amerykańskiego uczonego – W. Marstona²⁰. W almanachu encyklopedii Britannica z 2003 r. dzieło J. Larsona figuruje na liście 325 największych wynalazków wszechczasów²¹.

Wraz z rozwojem metodologii psychofizjologicznej detekcji nieszczerości próbowano przełamywać niechęć do niej środowiska prawniczego w różnych miejscach na świecie. Sceptycyzm wynikał nie tylko z wciąż wówczas słabych jeszcze podstaw naukowych w tej dziedzinie, ale także z obaw o ingerencję w ciało człowieka, jego prywatność oraz zagwarantowany w prawie zakaz zmuszania kogokolwiek do samooskarżenia. Podejście do tego tematu zmieniało się stopniowo. Zaczęto dopuszczać instrumentalną detekcję nieszczerości – na początek w sytuacjach wyjątkowych. Na przykład prokurator generalny Sądu Najwyższego w Holandii G. Langemeijer (sprawujący tę funkcję w latach 1957–1973) uważał, że gdy w grę wchodzi żywotne interesy obywateli, powinna istnieć możliwość wydobycia oświadczenia od podejrzanego albo niewspółpracującego świadka takimi środkami jak *wykrywacz kłamstw* czy narkoanaliza²². W grupie środków tego typu warto by wymienić także hipnozę²³.

¹⁹ J.A. Larson, *Modification of the Marston Deception Test*, „The Journal of Criminal Law and Criminology”, 12(3), 1922, s. 390–399.

²⁰ Zob. W.M. Marston, *Systolic Blood Pressure Symptoms of Deception*, „Journal of Experimental Psychology”, 2(2), 1917, s. 117–163.

²¹ *What were the greatest inventions of all time?*, www.edinformatics.com/inventions_inventors [dostęp: 19.05.2020].

²² F. Meyjes, *Scientific Criminal Investigation Techniques under Dutch Law*, „Journal of Criminal Law and Criminology”, 51(6), 1961, s. 658. Z kolei na temat samej narkoanalizy zob. F.E. Inbau, *Scientific Evidence in Criminal Cases*, „Journal of Criminal Law and Criminology”, 24(6), 1934, s. 1153–1157.

²³ Zob. M. Niedbała, *Próby stwierdzenia istnienia w pamięci śladów za pomocą hipnozy i ich wykorzystanie w procesie*, [w:] D. Moskal, P. Wąsik (red.),

W polskim procesie karnym poligraf (wariograf) zastosowano po raz pierwszy w listopadzie 1963 r. w sprawie o zabójstwo, która toczyła się przed Sądem Wojewódzkim w Olsztynie. Badanie przeprowadził wówczas prof. P. Horoszowski, a jego dwie opinie zaliczono w poczet dowodów w sprawie. Sąd w uzasadnieniu stwierdził:

Metoda badań wariograficznych jest metodą naukową i nie można jej pominąć tylko dlatego, że nie była dotychczas u nas stosowana [...] wyniki badań wariograficznych w konkretnej sprawie są zgodne z innymi przytoczonymi dowodami i stanowią jedno z ogniw dowodów²⁴.

W USA na szczeblu ogólnokrajowym dopiero w 1989 r. Federalny Sąd Apelacyjny z 11. okręgu – w sprawie *Stany Zjednoczone vs. Piccinonna* – orzekł, że sam z siebie dowód z badań poligraficznych nie jest już całkowicie niedopuszczalny i wskazał dwa wyjątki²⁵. Pierwszy – jeśli obie strony postępowania uprzednio zgodzą się w kwestii okoliczności i zakresu wykorzystania tego dowodu. Drugi – gdy podważy się zeznania świadka w procesie i jedna ze stron, która chce dopuszczenia dowodu, daje taką możliwość również stronie przeciwnej, z zaangażowaniem innego wskazanego przez nią biegłego. Trzeba przy tym zaznaczyć, że w procedurze anglosaskiej śledztwo nie jest częścią procesu. Ten zaczyna się przed sądem. W procesie poligraf wykorzystywany jest w celach dowodowych, natomiast w śledztwie – do celów eliminacyjnych. Stąd tak szerokie rozpowszechnienie badań poligraficznych w USA na etapie śledztwa i jedynie wyjątkowa dopuszczalność dowodu uzyskanego na ich podstawie w sądzie.

W 1993 r. zasada powszechnej akceptacji została zastąpiona w USA tzw. regułą Dauberta. Sąd Najwyższy USA w sprawie *Daubert vs. Merrell Dow Pharmaceuticals, Inc.* uznał, że dotychczasowe kryterium dopuszczalności dowodu naukowego jest zbyt restrykcyjne²⁶ i okre-

Ślady pamięciowe. Aspekty psychologiczne i kryminalistyczne, Kraków 2013, s. 17–23.

²⁴ IV K 94/63; celem biegłego było ustalenie „czy i jakie objawy charakterystyczne dla przeżyć emocjonalnych związanych z chęcią zatajenia pewnych faktów występują w krzywych zarejestrowanych przez wariograf w czasie badania podejrzanych”.

²⁵ 885 F.2d 1529 (11th Cir. 1989), [casetext.com/case/us-v-piccinonna?](http://casetext.com/case/us-v-piccinonna) [dostęp: 21.05.2020].

²⁶ 113 S Ct 2786 (1993), [casetext.com/case/daubert-v-merrell-dow-pharmaceuticals-inc?](http://casetext.com/case/daubert-v-merrell-dow-pharmaceuticals-inc) [dostęp: 21.05.2020].

ślił nowe warunki. Wśród nich była możliwość weryfikacji przedstawionej teorii czy metody przez formułowanie hipotez i zastosowanie eksperymentów oraz obserwacji. Ponadto sąd powinien mieć możliwość zapoznania się ze znanym lub potencjalnym poziomem błędu, standardami posługiwania się analizowaną metodą oraz stopniem jej akceptacji w określonym środowisku naukowym. Rozważenia wymaga również, czy metoda została opublikowana i poddana niezależnej recenzji. Żadne z tych kryteriów z osobna nie gwarantuje jeszcze odpowiedniej jakości dowodu naukowego i należy je rozpatrywać łącznie. Jeżeli mowa o szerokiej akceptacji, to – rzecz jasna – wskazane jest, aby wypowiedzieli się nie tylko przedstawiciele danej specjalności, ale również osoby, które nie mają w tym subiektywnych interesów. Niezależne recenzje nie świadczą jeszcze o niezawodności metody, ale ich brak od razu implikuje konieczność zachowania ostrożności. Najważniejsza wydaje się jednak empiryczna weryfikacja danej metody.

Zbliżone kryteria, niezbędne dla uznania jakiejś teorii za naukową, można spotkać w literaturze. Na przykład A. Kenny podaje, że musi być spójna, o ustalonej metodologii, kumulatywna i przewidywalna²⁷. Spójność oznacza, że w podstawowych kwestiach eksperci nie powinni regularnie przedstawiać sprzecznych odpowiedzi. Zastosowanie odpowiedniej metodologii ma prowadzić do tego, żeby procedura badań miała charakter powtarzalny. Kumulatywność należy rozumieć w ten sposób, że kolejni naukowcy nadbudowują własne ustalenia na istniejących podstawach, które zbudowali poprzednicy. Z kolei przewidywalność – wedle wskazanego autora – wiąże się z możliwością weryfikacji danej teorii – przede wszystkim testowania okoliczności, które mogłyby ją obalić.

W. Thomas zauważa trzy poziomy naukowe myślenia, istotne przy ocenie wartości dowodu: a) teorie lub zasady, b) techniki lub procedury, c) właściwe zastosowanie danej procedury w konkretnym przypadku²⁸. W niniejszej monografii autor zajmuje się każdym z trzech wskazanych poziomów naukowego myślenia w zakresie badań poligraficznych, najwięcej zaś uwagi poświęca drugiemu i trzeciemu

²⁷ A. Kenny, *The Psychiatric Expert i Court*, „Psychological Medicine”, 14(2), 1984, s. 291–302.

²⁸ W.A. Thomas, *Some Observations by a Scientist*, 115 F.R.D., 1984; za: D.L. Faigman, E. Porter, M.J. Saks, *Check Your Crystal Ball at the Courthouse Door, Please. Exploring the Past, Understanding the Present, and Worrying About the Future of Scientific Evidence*, „Cardozo Law Review”, 15, 1994, s. 1825.

z nich, ze szczególnym uwzględnieniem zagadnień metod interpretacji danych, zgodności ocen dokonywanych przez poligrafów oraz subiektywnych czynników rzutujących na przebieg diagnozowania. Jeśli sąd ma uwzględnić dowód przy orzekaniu, to żadna z powyższych okoliczności nie może budzić wątpliwości.

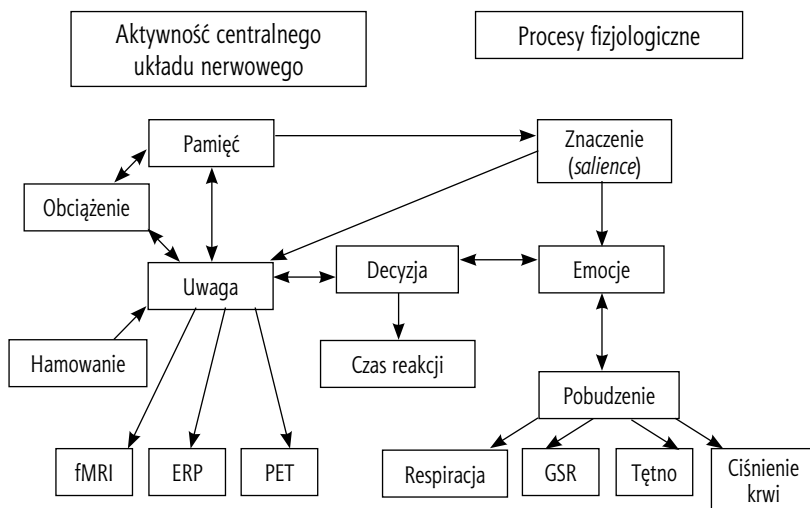
Pierwszy poziom ma charakter najbardziej abstrakcyjny, a przykładem może być zaprezentowana wyżej koncepcja zróżnicowanego znaczenia bodźców dla osoby poddawanej badaniu poligraficznemu; uznanie odruchu skórno-galwanicznego, określonych zmian cyklu oddechowego, ciśnienia krwi, tętna czy zmian temperatury twarzy za fizjologiczne korelaty emocji; ponadto: zmian przepływu krwi i utlenowania aktywnej okolicy mózgu, poznawczych potencjałów wywoływanych (fale P300) – jako niektórych spośród neuronalnych korelatów procesów uwagi (zob. ryc. 1)²⁹.

Drugi poziom naukowego myślenia obejmuje techniki badawcze wykorzystywane w badaniach poligraficznych oraz standardy praktyki – m.in. w zakresie organizacji przebiegu badania, prowadzenia wywiadu, doboru testów, formułowania pytań testowych, sposobu dokonywania pomiarów parametrów fizjologicznych, ich analizy i formułowania konkluzji. Zarówno koncepcje leżące u podstaw teoretycznych prowadzonych badań, jak i technika badawcza, która ma dostarczyć interesujących danych związanych z teorią, muszą mieć odpowiednie potwierdzenie naukowe. Można sobie wyobrazić, że

²⁹ Opisane symptomy wykorzystywane są w detekcji nieszczerości. Te związane z pobudzeniem fizjologicznym mierzone są m.in. za pomocą poligrafu i kamery termowizyjnej [szerzej na ten temat: M. Gołaszewski, P. Zajac, J. Widacki, *Thermal Vision as a Method of Detection of Deception. A Review of Experiences*, „European Polygraph”, 9, 1(31), 2015, s. 5–24; J. Widacki, Z. Mikrut, M. Widacki, J. Antos, *Próba wykorzystania zmian temperatury twarzy jako wskaźnika zmian emocjonalnych przy detekcji kłamstwa*, [w:] J. Widacki (red.), *Kierunki rozwoju instrumentalnej i nieinstrumentalnej detekcji kłamstwa. Problemy kryminalistyczne, etyczne i prawne*, Kraków 2018, s. 131–148]. Z kolei symptomy związane z aktywnością neuronalną bada się za pomocą elektroencefalografu, tomografu komputerowego czy funkcjonalnego rezonansu magnetycznego – na temat tego ostatniego zob. m.in. G.P. Lee, K.J. Meador, *Neural Substrates of Emotion as Revealed by Functional Magnetic Resonance imaging*, „Cognitive and Behavioral Neurology”, 17(1), 2004, s. 9–17; B.G. Bell, D. Grubin, *Functional Magnetic Resonance Imaging May Promote Theoretical Understanding of the Polygraph Test*, „The Journal of Forensic Psychiatry & Psychology”, 21(1), 2010, s. 52–65.

sama teoria charakteryzuje się solidnymi podstawami naukowymi, ale procedura badań jest nierzetelna i prowadzi do nietrafnych rezultatów. Tak samo jak możliwe jest zaprojektowanie niezawodnego testu, który jednak nie będzie osadzony na gruncie żadnej akceptowanej teorii.

Ryc. 1. Model procesu wprowadzania w błąd



Źródło: opracowanie własne podst. J.M.C. Vendemia, R.F. Buzan, S.L. Simon-Dack, *Reaction Time of Motor Responses in Two-stimulus Paradigms Involving Deception and Congruity with Varying Levels of Difficulty*, „Behavioral Neurology”, 16, 2005, s. 27.

Wreszcie trzeci poziom, brany pod uwagę przy ocenie dowodu z opinii biegłego, sprowadza się do konkretnych czynności wykonanych w związku z zastosowaniem wybranej metody badawczej – w odniesieniu do jednostkowego analizowanego przypadku w sprawie, do której biegły został powołany. Tu już w grę wchodzi ewentualne słabości w wyszkoleniu badającego czy wady sprzętu. To właśnie na tym poziomie kluczowe znaczenie ma dokonanie interpretacji danych przez poligrafera, wykorzystującego określone metody. Interpretacja obserwacji z badania poligraficznego odbywa się w kilku następujących po sobie etapach. Najpierw biegły analizuje zapisy reakcji fizjologicznych (wizualnie lub korzystając z narzędzi komputerowych), sprawdza ogólne tendencje i zmiany w poszczególnych miejscach testu dla każdego parametru rejestrowanego przez poligraf. Jeżeli zastosował system analizy numerycznej, interpretuje następnie uzyskane wyniki liczbowe. Wreszcie przekłada to na opisowy język i formułuje

konkluzje. Z kolei wyniki testów i wnioski wyrażone w opinii również podlegają dalszej interpretacji przez osoby zainteresowane ich treścią, uczestników danego postępowania. Zrozumienie danych przedstawionych przez biegłego jest konieczne dla przypisania im odpowiedniej wartości dowodowej.

Nawet jeśli jakaś metoda zasługuje na powszechną akceptację, to sposób, w jaki biegły się nią posłuży, może być wadliwy i obniżać jej trafność. Wnioski z interpretowanych pomiarów czy obserwacji nie będą wówczas miały solidnych podstaw ani w standardach, ani w aplikowanej teorii, a biegły – zamiast wspomagać proces decyzyjny – tylko go zagmatwa.

W części I monografii przeanalizowano ogólne czynniki, które rzutują na poprawność i zgodność ocen biegłych z zakresu badań poligraficznych. Opisano również spostrzeżenia dotyczące potencjalnych niedokładnych pomiarów. Poruszono wagę odpowiednich kwalifikacji biegłego, związanych z oczekiwanym standardem wykształcenia, doświadczeniem oraz osobistymi predyspozycjami. Omówiono także wybrane kwestie towarzyszące prowadzeniu badań, niewynikające z logicznego rozumowania i ścisłego postępowania zgodnie z przewidzianymi procedurami, a przez to wpływające ujemnie na niezawodność ocen poligraferów, skutkujące błędami poznawczymi. Snuto więc tutaj rozważania o stronie motywacyjnej i możliwym sugerowaniu się przez biegłych różnymi informacjami obok surowych wyników testów.

W części II przedstawiono różne metody interpretacji danych w badaniach poligraficznych – m.in. jakościowe i ilościowe, manualne i komputerowe. Przeprowadzono analizę ewentualnych różnic w dokładności identyfikacji i zgodności między oceniającymi, w zależności od zastosowanej metody. Zaprezentowano także badania empiryczne dotyczące marginesu subiektywizmu biegłych z zakresu badań poligraficznych – tak z przeglądu literatury, jak i własne. Jak pisze J. Moszczyński

„subiektywizm” biegłego oznacza ocenę odnoszącą się do pewnych aspektów badań kryminalistycznych, opartą z powodu braku możliwości zastosowania metod pomiarowo-statystycznych, na jego wiedzy, umiejętnościach i doświadczeniu, a także na wewnętrznym przekonaniu³⁰.

³⁰ J. Moszczyński, *Subiektywizm w badaniach kryminalistycznych*, Olsztyn 2011, s. 11.

Dalej uzupełnia, że

nawet w przypadkach stosowania wysoko zaawansowanych metod aparaturowych oraz statystycznych analiz wyników, z różnych powodów nie jest niekiedy możliwe uzyskanie jednoznacznych rezultatów i jest wymagana subiektywna interpretacja biegłego³¹.

Omawiając poszczególne metody analizy danych w badaniach poligraficznych, zasygnalizowano przynajmniej część obszarów, w których istnieje swoboda interpretacji.

Część III poświęcono potencjalnym sposobom ograniczania zjawiska subiektywizmu – m.in. przez ciągły proces walidacji i standaryzacji poszczególnych testów, którymi posługują się poligraferzy, ujednoczenie procedury przeprowadzania badań, doprecyzowanie kryteriów diagnostycznych, prowadzenie mniej lub bardziej sformalizowanej kontroli jakości.

Pojedyncze systemy oceniania danych testowych były przedmiotem rozmaitych publikacji³². W literaturze podejmowano ponadto próby porównania niektórych z nich, np. metody jakościowej i z wykorzystaniem algorytmów komputerowych³³. Przedstawiano zestawienie głównych systemów analizy numerycznej³⁴, a w podręcznikach uwzględniano nieco bardziej kompletny zarys różnych metod wykorzystywanych w praktyce zawodowej poligraferów³⁵. Powstawały

³¹ *Ibidem*.

³² Por. m.in.: R. Nelson, M. Handler, P. Shaw, M. Gougler, B. Blalock, C. Russell, B. Cushman, M. Oelrich, *Using the Empirical Scoring System*, „Polygraph”, 40(2), 2011, s. 67–78; M. Handler, R. Nelson, *Utah Approach to Comparison Question Polygraph Testing*, „Polygraph”, 38(1), 2009, s. 15–30.

³³ D.J. Krapohl, S.M. Senter, B.A. Stern, *An Exploration of Methods for the Analysis of Multiple-Issue Relevant/Irrelevant Screening Data*, „Polygraph”, 34(1), 2005, s. 47–62.

³⁴ Por. m.in.: R.S. Weaver, *The Numerical Evaluation of Polygraph Charts. Evolution and Comparison of Three Major Systems*, „Polygraph”, 9(2), 1980, s. 94–108; M. Gołaszewski, *Validated Techniques and Scoring Models for PDD Test Data Analysis – Conclusions from the 2011 APA Report*, „European Polygraph”, 6, 4(22), 2012, s. 227–240.

³⁵ Zob. m.in.: J. Widacki, *Wprowadzenie do problematyki badań poligraficznych*, Warszawa 1981; M. Kleiner, *Handbook of Polygraph Testing*, San Diego 2002; J.A. Matte, *Forensic Psychophysiology. Using The Polygraph. Scientific Truth Verification – Lie Detection*, New York 1996; J. Konieczny, *Badania poligraficzne. Podręcznik dla zawodowców*, Warszawa 2009; M. Gołaszewski (red.), *Współczesne standardy badań poligraficznych*, Warszawa

też opracowania dotyczące przesłanek diagnozowania (kryteriów diagnostycznych) i swoiste atlasy wzorcowych zapisów na poligramach, odzwierciedlających istotne zmiany reakcji organizmu³⁶. Nie było natomiast dotąd jednej pracy, w której współcześnie znane metody interpretacji danych w badaniach poligraficznych zostałyby w sposób kompleksowy usystematyzowane i poddane krytycznej analizie, biorąc pod uwagę zarówno ich historyczny rozwój, ale przede wszystkim aktualny stan nauki. W związku z powyższym autor niniejszej monografii podjął próbę tej systematyzacji, a przy tym rozważenia zakresu subiektywizmu w opiniowaniu biegłych omawianej specjalności. Pomocne w tym były artykuły (głównie anglojęzyczne) publikowane w periodykach naukowych i specjalistycznych (m.in. „Polygraph”³⁷, „European Polygraph”³⁸), książki, materiały konferencyjne, akty prawne, raporty – tak w formie drukowanej, jak i źródeł internetowych.

Na podstawie przeglądu piśmiennictwa oraz doświadczeń praktycznych autor niniejszej monografii postawił następujące pytania problemowe:

1. Czy poligrafer, oceniając zapisy na poligramach, sugeruje się (świadomie lub nieświadomie) innymi informacjami, dowodami zebranymi w toku prowadzonego postępowania?
2. Czy metody *numerycznej* interpretacji zapisów są lepsze od metody *jakościowej* (przy czym przez *lepsze* autor rozumie dające więcej poprawnych wyników i pozostawiające mniejszy margines subiektywizmu)?
3. Czy – w zależności od zastosowanej metody analizy numerycznej – uzyskuje się różne końcowe rezultaty testów (jeżeli chodzi

2013; D.C. Raskin, C.R. Honts, J.C. Kircher, *Credibility Assessment. Scientific Research and Applications*, San Diego 2014; D.J. Krapohl, P.K. Shaw, *Fundamentals of Polygraph Practice*, San Diego 2015.

³⁶ Por. J. Reid, F. Inbau, *Truth and Deception. The Polygraph (Lie-detector) Technique*, Baltimore 1977; J. Widacki, *Analiza przesłanek diagnozowania w badaniach poligraficznych*, Katowice 1982; T. Shurany, I. Ravid, *Evaluation of Polygraph Charts. Formats, Criteria and Scoring*, T. 1, Publications, Israel 2004; A.Y. Molchanov, N.A. Molchanova, *Atlas poligramm*, Jaroslav 2007; M. Gołaszewski (red.), *Współczesne standardy...*, *op. cit.*; J.A. Matte, *Forensic Psychophysiology...*, *op. cit.*, s. 371–397.

³⁷ Wydawnictwo American Polygraph Association. Od 2017 r. czasopismo zmieniło nazwę na „Polygraph & Forensic Credibility Assessment. A Journal of Science and Field Practice”.

³⁸ Półrocznik wydawany w Krakowie przez Oficynę Wydawniczą AFM.

o przyporządkowanie reakcji osoby badanej na pytania krytyczne jako typowych dla referencyjnych populacji osób odpowiadających szczerze albo nieszczerze)?

4. Czy metody numerycznej analizy danych testowych, zawierające węższą skalę ocen (*3-pozycyjną*), są bardziej obiektywne niż te o szerszych skalach (*7-pozycyjnych*)?
5. Która z ocen zapisów na poligramach jest dokładniejsza – przeprowadzana metodą *ślepej interpretacji* czy dokonywana przez poligrafera, który przeprowadził badanie?

W poszukiwaniu odpowiedzi na powyższe pytania autor niniejszej monografii zdecydował się na weryfikację niżej wymienionych hipotez:

1. Informacje na temat osoby poddawanej badaniu poligraficznemu, z którymi biegły zapoznaje się przed przeprowadzeniem analizy zarejestrowanych przez poligraf danych, mają wpływ na przebieg późniejszej interpretacji zapisów.
2. Metody numerycznej interpretacji zapisów na poligramach są dokładniejsze i pozostawiają mniejszy margines subiektywizmu niż metody jakościowe.
3. Pośród potwierdzonych naukowo metod analizy numerycznej nie ma istotnych różnic po ich zastosowaniu pod względem końcowych rezultatów uzyskiwanych w testach z pytaniami porównawczymi (typu CQT).
4. Metody numerycznej analizy zapisów na poligramach o węższej skali (*3-pozycyjnej*) są bardziej obiektywne niż te o skalach szerszych (*7-pozycyjnych*).
5. Oceniający metodą *ślepej interpretacji* – poprzez wyeliminowanie innych czynników niż dane zarejestrowane przez poligraf, świadomie lub nieświadomie branych pod uwagę przy wydawaniu opinii – z założenia powinni diagnozować z większym obiektywizmem, a przez to również z większą dokładnością niż eksperci, którzy osobiście przeprowadzili badania poligraficzne.

Hipotezy weryfikowano, opierając się na aktualnym stanie nauki, a w części dzięki własnemu udziałowi w badaniach empirycznych. Dane z tych badań zostały przetworzone z zastosowaniem analizy statystycznej w kierunku uzyskania wskaźników interesujących z punktu widzenia celów pracy. Przy okazji nasuwa się na myśl swoisty paradoks – polegający na tym, że dowody naukowe mają z założenia

do dodawania pewności w procesie podejmowania prawomocnych decyzji, a w istocie sama nauka, w tym rozmaite analizy statystyczne, wskazują na wiele niepewnych zmiennych, dlatego obserwuje się też tendencję do ostrożniejszego wyrażania stanowiska przez biegłych w kierunku wnioskowań probabilistycznych zamiast kategoriowych, opiniowania – a nie stwierdzania jakiegoś stanu rzeczy.

I. Czynniki rzutujące na poprawność i zgodność ocen poligraferów

1. Obszary subiektywizmu i potencjalne elementy pozadyskursywne przy ekspertyzach z zakresu badań poligraficznych

Poprawność ocen dokonywanych przez biegłego rzutuje na późniejsze postępowanie tych, którzy w mniejszym lub większym stopniu wspierając się tymi ocenami – podejmują decyzje dotyczące konkretnych ludzi, organizacji czy środowiska, w którym żyjemy. Ocena biegłego z zakresu badań poligraficznych może zmienić los osoby poddawanej badaniu – podejrzanego o popełnienie przestępstwa, kandydata do pracy lub służby, a nawet wiernego lub niewiernego partnera, który wybrał akurat taki sposób na (skądinąd wątpliwą) odbudowę zaufania do siebie w prywatnym związku.

Źródła nierzetelnych opinii¹ biegłych mogą być różnorakie. Należy ich upatrywać tak po stronie metodologii badań, jak i w czynniku ludzkim. W tym kontekście np. C. Adam dostrzega problemy związane m.in. ze statusem zawodowym, faktycznym doświadczeniem biegłych, ich stronniczą i niekiedy nieetyczną postawą². Truizmem jest stwierdzenie, że w środowisku biegłych rozmaitych specjalności – tak w Polsce, jak i na świecie – znajdują się osoby, które *de facto* są pseudo-biegłymi, nawet w sytuacji formalnie potwierdzonych kwalifikacji, np.

¹ „Rzetelność” – rozumiana tutaj dwojako: jako należyte wypełnienie obowiązków przez biegłego, ale także – a właściwie przede wszystkim – jako powtarzalność wyników badań (osłabiana przez potencjalny subiektywizm w ekspertyzie).

² C. Adam, *Forensic Evidence in Court. Evaluation and Scientific Opinion*, Oxford 2016, s. 11–16.

przez ustanowienie biegłym przy danym sądzie okręgowym, a w istocie nie posiadają dostatecznych kwalifikacji merytorycznych albo etycznych, aby taką funkcję sprawować. To jednak, jakie wiadomości specjalne powinien posiadać biegły z zakresu badań poligraficznych, omówiono w rozdziale 3.

W tym miejscu autor pragnie zasygnalizować zjawisko nieco inne, a mianowicie polegające na wykorzystywaniu przez biegłego uznanej pozycji w danym obszarze nauki – np. kryminalistyce czy nauce prawa karnego, a jednocześnie wypowiedaniu się na tematy, które w rzeczywistości nigdy nie znajdowały się wśród jego głównych zainteresowań badawczych i w których nie ma realnego doświadczenia ani specjalistycznego przeszkolenia. Mamy tu do czynienia z sytuacją, którą J. Bocheński nazwał *zabobonem intelektualisty*, który

polega na mniemaniu, że intelektualistcie przysługuje jako takiemu autorytet w dziedzinie etyki, polityki i poglądu na świat. [...] wierzenie w autorytet intelektualisty jest dosłownie na niczym nieoparte. Tak np. profesor wykładający historię nowożytną jest zapewne autorytetem (epistemicznym), gdy chodzi o rewolucję francuską, ale nie jest w dziedzinie użytkowania energii atomowej. Skoro więc taki profesor podpisuje razem z kolegami wyspecjalizowanymi w ceramice chińskiej, zoologii, względnie w rachunku prawdopodobieństwa deklaracje dotyczące tej energii, popełnia jaskrawe nadużycie autorytetu, tym gorsze, że wywołuje wrażenie, iż to sama „nauka” się wypowiada³.

Z kolei brak bezstronności biegłego może przejawiać się faworyzowaniem oskarżyciela i pomijaniem lub umniejszaniem znaczenia okoliczności oraz alternatywnych rozstrzygnięć, które mogłyby dostarczyć wątpliwości co do winy oskarżonego. Takie postawy kształtują się albo nieświadomie – w związku z subiektywizmem, wpływem informacji przekazywanych przez śledczych – albo rozmyślnie: w rezultacie postępowania sprzecznego z zasadami etyki zawodowej. Również ekspert zatrudniony przez stronę realizującą prawo do obrony może mieć skazę stronniczości. Ryzyko manipulacji informacjami przekazanymi na potrzeby badań, kwestia wynagrodzenia łącząca eksperta i stronę, a nawet zwyczajne przekupstwo, sprawiają, że zawsze należy liczyć się z możliwością nierzetelnego opiniowania, a przez to efekty pracy takiego eksperta wymagają ostrożnej i wnikliwej oceny.

³ J. Bocheński, *Sto zabobonów. Krótki filozoficzny słownik zabobonów*, Kraków 1994, s. 60–61.

Znane są nadto zjawiska wyolbrzymionego oszacowania znaczenia uzyskanego dowodu, zbytniego zaufania do nowych technik badawczych, błędów w badaniach identyfikacyjnych wynikających z nieprawidłowej interpretacji badanego śladu (C. Adam zilustrował to przykładem z daktyloskopii) lub jego kontaminacji (co można sobie wyobrazić, m.in. opierając się na historycznych przypadkach z badań laboratoryjnych śladów substancji wybuchowych)⁴.

Z kolei niedostateczna walidacja metody, a przez to niezweryfikowana wartość diagnostyczna, znajdują się w grupie metodologicznych uwarunkowań rzetelności opiniowania. Kontrowersyjne pod tym względem są np. kryminalistyczne badania śladów ugryzień⁵. U wielu przedstawicieli środowiska prawniczego badania poligraficzne także budzą takie kontrowersje pomimo współcześnie już solidnie dopracowanych podstaw teoretycznych i bardzo dobrze sprawdzonej empirycznie wartości diagnostycznej różnych technik stosowanych w tych badaniach. W zasadzie każda nowa metoda przechodzi przez kilka faz akceptacji. Ze strony prawników najpierw pojawia się nieufność, następnie ograniczona akceptacja, bezkrytyczna akceptacja i wreszcie – krytyczna akceptacja. Natomiast od strony kryminalistyki zaczyna się od nadmiernej akceptacji przez biegłych, zderzającej się z sądową nieufnością⁶.

Gorzej, jeśli sami biegli nie są ze sobą zgodni. Gdy spierają się ze sobą co do konkretnych ustaleń w określonej sprawie, dowód oparty na tych ustaleniach z pewnością traci procesowo na znaczeniu. Oszacowanie stopnia zgodności między biegłymi przy danej metodzie badawczej jest szczególnie ważne, ponieważ ma znaczenie dla oceny wartości dowodowej sporządzanych przez nich opinii. O ile w badaniach laboratoryjnych można tak dobrać sędziów kompetentnych, aby mieli oni np. zbliżone kwalifikacje, o tyle w analizach dotyczących rzeczywistych spraw bywa z tym różnie. Oczywiście – jak już powiedziano wcześniej – źródła rozbieżności mogą wynikać zarówno z niedoskonałości samej metody badawczej, popełnianych błędów (np. omyłek obliczeniowych, niewłaściwego zastosowania cech diagnostycznych czy reguł decyzyjnych – niezgodnie z wytycznymi danego systemu analizy), jak i czynników subiektywnych po stronie sporządzających ekspertyzy (wszędzie tam, gdzie brakuje precyzyjnych obiektywnych kryteriów oceny).

⁴ Zob. C. Adam, *Forensic Evidence in Court...*, *op. cit.*, s. 13–15.

⁵ *Ibidem*, s. 15.

⁶ J. Widacki (red.), *Kryminalistyka*, wyd. 4, Warszawa 2018, s. 461 i nast.

1.1. Ogólne czynniki ujemnie wpływające na miarodajność testów poligraficznych, w tym subiektywizm badającego

Tak jak przy każdym innym badaniu kryminalistycznym w badaniu poligraficznym również mogą zaistnieć okoliczności ujemnie wpływające na dokładność ustaleń dokonywanych przez biegłych. Należą do nich te po stronie biegłego – związane z istniejącymi obszarami subiektywizmu w metodologii badań oraz psychologicznymi aspektami opiniowania, a także po stronie osoby badanej – czynniki kontaminujące te procesy, które są istotne przy odbiorze oraz przetwarzaniu bodźców testowych, a uwarunkowane poznawczo, emocjonalnie lub aktualną kondycją fizyczną tej osoby⁷. W konsekwencji – niewielki odsetek osób badanych nie zalicza się do zbioru przypadków typowych i tworzy przypadki szczególne. Czynniki kontaminujące należą więc do źródeł fałszywych wyników testów i zwiększają ryzyko wystąpienia błędu – zwłaszcza fałszywej identyfikacji pozytywnej (obciążającej badanego). Częściej – nie tyle doprowadzą do błędu – co utrudnią uzyskanie jednoznacznego rozstrzygnięcia. Wśród nich można wymienić m.in.:

- a) gdy badany – mimo udzielania szczerych odpowiedzi – reaguje w sposób obserwowany zwykle u osób nieprawdomównych:
 - ukrywanie innych informacji niż te, których literalnie dotyczyły krytyczne pytania testowe, ale w dalszym ciągu w jakiś sposób związanych ze sprawą, będącą przedmiotem postępowania karnego;
 - obawa o fałszywie obciążające wyniki badania;
 - krótki okres inkubacji (czas na przetworzenie informacji o planowanym badaniu poligraficznym, także na poziomie nieświadomym);
- b) gdy badany – mimo faktycznego wprowadzania w błąd przy pytaniach krytycznych – reaguje w sposób typowy dla osób prawdomównych:
 - mechanizm obronny wtórnego wyparcia ze świadomości części wspomnień i odczuć zagrażających spójności osobowości;
 - skrajnie niepoważne, zobojętniałe podejście do testów;
 - zinterpretowanie pytań porównawczych (kontrolnych) przez badanego jako subiektywnie istotniejszych względem pytań relewantnych (krytycznych) – wbrew obiektywnym założeniom dotyczącym odmiennej wagi różnych typów odpowiednio sformułowanych pytań testowych dla osób szczerych i nieszczerych

⁷ T.T. Amsel, *Practicing Polygraph. Best Practice Guide*, North Charleston 2017, s. 45–96.

(taka sytuacja jest możliwa np., gdy pytanie porównawcze obejmuje swoim zakresem jakieś okoliczności, obarczone – zdaniem osoby badanej – większym ryzykiem skutecznego wykrycia i osądzenia, albo których ewentualne ujawnienie jest postrzegane jako bardziej dolegliwe niż w przypadku czynu, którego dotyczy pytanie krytyczne);

- c) poza ww. czynnikami – gdy w tzw. testach rozpoznania (ukrytych informacji, szczytowego napięcia) występują istotne zmiany reakcji fizjologicznych przy bodźcu, który w rzeczywistości nie jest krytyczny w sprawie:
- rozpoznanie jakiegoś szczegółu w teście nie na gruncie bezpośrednich, osobistych doświadczeń – lecz w sposób wtórny, za pośrednictwem osób trzecich (np. z mediów czy przez wcześniejsze ukierunkowanie badanego przez organ przesłuchujący);
 - łączenie z jakimś bodźcem (pytaniem o osobę, przedmiot, czynność, zaprezentowanym wizerunkiem itp.) szczególnych przeżyć emocjonalnych i nieprzedstawienie ich biegłemu przed przeprowadzeniem testu;
 - ukrywanie istotnych w sprawie informacji, z którymi w pewnym stopniu wiąże się dany bodziec (pytanie), choć wprost ich nie dotyczy.

Należy jednak wyraźnie podkreślić, że – co do zasady – prawdopodobieństwo przyczynienia się przedstawionych wyżej hipotetycznych okoliczności do jakiejś fałszywej konkluzji jest zwykle relatywnie niewielkie. W konkretnych analizowanych przypadkach mieści się w granicach statystycznego poziomu błędu, ustalonego dla wykorzystanych w danym badaniu typów testów oraz ich rezultatów (jeśli zostały wyrażone liczbowo).

Jak zauważa M. Kulicki – badanie poligraficzne może być w pewnym stopniu skażone subiektywizmem badającego: „Przekonania nie-doświadczonego eksperta co do występowania bądź niewystępowania związków badanego ze zdarzeniem mogą w istotny sposób wpływać na wnioski wyrażone w jego opinii, zwłaszcza w sytuacji niejednoznacznych reakcji badanego”⁸. Problem subiektywizmu dotyczy różnych nauk, choć zapewne w niejednakowym stopniu, z pewnością mniejszym w naukach ścisłych, jak fizyka czy biologia, a większym – w naukach społecznych, np. psychologii. Różna jest też waga konsekwencji subiektywizmu,

⁸ M. Kulicki, *Ekspertyza wariograficzna*, [w:] J. Wójcikiewicz, *Ekspertyza sądowa. Zagadnienia wybrane*, Warszawa 2007, s. 278.

które możemy sobie wyobrazić – inne będą w medycynie, naukach sądowych, a inne w socjologii, kulturoznawstwie. W naukach ścisłych przestrzeń subiektywizmu ogranicza się właściwie do niepewności pomiarów – stąd takie pojęcia jak *niepewność standardowa*, *błąd systematyczny*, *wartość prawdziwa* i wyznaczanie poziomów ufności w analizie statystycznej. W naukach społecznych zaś mamy do czynienia z wielością, niekiedy konkurencyjnych wobec siebie teorii – jak np. analityczna teoria prawa i ekonomiczna analiza prawa⁹. Teoria zaś rzutuje na formułowanie hipotez, dobór metod badawczych i interpretację wyników.

Badania w naukach sądowych polegają na obserwacjach, pomiarach, dokonywaniu ocen różnych zjawisk. Nie tworzy się na tym poziomie ogólnych teorii zjawisk, lecz uwaga badacza skierowana jest na jednostkowy przypadek. Przez fakt naukowy rozumie się „stwierdzenie konkretnego stanu rzeczy lub zdarzenia w określonym czasie i przestrzeni”¹⁰. Według austriackiego filozofa K. Poppera – autora postulatu falsyfikowalności jako kryterium oceny naukowości – fakty jednostkowe są stwierdzane przez zdania bazowe, które mają postać zdań egzystencjalnych, głoszących, że jakieś obserwowane zdarzenie zachodzi w konkretnym obszarze czasu i przestrzeni. Odróżnienia wymaga przy tym to, co dzieje się na poziomie percepcji badacza, czyli doznania subiektywne, uczucia czy przekonania, od obiektywnych relacji logicznych. Filozof zauważał, że te pierwsze nie dostarczają uzasadnienia dla zdań (choć stanowić mogą przedmiot badania psychologicznego). Zdania mogą być logicznie uzasadniane jedynie przez zdania i być intersubiektywnie sprawdzalne przez obserwację¹¹.

Sam pomiar jest zawsze obarczony niepewnością, która powinna być oszacowana. Z kolei tam, gdzie kończy się pomiar aparaturą, a zaczyna się analiza jego rezultatów, pojawia się czynnik ludzki, uwarunkowany wpływem osobistych doświadczeń i zdolnościami badacza oraz bodźcami zewnętrznymi, które na niego oddziałują. Problem jest dostrzegany przez naukowców i poruszany w literaturze¹². Co więcej, w nie-

⁹ Zob. G. Góralczyk, *Ekonomiczna analiza prawa jako wyraz zmiany tradycyjnego paradygmatu*, [w:] A. Samonek, *Teoria prawa. Między nowoczesnością a ponowoczesnością*, Kraków 2013, s. 161–172.

¹⁰ *Fakt naukowy* [hasło], sjp.pwn.pl/sjp/fakt-naukowy;2458324.html [dostęp: 13.05.2020].

¹¹ Zob. K.R. Popper, *The Logic of Scientific Discovery*, London 2002, s. 57–94.

¹² Por. B. Robertson, G.A. Vignaux, Ch. Berger, *Interpreting Evidence. Evaluating Forensic Sciences in the Courtroom*, 2nd edition, Oxford 2016, s. 102–104; C. Adam, *Forensic Evidence in Court...*, *op. cit.*, s. 117–124.

których dyscyplinach w ogóle nie spotykamy pomiarów ilościowych, a jeśli nawet – to w ograniczonym zakresie, zaś podstawą opiniowania są metody jakościowe, np. w psychiatrii sądowej czy badaniach pisma ręcznego. W badaniach poligraficznych współwystępują zarówno metody jakościowe, jak i ilościowe, o czym szerzej w rozdziale drugim. Posługując się metodami jakościowymi, poligrafer dokonuje interpretacji poczynionych obserwacji, opierając się na zbiorach cech najczęściej występujących, charakteryzujących dany przypadek i pozwalających na ustalenie pewnego wzorca, któremu można przyporządkować przedmiot badań lub nie. Jeżeli biegły wskazuje w opinii np., że osoba badana reagowała przy udzielaniu odpowiedzi na pytania krytyczne w sposób, jaki zwykle obserwuje się osób odpowiadających szczerze – to czyni nic innego jak porównanie zapisów na poligramach do tzw. przypadku typowego. Nawet jeśli dokonał precyzyjnych pomiarów przy użyciu poligrafu i zinterpretował uzyskane dane zgodnie z regułami sztuki, to w dalszym ciągu istnieje jakaś liczba przypadków nietypowych i możliwość, że to hipoteza zerowa jest prawdziwa – tym bardziej że nie istnieje coś takiego jak jeden wzór kłamstwa dla całej ludzkości. Podobnie jest zresztą np. w medycynie, gdzie rozmaite choroby mogą objawiać się w różny sposób pomimo wyodrębnienia najczęstszych symptomów, a nawet przebiegać bezobjawowo. Wiarygodność diagnozy ma szersze uwarunkowania – zależy od ogólnego rozwoju nauki w danej dyscyplinie i przyjętej dla niej metodologii badań.

Jeżeli w badaniu poligraficznym poza pomiarem danych konieczna jest ich interpretacja, rzeczywiście pojawia się przestrzeń do zaistnienia subiektywizmu poligrafera – przy czym przez element subiektywizmu w ekspertyzie autor niniejszej monografii rozumie jako margines swobody biegłego przy owej interpretacji, niepodlegający obiektywnym, wyraźnie określonym kryteriom. Należy założyć, że margines ten będzie inny dla poszczególnych systemów analizy, biorąc pod uwagę różnice metodologiczne (przyjęte cechy diagnostyczne zapisów, sposób ich ewaluacji, reguły decyzyjne). Powinien być jak najlepiej poznany, by trafnie oceniać efekty pracy poligrafera.

Warto dodać, że subiektywizm może przejawiać się także w związku z interpretacją wyników badań (zwłaszcza w formie probabilistycznej, liczbowej) przez strony postępowania. Mamy wówczas do czynienia z tzw. sofizmatami – argumentowaniem pozornie poprawnym, ale w istocie zawierającym ukryte błędy logiczne. Dzieje się tak albo

z powodu braku poprawnego zrozumienia danych statystycznych, albo jest to działanie rozmyślne, ze względu na interes strony.

Przykładowo w badaniach genetycznych z wykorzystaniem DNA biegły podaje zwykle tzw. iloraz wiarygodności¹⁾ (*szansa ustalenia danej zgodności, jeżeli to nie podejrzany, a przypadkowa osoba pozostawiła ślad na miejscu zdarzenia wynosi X*). Szansy identyfikacji jakiejś cechy nie wolno jednak utożsamiać wprost z szansą winy. Jeżeli profil DNA podejrzanego zgadza się ze śladem biologicznym z wymazu z pochwy ofiary gwałtu, to nie jest to jeszcze równoznaczne z tym, że podejrzany jest gwałtciwicielem. Szansa winy będzie tutaj zależała od uwzględnienia innych dowodów w sprawie²⁾. Konieczne jest nadto łączne rozważenie prawdopodobieństwa ustalenia dowodu w sytuacji, gdy hipoteza oskarżyciela byłaby prawdziwa i zarazem prawdopodobieństwa hipotezy obrony. Tymczasem jeżeli przyjmiemy, że X z powyższego przykładu wynosi 0,000001 – to prokurator może chcieć przekonywać, że prawdopodobieństwo winy podejrzanego (czyli że to on naniósł ślad) wynosi 99,9999% i graniczy z pewnością. Z drugiej strony – obrońca stanie na stanowisku, że jego klient jest jednym z 800 tys. ludzi na świecie (przy ogólnej liczbie ludności ok. 8 mld), którzy mogą posiadać zidentyfikowane cechy DNA i nawet, jeżeli z okoliczności sprawy wynika, że sprawcą był mieszkaniec ok. dwumilionowej Warszawy – to w dalszym ciągu równie prawdopodobne jest, że co najmniej jeszcze jedna niespokrewniona z podejrzanym osoba pozostawiła badany ślad, a zatem prawdopodobieństwo winy jego klienta wynosi tylko 50%.

$$LR = \frac{p(E|H_p)}{p(E|H_d)}$$

¹⁾ Wzór na iloraz wiarygodności (LR – ang. *likelihood ratio*), gdzie porównuje się prawdopodobieństwa warunkowe. Pierwsze prawdopodobieństwo (w liczniku) – wskazujące, że dowód (E) pochodzi od tej samej osoby, tego samego obiektu itd., wspiera tezę o winie podejrzanego (Hp – hipoteza prokuratora). Natomiast w mianowniku mamy prawdopodobieństwo, że dowód (E) świadczy o tym, że porównywane cechy dotyczą różnych osób, przedmiotów czy też wynik jakiegoś testu jest dziełem przypadku, wspiera tezę o niewinności podejrzanego (Hd – hipoteza obrońcy).

²⁾ Mamy tu typowy przykład dowodu pośredniego (*circumstantial evidence*) – por. J. Widacki, *Results of Polygraph Examinations. Direct or Circumstantial Evidence?*, „European Polygraph”, 8, 2(28), 2014, s. 61–67.

Sofizmaty (w powyższym przykładzie – prokuratora i adwokata) mogą wypaczać właściwy sens wyników badań. Podobne błędy może popełnić także biegły. Rozważmy hipotetyczne opisy wartości diagnostycznej testów wykonanych przez poligrafera. Przykładowo może on zawrzeć w opinii, że zaobserwowany sposób reagowania badanego na pytania krytyczne (uzyskany wynik testu) jest 20-krotnie bardziej prawdopodobny, jeśli badany wprowadzał błąd lub zataił istotne okoliczności, niż gdyby takie zmiany reakcji były dziełem przypadku. Co innego natomiast, gdy na podstawie tych samych danych biegły napisałby, że w świetle zarejestrowanych reakcji jest 20 razy bardziej prawdopodobne, że badany popełnił zarzucany mu czyn. Zdania brzmią podobnie, ale tylko pierwszy z opisów należy uznać za w pełni poprawny¹³.

Biegły udziela odpowiedzi na zadane pytania problemowe. Jego proces myślowy może przy tym uwzględniać inne kwestie znane mu w danej sprawie i wstępne wyobrażenia na temat przedmiotu badań¹⁴. Elementy pozadyskursywne występujące przy wyjaśnianiu faktów jednostkowych zależą od indywidualnych predyspozycji poligraferów, ich intuicji¹⁵, zdolności kojarzenia faktów, wcześniejszych doświadczeń (por. podrozdział 1.4). Słusznie przy tym zauważył J. Życiński: „Przecenianie roli czynników pozadyskursywnych w poznaniu może prowadzić łatwo do subiektywizmu”¹⁶. Subiektywizm może z kolei skutkować tym, że w pewnych warunkach te same dane z badań będą interpretowane przez biegłych niejednako, obniżając rzetelność metody badawczej.

¹³ Więcej możliwych błędów w interpretacji statystycznych wyników badań zob. J.L. Gastwirth (ed.), *Statistics for Social Science and Public Policy*, New York 2000, s. 4–9.

¹⁴ Por. R.J. Joling, *Legal Commentaries on Forensic Scientist in Court*, „Journal of Forensic Sciences”, 8(3), 1963, s. 341.

¹⁵ Termin „intuicja” użyty na określenie podświadomego procesu rozumowania „doświadczeniowego”, skojarzeniowego. Intuicja sprzyja kreatywności badającego, ale nie powinna oczywiście zastępować rozumowania racjonalnego, opartego na regułach określających metodykę badań.

¹⁶ Zob. J. Życiński, *Problem wyjaśniania w naukach nomotetycznych a poznanie faktu jednostkowego*, [w:] J. Widacki (red.), *Z zagadnień teorii opinii biegłego*, Materiały IV Sympozjum Metodologii Kryminalistyki i Nauk Pokrewnych, Chęciny 24–25 VI 1983, Katowice 1983, s. 60 i nast.

Zgodnie ze stanowiskiem przedstawicieli konstrukcjonizmu społecznego świat zewnętrzny nie jest odwzorowywany, lecz *konstruowany* w umyśle z informacji dochodzących z receptora oraz oczekiwań, uwarunkowanych przekonaniem, celami, dotychczasowymi doświadczeniami, kulturą¹⁾. W związku z tym człowiek rozpatruje rzeczywistość w określonym kontekście i nadaje jej subiektywne znaczenie. Czysto obiektywny stan rzeczy nie może więc być przez nikogo poznany.

Uważa się, że struktury poznawcze (reprezentacje umysłowe, składniki obrazu świata w umyśle) tworzy się w taki sposób, aby dostarczały maksymalnie użytecznych informacji minimalnym kosztem (zasada oszczędności poznawczej)²⁾. Reprezentacje umysłowe uzależnione są od dostępnych człowiekowi ram pamięciowych. W magazynowym modelu pamięć dzieli się na ultrakrótką (zmysłową), krótkotrwałą (operacyjną) i trwałą. W ramach tej ostatniej, ze względu na format przechowywanych informacji, wyróżniamy pamięć epizodyczną, semantyczną i proceduralną. Pamięć epizodyczna dotyczy tych reprezentacji umysłowych, jakimi są wspomnienia, czyli zdarzenia, z którymi mieliśmy bezpośredni kontakt w przeszłości i mającymi wyraźne odniesienie do konkretnego miejsca lub czasu. Pamięć semantyczna (typu *wiem*) związana jest z przyswajaniem, przechowywaniem i wykorzystaniem wiedzy o charakterze pojęć, przekonań, sądów. Z kolei pamięć proceduralna (*umiem*) odpowiedzialna jest za przyswajanie umiejętności, nawyków i procedur³⁾. Realizacja zasady oszczędności poznawczej odbywa się m.in. przez schematy poznawcze, abstrakcyjne uogólnienia wynikające z wcześniejszych doświadczeń dotyczących jakichś zdarzeń czy podmiotów, nadające sens sytuacjom typowym.

Jak zauważa B. Wojciszke: „Duża część zachowania społecznego wyznaczana jest naszymi bieżącymi sądami na temat aktualnej sytuacji i jej uczestników”⁴⁾. Sądy te zależą od struktur wiedzy, a także od bieżącego kontekstu, w którym są wydawane. Inne obiekty kontekstu mogą wpływać dodatnio na ten oceniany, albo ujemnie. Przesunięcie oceny w kierunku informacji kontekstowej nazywamy efektem asymilacji. Przeciętny słuchacz kursu oficerskiego w ośrodku szkolenia służb specjalnych oceniany razem z dwoma prymusami podczas wspólnej prezentacji efektów pracy analitycznej w grupie zyska na własnej ocenie. Gdy jednak po kilku miesiącach nadejdzie czas egzaminów ustnych przeprowadzanych pojedynczo, nastąpi efekt kontrastu – odsunięcie oceny od informacji kontekstowej i ocena przeciętnego podoficera spadnie na tle lepszych od niego słuchaczy.

Niewielka różnica między kontekstem a obiektem ocenianym prowadzi zazwyczaj do asymilacji, zaś duża – do kontrastu. Gdy na badanie poligraficzne w służbie specjalnej stawi się pragnący szybszego rozwoju zawodowego kandydat wywodzący się z np. Policji, bez jakichś szczególnych osiągnięć czy umiejętności – wyobrażenie na jego temat na podstawie lektury akt opracowania kadrowego nie będzie odbiegało od ogólnego obrazu kandydatów tego typu, może nieco stereotypowe, ale powstałego na podstawie wcześniejszych doświadczenia z badań. Z kolei, jeśli pojawiłby się kandydat – również z Policji – który już na starcie zademonstruje historię swoich sukcesów, wysokie kwalifikacje merytoryczne i – przynajmniej deklaratywnie – moralne, może liczyć na wyższą ocenę na zasadzie kontrastu z poprzednikami.

Trzeba też pamiętać, że identyczny obiekt kontekstowy może oddziaływać na oceny albo przez kontrast, albo przez asymilację w zależności od tego, jakiego rodzaju kategoryzacje dominują w danej sytuacji⁵⁾. Jeżeli na ławie oskarżonych zasiadają wspólnie mężczyźni wymuszający brutalnie w warunkach recydywy spłatę lichwiarskich zobowiązań oraz jego *partner biznesowy* prowadzący sprawy księgowe, to z punktu widzenia zewnętrznego obserwatora relacji medialnej z procesu oceny *księgowego* najpewniej zasymilują negatywny obraz *agresywnego recydywisty* przez potraktowanie obu osób w kategorii *brutalni lichwiarze*. Natomiast jeśli mężczyźni trafią do zakładu karnego, *księgowy – biały kołnierzyk* będzie najprawdopodobniej oceniany przez pracowników Służby Więziennej korzystniej, łagodniej na tle *bandyty – recydywisty*.

¹⁾ Por. K. Najder, *Schematy poznawcze*, [w:] M. Materska, T. Tyszka (red.), *Psychologia i poznanie*, Warszawa 1992, s. 39.

²⁾ Zob. *ibidem*, s. 41.

³⁾ A. Stępnik, *Reprezentacje umysłowe, rodzaje pamięci a wiedza*, „Przegląd Filozoficzno-Literacki”, 2(39), 2014, s. 167–187.

⁴⁾ B. Wojciszke, *Psychologia społeczna*, Warszawa 2011, s. 84.

⁵⁾ Zob. *ibidem*, s. 85–86.

1.2. Uproszczone reguły myślenia, błędy postrzegania

Wobec obfitości dopływających z otoczenia informacji do przetworzenia umysł ludzki stosuje uproszczone reguły myślenia (heurystyki), które pozwalają na formułowanie ocen, podejmowanie decyzji w sposób sprawny i niewymagający nadmiernego wysiłku. Człowiek kieruje się w swoich obserwacjach w pierwszej kolejności tym, co jest najłatwiej dostępne i dobrze zapamiętane. Ocenę częstości lub prawdopodobieństwa

zdarzeń na podstawie łatwości, z jaką przychodzą nam na myśl ich przykłady, nazywamy *heurystyką dostępności*¹⁷. Większość ludzi jest przekonana co do tego, że bardzo łatwo jest „pokonać” poligraf. Historie tego typu, zwłaszcza z aferą szpiegowską w tle, są szeroko nagłaśniane i budzą sensację¹⁸. Tymczasem są to pojedyncze przypadki, a potwierdzona w badaniach naukowych ogólna dokładność wyników badań poligraficznych ze spraw rzeczywistych oraz aranżowanych laboratoryjnie jest wysoka i to pomimo tego, że z różnych analiz wynika, iż nawet co druga osoba badana stosuje jakąś formę środków zakłócających. Choć próby te są trudne do wykrycia, okazuje się, że przynajmniej spontaniczne przeciwdziałanie uzyskaniu miarodajnych wyników testów jest zupełnie nieskuteczne¹⁹.

Spśród innych znanych przykładów uproszczonego myślenia warto wymienić *heurystykę reprezentatywności*, czyli ocenę przynależności obiektu do danej kategorii na podstawie podobieństwa do przypadków typowych. Badany za pomocą poligrafu, wykonujący serię powolnych i głębokich oddechów, zostanie szybko zakwalifikowany przez wielu poligraferów jako osoba podejmująca próbę zniekształcenia prawidłowych pomiarów i niewykluczone, że nawet w tym wytrenowana (równoległe może tu występować także wspomniana wcześniej heurystyka dostępności w postaci odtwarzanych w pamięci studiów przypadków ze szkoleń specjalistycznych). Diagnoza ta pewnie nieraz okazałaby się trafna, ale równie dobrze mógłby to być ktoś uprawiający regularnie jogę, nurek czy muzyk grający na instrumentach dętych, przyzwyczajony do specyficznego sposobu oddychania. Przyczyn rozpatrywanych zdarzeń upatruje się w sytuacjach podobnych. Na przykład utrata niejawnego dokumentu przez funkcjonariusza wzbudza hipotezę celowego przekazania obcym służbom albo dziennikarzom, a więc dla poważnego

¹⁷ B. Wojciszke, *Psychologia społeczna, op. cit.*, s. 90.

¹⁸ Por. Senate Select Committee on Intelligence, *An Assessment of the Aldrich H. Ames Espionage Case and Its Implications for U.S. Intelligence*, 1.11.1994, fas.org/irp/congress/1994_rpt/ssci_ames.htm [dostęp: 4.06.2020]; *Ana Montes: Cuban Spy*, www.fbi.gov/history/famous-cases/ana-montes-cuba-spy [dostęp: 4.06.2020].

¹⁹ Zob. m.in.: C.R. Honts, D.C. Raskin, J.C. Kircher, R.L. Hodes, *Effects of Spontaneous Countermeasures on the Physiological Detection of Deception*, „Journal of Police Science and Administration”, 16(2), 1988, s. 91–94; K.D. Otter-Henderson, C.R. Honts, *Spontaneous Countermeasures during Polygraph Examinations. An Apparent Exercise in Futility*, „Modern Psychological Studies”, 8(1), 2001, s. 11–15.

skutku poszukuje się poważnej przyczyny, gdy faktycznie nastąpiło np. przypadkowe przyklejenie się kartki papieru do pliku materiałów roboczych wkładanych do niszczarki. Tego typu myślenie jest jednym z zagrożeń przy interpretacji danych testowych – zwłaszcza metodą jakościową. Może skłaniać w kierunku poszukiwania dowodów heurystycznie przyjętej hipotezy, lekceważąc inne rozwiązania.

Wyjaśnieniem uproszczonych, zdroworozsądkowych sposobów postrzegania zachowania innych ludzi i przyczyn zdarzeń są teorie atrybucji, skonstruowane w latach 50. przez F. Heidera. Atrybucje opierają się na założeniu, że bodźce nie wywołują bezpośrednio reakcji, lecz między nimi pośredniczą procesy poznawcze. Ludzie starają się rozumieć przyczyny zjawisk w sposób realistyczny, podobny do naukowego, czyli przez tworzenie hipotez, sprawdzanie ich na podstawie własnych obserwacji, a następnie ewentualne korygowanie. Co więcej – jest to użyteczne, ponieważ pozwala zrozumieć świat, dokonywać przewidywań i kontrolować przebieg zdarzeń¹⁾.

Heider zakładał, że zachowanie jest łączną konsekwencją czynników zewnętrznych i wewnętrznych, które dodają się do siebie. Atrybucja wewnętrzna zakłada, że przyczyna czyjegoś zachowania leży po stronie tej osoby, natomiast zewnętrzna za decydujące przyjmuje oddziaływanie otoczenia. Jeżeli kandydat do służby spóźni się na umówione badanie poligraficzne, pierwszy rodzaj atrybucji będzie prowadził do wniosków, że jest to osoba spóźnialska i źle zorganizowana. Niewykluczone, że w konsekwencji badający subiektywnie przyjmie przez to mniej pomocną postawę wobec badanego.

Drugi zaś rodzaj atrybucji wymagałby wzięcia pod uwagę innych możliwych okoliczności, jak wypadek po drodze i zakorkowana trasa dojazdu na miejsce badania. „Rodzaj dokonanej atrybucji wpływa na oczekiwania dotyczące przyszłości, a także na przeżywane emocje i poziom wykonania”²⁾. W ślad za B. Wojciszke warto zauważyć, że w podobnym kierunku jak teoria atrybucji zmierza teoria wniosków korespondentnych Jonesa i Davisa, w której ramach wyjaśnianie ludzkiego zachowania jest rozumiane jako poszukiwanie odpowiedniości między działaniem, intencją i cechami osobowości. Wniosek korespondentny oznacza, że za obserwowanym zachowaniem kryje się jakaś odpowiadająca mu trwała cecha aktora. „Dopóki człowiek robi to, co »się robi«, co jest społecznie oczekiwane i aprobowane, dopóty jego zachowanie nie daje podstaw do wnioskowania o cechach tej osoby. Silną podstawą do wnioskowania

o cechach osobowości jest natomiast zachowanie sprzeczne z oczekiwaniami³⁾. Przy czym – formułowanie wniosków korespondentnych odbywa się w sposób bardziej zdecydowany niż ma to faktyczne uzasadnienie, nie doceniając roli ograniczeń sytuacyjnych.

Zwykle najpierw automatycznie, w znacznej mierze nieświadomie, upatrujemy przyczyn ludzkiego zachowania w czynnikach wewnętrznych – głównie subiektywnie postrzeganych cechach osobowości, a następnie w sytuacji zewnętrznej, złożonej – co do której trudniej uzyskać pełny zakres informacji. W ten sposób popełniamy tzw. podstawowy błąd atrybucji lub inaczej – błąd korespondencji, którego ryzyko rośnie wraz obciążeniem zasobów poznawczych innymi zadaniami, a słabnie, jeśli obserwator koncentruje się maksymalnie tylko na osobie ocenianej i wtedy, gdy uświadamia sobie konieczność uzasadnienia swoich sądów.

¹⁾ Zob. F. Försterling, *Atrybucje. Podstawowe teorie, badania i zastosowanie*, tłum. J. Radzicki, Gdańsk 2005, s. 21–24.

²⁾ B. Wojciszke, *Psychologia społeczna, op. cit.*, s. 134–135.

³⁾ *Ibidem*, s. 135.

Odmianą podstawowego błędu atrybucji jest tzw. efekt *halo*, czyli tendencja do dokonywania ocen na temat cech osobowościowych pod wpływem pierwszego wrażenia – tak pozytywnych (efekt *aureoli*), jak i negatywnych (efekt *diabla*). Przypisanie komuś jednej właściwości wpływa na kolejne, które są zgodne z zabarwieniem emocjonalnym tej pierwszej i nie wynikają z obiektywnej obserwacji. W ten sposób mogą zadziałać m.in. wstępne odczucia na temat osoby poddawanej badaniu poligraficznemu, wyrobione np. na podstawie charakterystyki przedstawionej na temat informatora przez oficera prowadzącego czy w związku z lekturą akt sprawy przed badaniem podejrzanego. Tak samo nie bez znaczenia jest ubiór kandydata do służby czy styl jego przywitania po przybyciu na badanie przedzatrudnieniowe. Na podobnej zasadzie działa efekt pierwszeństwa – czyli silniejsze oddziaływanie pierwszej informacji niż późniejszych.

Ważny przy formułowaniu opinii z badań poligraficznych okazuje się także *błąd tendencji centralnej*, polegający na ignorowaniu różnic indywidualnych, a jeżeli chodzi o kontekst niniejszych rozważań – przede wszystkim przejawiający się w unikaniu skrajnych ocen i kategorycznych konkluzji poprzez wyrażenia typu *ma związek ze sprawą, nie można wykluczyć, że...* itp. Ignorowanie różnic indywidualnych

stwarza ponadto tzw. *ryzyko Brokawa*²⁰, czyli uznania osoby prawdomównej za kłamcę z powodu zachowania postrzeganego jako typowe dla osoby winnej. P. Ekman przypomina, że są tacy ludzie, którzy zawsze odpowiadają pokrętnie, mają taki styl bycia:

Jakiegokolwiek zachowanie, które jest użyteczną wskazówką kłamstwa, u niektórych osób będzie ich typowym zachowaniem. [...] Wykrywający kłamstwa są narażeni na ryzyko Brokawa, gdy ktoś, kogo podejrzewają, jest dla nich osobą obcą i nie wiedzą, jakie jest jej typowe zachowanie²¹.

Podobne skutki jak ryzyko Brokawa przynosi tzw. *błąd Otella*, polegający na uznaniu osoby niewinnej za winną z powodu symptomów napięcia emocjonalnego związanego z obawą o konsekwencje nietrafnego osądu. Ekman krytykuje przez to badania poligraficzne, spływając ich istotę do wykrywania pobudzenia emocjonalnego. Poligraf nie jest jednak *miernikiem* stresu. Specyficzne zmiany reakcji fizjologicznych, brane pod uwagę przez biegłego z zakresu badań poligraficznych, nie są wyłącznie przejawem emocji. Co więcej – subiektywnie uważam, że oceniający tylko na podstawie oznak behawioralnych, np. analizy tzw. *przecieków*, mikroekspresji – są narażeni na błąd Otella o wiele bardziej niż poligraferzy.

Generalnie dobrze jest, by możliwie najwięcej opinii czy podejmowanych decyzji zawierało uzasadnienie. Jest wówczas szansa na ograniczenie decyzji pochopnych, arbitralnych i nieracjonalnych. Od biegłych generalnie wymaga się, by ich opinie miały uzasadnienie, więc i ryzyko błędów atrybucyjnych wydaje się nieco mniejsze. Poza tym w przypadku badań poligraficznych przedmiotem zainteresowania jest to, czy dane zachowanie nastąpiło, czy nie, a mniej – z jakiego powodu. To jest już istotniejsze, gdy trzeba wydać określoną decyzję kadrową czy wyrok sądowy, gdzie strona motywacyjna i inne okoliczności towarzyszące jakiemuś zdarzeniu wpływają na ocenę stopnia ewentualnej winy.

²⁰ Tom Brokaw – znany dziennikarz telewizyjny w USA, prowadzący programy typu talk-show, m.in. „Today Show” w stacji NBC.

²¹ P. Ekman, *Kłamstwo i jego wykrywanie w biznesie, polityce i małżeństwie*, tłum. S.E. Draheim, M. Kowalczyk, Warszawa 2015, s. 83–84.

Reakcje na zachowania, postawy czy np. wyniki uzyskiwane w testach przez osobę badaną zależą m.in. od tego, jak postrzegamy ich przyczyny. W związku z tym będziemy albo chwalić, albo krytykować. Wykażemy gniew albo empatię. Potraktujemy kogoś z przychylnością lub przeciwnie – zechcemy zaszkodzić. Oceniający odmiennie podchodzą do wyjaśnienia zachowań własnych i cudzych – w sposób egotystyczny. U siebie upatrują raczej wpływów czynników zewnętrznych, w szczególności przy zdarzeniach negatywnych, porażkach. Odwrotnie przy sukcesach. Postępowanie innych natomiast wyjaśnia się z reguły ich osobistymi właściwościami. Jest to więc podejście dalece subiektywne. Do tego dochodzi jeszcze błąd fałszywej powszechności, przez który uznajemy, że nasze poglądy są podzielane przez większość ludzi. Dzieje się tak dlatego, że częściej rozważamy własne stanowisko i nie zajmujemy się specjalnie tym, co sądzą inni, a ponadto sami otaczamy się takimi ludźmi, z którymi „nadajemy na podobnych falach”.

Uleganie różnym heurystykom narasta w szczególności przy presji czasu. Gdy nie jest to naprawdę konieczne, nie należy przystępować do badań poligraficznych w pośpiechu. Inaczej biegły oprze się z konieczności na informacjach tylko najłatwiej dostępnych, pobieżnych, uruchomi schematy, w większym stopniu może ulegać osobistym wrażeniom.

Na treść sądów wpływa też aktualny stan emocjonalny ocenianego, jego nastrój. Dlatego m.in. poligraferom zaleca się na szkoleniach, aby nie decydowali się na przeprowadzenie badania w sytuacji, gdy coś zaburza ich aktualną równowagę emocjonalną, np. przez silne wzburzenie czy głęboki smutek w wyniku jakiegoś zdarzenia w życiu prywatnym. Jeżeli nastrój jest dobry, sprzyja podnoszeniu ocen innych ludzi i nasila skłonność do pomagania innym²². Biegły z zakresu badań poligraficznych w pogodnym nastroju, w razie występujących w toku badania wątpliwości dotyczących wiarygodności badanego, może przejawiać większe zaangażowanie, aby pomóc mu te wątpliwości usunąć. Jeżeli po badaniu przedstawi wynik liczbowy przeprowadzonego testu – to nie pozostawi wielkiego pola manewru, lecz jeśli dołoży do tego wynik opisowy, jakościowo odnoszący się do przedmiotu badań – ostateczny wydzźwięk takich wniosków może różnić się w zależności od tego, jakie jest subiektywne postrzeżenie osoby badanej i co na nie wpływa – np. nastrój podczas pisania opinii.

²² B. Wojciszke, *Psychologia społeczna*, op. cit., s. 86.

Przy wynikach wskazujących na prawdomówność nie będzie to miało większego znaczenia, natomiast przy rezultatach występujących zwykle u osób odpowiadających nieszczerze na pytania krytyczne – dodatkowa interpretacja może takie wyniki wzmocnić albo złagodzić.

Zły nastrój nie przynosi tak jednoznacznych skutków jak dobry, ponieważ z jednej strony – obniża oceny, ale z drugiej – człowiek może dążyć do poprawy własnego samopoczucia przez działania pozytywne i pomoc innym. Co ciekawe, to w nastroju negatywnym ludzie stają się bardziej uważni i lepiej rozróżniają argumenty słabe od silnych, dając się przekonać bardziej tym ostatnim²³.

Znanym w przeszłości z praktyki niektórych poligraferów było (i być może nadal bywa) stwierdzenie brzmiące mniej więcej w ten sposób, że reakcje osoby badanej przy pytaniach krytycznych mogły wiązać się z przedstawionymi przez nią w toku badania oświadczeniami. Pomijam w tym miejscu głębsze rozważanie sensu tak sformułowanego zdania, ponieważ jest ono metodologicznie niepoprawne, polega na spekulowaniu i wyraża tylko jedną z niezweryfikowanych hipotez. Chodzi mi o pokazanie, że taki przykładowy zabieg wyraża raczej pomocne nastawienie badającego do osoby badanej przez próbę usprawiedliwienia znaczących zmian w reakcjach. W sposób poprawny natomiast może się to przejawiać m.in. przez uwypuklenie innych hipotez niż ta przyjęta na podstawie wyniku testu, przedstawienia potencjalnych czynników kontaminujących itp.

Sklonność do pomocy rośnie, gdy czyjeś położenie zdaje się wynikać z okoliczności pozostających poza kontrolą tej osoby, a zatem nie powinna ona ponosić za to odpowiedzialności. Również przy usprawiedliwianiu własnych niedociągnięć, naruszeniu jakiejś normy społecznej, odwołujemy się do niekontrolowanych (choćby tylko pozornie) przyczyn, starając się zredukować negatywne oceny. Atrybucje wywołują emocje, a te kierują dalej zachowaniem. Można wyobrazić sobie, że badaniu poligraficznemu poddawany jest młody człowiek z małej miejscowości, który wychowywał się w rodzinie patologicznej, w bardzo skromnych warunkach, obracał się w nieciekawym towarzystwie rówieśniczym i próbował zerwać ze swoim środowiskiem i przeszłością, zrobić krok naprzód. W wywiadzie inteligentny, silnie zmotywowany do podjęcia nowej pracy, która może radykalnie zmienić jakościowo jego życie. Kandydat opowiada o faktach z okresu burzliwej wczesnej

²³ Zob. *ibidem*.

młodości – drobnym incydencie z kradzieżą, a także udziałem w bójce. W czasie testów przesiewowych następuje potknięcie, pojawiają się wątpliwości, np. w kwestii dopuszczenia się pobicia. Wobec tego, w zgodzie z procedurami i zasadą ekonomiki pracy, badanie mogłoby zostać zakończone, a w miejsce tego kandydata szansę otrzymałby kolejny, jednak subiektywne doświadczenie współczucia, litości przez badającego może prowadzić do udzielenia pomocy – podjęcia dodatkowych prób usunięcia zaistniałych wątpliwości, przez co ostateczne konkluzje z badania byłyby na tyle korzystne dla badanego, aby nie przerywać procesu kadrowego. Taka adwokacka postawa wobec osób badanych może być nieświadoma – tym bardziej że większość ludzi na starcie podchodzi pozytywnie do swoich rozmówców i łatwiej zauważa zalety aniżeli wady (taką deformację informacji nazywamy błędem łągodności)²⁴.

1.3. Presja otoczenia, wpływ informacji przekazywanych biegłemu na późniejsze opiniowanie

A. Krzyścin słusznie wskazuje, że tak osoby badane, jak i biegli podlegają oddziaływaniu różnych grup społecznych, przy czym istnieją grupy nacisku specyficzne dla badanego i badającego. W przypadku badanych będzie to presja wywierana przez osoby z bliższego i dalszego jej otoczenia, wpływające na jej zachowania, przyjmowany przez nią system wartości oraz przekonania.

²⁴ Spostrzeżenia te mają zastosowanie właściwie jedynie w odniesieniu do badań przedzatrudnieniowych, które z natury rzeczy mają charakter przesiewowy i wymagają usuwania potencjalnych wątpliwości względem wiarygodności kandydata „krok po kroku”. W sytuacji, gdy wyniki zasadniczych testów nie są od razu korzystne dla osoby badanej i wskazują na jakieś problematyczne zagadnienia – w znacznej mierze to od poligrafera zależy, czy na tym zakończy badanie, czy będzie próbował te zagadnienia dalej weryfikować, poświęcać czas i energię na dodatkowe testy i rozmowę z badanym. W tym sensie „pomoc” oznacza niejako szansę dla badanego, aby uzupełnił pominięte informacje lub je doprecyzował, co może ostatecznie przełożyć się na lepszy dla niego wynik końcowy. Badania procesowe mają inny charakter, są konkretne, tj. odnoszą się do znanego problemu, co zwykle przekłada się też na inne rodzaje testów niż w badaniach przesiewowych. Rozstrzygnięty test diagnostyczny zamyka sprawę z punktu widzenia badającego. Ograniczone są też możliwości rozmowy z osobą badaną w warunkach kodeksowego zakazu dowodowego związku badania poligraficznego z przesłuchaniem.

Specyficzną grupą nacisku na eksperta są zleceniodawcy badania, a przede wszystkim przekazane mu informacje na temat faktów i okoliczności zdarzenia oraz posiadanych przez nich dowodów, poszlak i wiadomości ze źródeł operacyjnych²⁵.

W warunkach idealnych biegły z zakresu badań poligraficznych powinien formułować wnioski według ściśle określonych zasad, opierając się wyłącznie na logicznej syntezie danych uzyskanych w toku ekspertyzy, w oderwaniu od nacisków, zewnętrznych informacji zniekształcających procesy atrybucyjne oraz wewnętrznych motywacji innych niż te związane z potrzebą profesjonalnego wykonania zadania. Wiemy jednak, że natura ludzka jest ułomna i procesy poznawcze mają swoje ograniczenia. W konsekwencji swobodne interpretacje, subiektywne oceny narażone są na błędy. Subiektywizm w ekspertyzie dotyczy więc z jednej strony zakresu swobody biegłego, na który pozwala mu metoda badawcza, z drugiej zaś zależy od czynników psychologicznych wpływających na proces podejmowania decyzji²⁶.

Biegły z zakresu badań poligraficznych nie tylko przetwarza pomiary fizjologiczne zarejestrowane w czasie testów z wykorzystaniem poligrafu, ale pozostaje pod wpływem wielości innych informacji zebranych przed przystąpieniem do badania, a także w trakcie jego przeprowadzania. Jest odpowiedzialny za wiele czynności obejmujących m.in. wywiad z osobą badaną, formułowanie pytań, ewaluację danych i podjęcie decyzji o finalnej konkluzji. Punktem wyjścia są informacje przekazane przez zleceniodawcę ekspertyzy. E. Elaad i *et al.* stwierdzają więc, że w metodologię badań poligraficznych, w której ramach zasadniczym podejściem jest technika pytań porównawczych, inherentnie wpisana jest kontaminacja²⁷. Przez to pojawiają się czynniki subiektywne, tendencyjność. Od tego, z czym wcześniej zapozna się poligrafer, zależy treść zadawanych pytań i generalnie – taktyka badania. Można by również podejrzewać, że w jakimś sensie także

²⁵ A. Krzyścin, *Ocena wyników badania poligraficznego* [niepublikowany skrypt w dyspozycji Agencji Bezpieczeństwa Wewnętrznego].

²⁶ Por. M. Leśniak, *Wartość dowodowa opinii pismoznawczej*, Pińczów 2012; B. Leśniak, M. Leśniak, *Subiektywność w opiniach biegłych sądowych – prawne, etyczne i psychologiczne aspekty*, „Zeszyty Naukowe Politechniki Śląskiej”, 131, 2018, s. 287–297.

²⁷ E. Elaad, A. Ginton, G. Ben-Shakhar, *The Effects of Prior Expectations and Outcome Knowledge on Polygraph Examiner's Decisions*, „Journal of Behavioral Decision Making”, 7, 1994, s. 280–281.

ewaluacja zapisów na poligramach i opinia końcowa. Nie jest to jeszcze równoznaczne z istotnym zwiększeniem poziomu błędu w testach – tym bardziej że poligraferzy są szkoleni w tym, co robić, aby zmaksymalizować dokładność decyzji. Pozostaje więc pytanie: jak wielki może być ten wpływ wcześniejszych informacji na dokonywanie ocen? Problem jest tym donioślejszy, że niektórzy uważają, iż na większości wykresów z badań można odszukać jakieś dane potwierdzające niemal każdą z możliwych hipotez²⁸.

G. Barland w swojej rozprawie doktorskiej, obronionej w 1975 r. na Uniwersytecie Utah w Salt Lake City, przedstawił, że 17 z 19 poligraferów, którzy podzielili się swoim wstępnymi opiniami jeszcze przed badaniami (na podstawie materiałów, które otrzymali) – podtrzymało swoje opinie również po testach z wykorzystaniem poligrafu²⁹. W 1986 r. trzech prywatnych poligraferów ośmieszyła amerykańska stacja CBS. Dziennikarze podali się za przedstawicieli magazynu fotograficznego, zlecając im badania w sprawie fikcyjnej kradzieży aparatu i soczewek z biura. Powiedzieli, że z okoliczności sprawy wynika, że musiał to zrobić ktoś z czterech pracowników i każdemu z poligraferów zasugerowano, kto konkretnie mógłby to być, przy czym poszczególnym *ekspertom* wskazano inną osobę. Wszyscy wydali opinie zgodne z wcześniejszymi oczekiwaniami, obciążając *podejrzanego* i oczyszczając pozostałych³⁰.

Wpływ uprzednio przekazywanych biegłemu informacji badano szerzej od początku lat 90. w Izraelu. W pierwszym eksperymencie poligraferom z co najmniej trzyletnim stażem zawodowym przekazano poligramy z zapisami z testów pytań porównawczych (CQT) z 14 rzeczywistych spraw kryminalnych. Prawidłowa analiza danych metodą numeryczną (7-pozycyjną) w opinii trzech doświadczonych biegłych (z ponad 15-letnim stażem), którzy tworzyli swoistą grupę kontrolną w eksperymencie, powinna prowadzić do wyników nierozstrzygniętych. Z kolei grupa 10 poligraferów została podzielona na kolejne dwie (równe), w których ramach każdy miał dokonać indywidualnej analizy danych testowych z 7 badań na zasadzie ślepej oceny (bez otrzymania opisu sprawy, treści pytań testowych i oryginalnych

²⁸ *Ibidem*, s. 281.

²⁹ G.H. Barland, *Detection of Deception in Criminal Suspects. A Field Validation Study*, Salt Lake City 1975.

³⁰ E. Elaad, A. Ginton, G. Ben-Shakhar, *The Role of Prior Expectations in Polygraph Examiners Decisions*, „Psychology, Crime & Law”, 4, 1998, s. 4.

wyników). Pierwszą z tych grup poinformowano, że badany ostatecznie przyznał się do zarzucanego mu czynu. Drugą natomiast, że do przestępstwa przyznała się inna osoba. Aby uwiarygodnić zadanie, wszystkim powiedziano, że część zapisów została niepoprawnie oceniona przez biegłych, którzy osobiście przeprowadzali te badania poligraficzne i zrezygnowali z zarejestrowania dodatkowych wykresów. Zalegającym celem eksperymentu było ustalenie, czy ograniczenie zapisów do trzech wykresów zwiększy dokładność ślepych ocen.

Rezultaty potwierdziły występowanie błędu konfirmacji w analizach poligraferów. W grupie nakierowanej na niewinność badanego 21% wyników sklasyfikowano jako NDI (z j. ang. *no deception indicated* – brak wskazania na wprowadzanie w błąd). Żaden zestaw poligramów nie został oceniony jako DI (z j. ang. *deception indicated* – wskazanie na wprowadzanie w błąd), a pozostałe oceniono jako INC (z j. ang. *inconclusive* – nierozstrzygnięte). Średnia ocena numeryczna dla pojedynczego zapisu przy pytaniu relewantnym wyniosła +0,369.

Natomiast w grupie z wyreżyserowaną winą osób badanych uzyskano 8,6% opinii typu NDI i 4,3% DI, przy średniej ocenie dla pojedynczego zapisu -0,368. W grupie kontrolnej było to +0,076³¹. Dało się zaobserwować ponadto przypadki skrajne – tzn. byli tacy poligraferzy, którzy silnie ulegli sugestii i tacy, którzy okazali się na to całkowicie odporni, co może oznaczać, że podatność na błąd konfirmacji jest różna między biegłymi. Nie wykazano z kolei istotnych różnic w sposobie oceniania, mając na względzie kryterium doświadczenia zawodowego.

Powyższy eksperyment został zreplicowany przez tych samych badaczy, aby sprawdzić, czy wyjściowa wiedza przed badaniem rzutuje na interpretację danych testowych również w sposób wyraźnie kontrastujący z tym, co zarejestrował poligraf. Poprzednio oceniającym przekazano wyłącznie poligramy z zapisami niejednoznaczными, trudnymi do analizy. Tym razem wykorzystano materiały z badań zakończonych rozstrzygniętymi wynikami. Przekazano tym samym poligraferom, którzy uczestniczyli w pierwszym eksperymencie, po sześć zestawów poligramów wyraźnie wskazujących na szczerść i nieszczerść osób badanych. Na końcu ujawniono wszystkim uczestnikom prawdziwe cele eksperymentów, co spotkało się z ich zaskoczeniem, ale pełnym zrozumieniem zasadności takiego postępowania, włącznie z manipulacją.

³¹ *Eidem, The Effects of Prior Expectations and Outcome Knowledge...*, op. cit., s. 284–285.

W eksperymencie drugim nie wystąpiły istotne statystycznie różnice między grupami oceniających, jeżeli chodzi o ewaluację numeryczną zapisów i konkluzje po testach, choć kierunek przypisywania ocen (dodatni lub ujemny) okazał się zgodny z przewidywaniami (w grupie, na którą oddziaływano informacjami o winie osób badanych, średnia ocena zapisu w jednym analizowanym parametrze wyniosła $-0,063$, zaś w grupie ukierunkowanej na niewinność badanego: $+0,064$). Badacze postulowali na przyszłość poszerzenie eksperymentów dotyczących wpływu wcześniejszych oczekiwań nie tylko na samą (ślepią) interpretację zapisów, ale na cały proces badania poligraficznego – tak, aby uwzględnić ewentualny wpływ tych oczekiwań również na postępowanie badających wobec osób poddawanych testom³². Podjęli taką próbę kilka lat później³³.

Zaaranżowali sytuację, w której 38 studentom przydzielono pewne zadanie do wykonania w taki sposób, że część mogłaby się zdecydować na oszukiwanie – zwłaszcza że do osiągnięcia określonego wyniku zostali zmotywowani werbalnie i materialnie. Z kolejnego etapu eksperymentu wykluczono tych studentów, którzy zawyżyli swój rzeczywisty wynik (co ustalono dzięki kalkom, które ujawniły co i kiedy probanci pisali) i tych, którzy nie rozwiązali poprawnie żadnego zadania. W ten sposób wyłoniono grupę 28 osób niewinnych (podzieloną na pół – z punktacją poniżej 15 pkt i od 15 pkt wzwyż). Osoby te poddano badaniom poligraficznym przeprowadzonym przez siedmiu poligraferów z izraelskiej policji, którzy sądzili, że odbywa się to w związku z postępowaniem komisji dyscyplinarnej Uniwersytetu Hebrajskiego. Dodatkowo powiedziano badającym, że osiągnięcie wyniku na poziomie co najmniej 15 pkt jest całkiem nieprawdopodobne, o ile ktoś nie oszukiwał. Każdy z poligraferów przeprowadził po cztery badania. Aby uwiarygodnić manipulacyjny scenariusz eksperymentu, pierwsza z osób badanych poinformowała poligrafera, że uzyskała 20 pkt i oszukiwała podczas zadania. Pozostałe odpowiadały już tylko szczerze. Eksperymentatorzy próbowali wpływać na poligraferów stwierdzeniami typu: *Nie ma sensu badać tej osoby i tak będzie miała wynik wskazujący na prawdomówność* (przy trzech studentach z niskimi wynikami), albo – *Czy to możliwe, by ta osoba osiągnęła tak wysoki wynik bez oszukiwania?* (przy pięciu studentach z najwyższą

³² *Ibidem*, s. 289.

³³ *Eidem*, *The Role of Prior Expectations in Polygraph...*, *op. cit.*, s. 1–16.

punktacją). Poligraferzy mieli ponadto sporządzić notatki z rozmowy przedtestowej i opisać swoje wrażenia na temat poszczególnych badanych, które następnie zostały sklasyfikowane na kilka kategorii przez niezależnych ekspertów pod względem ich treści (typu: *badany spokojny i pewny siebie; przejęty pytaniami porównawczymi; podenerwowany, bez wskazówek, że ma to związek z pytaniami porównawczymi; wydaje się nieszczerzy*). Kategorie podkreślające stres badanego lub podejrzliwość wobec niego świadczyły o możliwości oczekiwania wyniku testu wskazującego na nieszczerłość. Na koniec testy zostały poddane analizie przez badających oraz dwóch innych doświadczonych poligraferów zaangażowanych do ślepej oceny. Współczynniki korelacji wyników między jedną i drugą grupą były wysokie (0,92 i 0,83). Ponadto odsetek wyników fałszywych pozytywnych wyniósł u siedmiu sprawdzanych poligraferów jedynie 7,1%, a liczba badań nierozstrzygniętych – 10,7%. Zatem manipulacja wcześniejszymi sugestiami nie wywarła efektu na badających, którzy weszli w interakcję z osobami badanymi (przeciwnie do poprzednich eksperymentów, gdzie nastąpił silny efekt przy ślepo oceniających).

Potwierdzono natomiast, że identyfikacje poczynione przez poligraferów korespondowały z ich wrażeniami na temat osób badanych z rozmowy przedtestowej. To wszystko oznacza, że badający integrują różne informacje – zarówno te związane z wcześniejszymi oczekiwaniami przed badaniem, wskazówki werbalne i niewerbalne z wywiadu, jak i dane fizjologiczne zarejestrowane na poligramach. Autorzy stawiają hipotezę: jeśli dane behawioralne i fizjologiczne stoją w sprzeczności z wcześniejszymi oczekiwaniami, efekt tych ostatnich zostaje zminimalizowany. Przyznają też, że mogły wystąpić i inne okoliczności – np. stopień manipulacji wcześniejszymi oczekiwaniami okazał się zbyt słaby albo badający byli sceptyczni odnośnie do natury eksperymentu. Wreszcie trzecia możliwość – bardzo prawdopodobna – większość zapisów mogła wyraźnie wskazywać na zmiany reakcji występujące zwykle u osób prawdomównych (czyli były łatwe do oceny, pozbawione niejednoznaczności), a jak już udowodniono – w takich warunkach górę biorą ściśle dane fizjologiczne.

W 2009 r. wyniki swojego eksperymentu opublikowali T. Shurny *et al.* Grupa 82 poligraferów z Meksyku, USA i Bułgarii otrzymała do analizy metodą numeryczną w skali 3-pozycyjnej osiem zestawów składających się z czterech poligramów z testów typu ZCT z dwoma

pytaniami relewantnymi (R5 i R7). Pierwszy i ósmy zestaw obejmował te same zapisy wskazujące na reakcje typowe dla osoby odpowiadającej szczerze na pytania relewantne, z potwierdzonej sprawy. Pozostałe, wykorzystane ze spraw niepotwierdzonych, wskazywały na różne wyniki (zarówno DI, jak i NDI) – potraktowano je jako bufor – tak aby oceniający nie zorientowali się, że dwa wymienione nie różnią się. Gdy wyświetlono zestaw nr 1, nie opatrzone go żadnym komentarzem. Przed prezentacją każdego kolejnego materiału przedstawiano dane ze sprawy. Przy zestawie nr 8 przekazano fikcyjną informację, że policja zabezpieczyła na miejscu przestępstwa odciski palców badanego, z 90-procentowym prawdopodobieństwem zgodności. Okazało się, że punktacja numeryczna dotycząca zmian reakcji przy pytaniach relewantnych istotnie obniżyła się (a więc w kierunku obciążającym osobę badaną) w przypadku ostatniego zestawu zapisów w porównaniu do pierwszego (średnio o 37% dla pytania R5 i o 26% dla pytania R7), zaś odsetek wyników fałszywych pozytywnych wzrósł z 1,2% do 11%. O niemal połowę spadła swoistość testu, jeśli wziąć pod uwagę również rezultaty bez rozstrzygnięcia (zamiast 67,1% wyników NDI – uzyskano 34,16%). Nie wykazano natomiast istotnych statystycznie różnic w ocenach, opierając się na kryterium kraju pochodzenia biegłego³⁴.

Jeden z ostatnich raportowanych w literaturze eksperymentów dotyczących wpływu informacji przekazywanych poligraferom na ich subiektywne procesy analizy danych zapisanych na poligramach przeprowadzili wspólnie D. Krapohl i D. Dutton³⁵. Grupie 102 poligraferów przekazali do ślepej interpretacji zestawu poligramów z 15 spraw. Zapisy z części spraw były niejednoznaczne, a pozostałe wyraźnie wskazywały na prawdomówność lub wprowadzanie w błąd przez osoby badane. Wszyscy poligraferzy mieli posłużyć się 7-pozycyjną metodą numeryczną, nauczaną przez amerykańskie Narodowe Centrum Oceny Wiarygodności (National Center for Credibility Assessment, NCCA), czyli tzw. systemem federalnym.

Połowie z oceniających przekazano, że stan faktyczny we wszystkich sprawach potwierdzał nieszczerść badanych, a pozostałym zaznaczono, że każdy z badanych został zweryfikowany jako prawdomówny.

³⁴ T. Shurany, J.A. Matte, E. Stein, *Influence of Case Facts on Blind Scorers of Polygraph Tests*, „European Polygraph”, 3–4(9–10), 2009, s. 133–139.

³⁵ Zob. D.J. Krapohl, D.W. Dutton, *Believing Is Seeing. The Influence of Expectations on Blind Scoring of Polygraph Data*, „Polygraph”, 47(2), 2018, s. 91–107.

Spośród poligraferów, do których zaadresowano poligramy, odpowiedziało 34 z grupy z wcześniejszymi oczekiwaniami nieszczerości (DE – *deception expectation*) i 39 z grupy z oczekiwaniami prawdziwości (TE – *truthful expectation*). Przedstawione przez nich wyniki świadczą o tym, że generalnie punktacja numeryczna i końcowe decyzje zmierzały w kierunku uprzedzających informacji. Spośród 15 możliwych decyzji dla każdej z grup w grupie TE uzyskano średnio 7,4 wyników NDI (bez wskazań na wprowadzanie w błąd), z kolei w grupie DE już tylko 3,2. Natomiast wyników DI (wskazujących na nieszczerść) w grupie TE było średnio 4,0 – w porównaniu do 7,8 w grupie DE. Różnice te okazały się statystycznie znaczące. Widać je jeszcze dobitniej, gdy weźmie się pod uwagę średnie sumy ocen numerycznych przypisywanych przez poligraferów reakcjom na testowe pytania krytyczne (zob. tab. 1). W sytuacji, gdy wcześniejsze oczekiwania oceniających nie pasowały do tego, co można było względnie obiektywnie stwierdzić na podstawie danych zarejestrowanych przez poligraf, znaczna część tych rozbieżności przejawiała się w klasyfikowaniu wyników testów jako nierozstrzygniętych, a zdecydowanie rzadziej – przeciwnych do tego, co wynikałoby z analiz nieobciążonych wcześniejszymi oczekiwaniami.

Tab. 1. Średnie sumy całkowite dla pojedynczego testu – w grupach poligraferów z wcześniejszymi oczekiwaniami szczerości i nieszczerości – w badaniach, których wyniki zostały pierwotnie uznane za wskazujące na szczerść, nieszczerść oraz nierozstrzygnięte

	Badania z wynikami uznanymi wyjściowo za NDI	Badania z wynikami uznanymi wyjściowo za DI	Badania nierozstrzy- gnięte
Oczekiwanie szczerości	+10,41	-6,25	+2,23
Oczekiwanie nieszczerości	+3,68	-13,42	-3,33
Różnica	6,73	7,17	5,56

Źródło: oprac. na podst. D.J. Krapohl, D.W. Dutton, *Believing Is Seeing. The Influence of Expectations on Blind Scoring of Polygraph Data*, „Polygraph”, 47(2), 2018, s. 99.

Autorzy eksperymentu opracowali *wskaźnik tendencyjności (bias index)*, zestawiając wyniki poligraferów-probantów (na których oddziaływano przekazanymi przed analizami zapisów informacjami) i tych, którzy dokonali ocen wyjściowych (pierwotnych). Jeśli poligraf podążył w kierunku manipulowanych informacji, sprzecznym

z decyzją oryginalnego badającego, otrzymywał 2 pkt, albo 1 pkt – gdy wskazał na wynik nierozstrzygnięty. Zatem przy pięciu sprawach w danej kategorii wyników (NDI lub DI) mógł otrzymać maksymalnie 10 pkt. Kolejna możliwość dotyczyła pięciu testów, które badający uznali najpierw za nierozstrzygnięte. Jeśli oceniający w ramach eksperymentu wydał dla takiego testu opinię rozstrzygającą, zgodną z wcześniejszymi oczekiwaniami – otrzymywał 1 pkt. W skrajnym przypadku łączna wartość wskaźnika tendencyjności osiągnęłaby 15 pkt (10 + 5). Z kolei wartość zerowa oznaczałaby, że oceniający w ogóle nie uległ wcześniejszym oczekiwaniam co do wyników testów. W zależności od wartości wskaźnika wyodrębniono trzy grupy poligraferów: z niską, średnią i wysoką tendencyjnością (rozkład wyników przedstawia tab. 2).

Tab. 2. Liczba poligraferów przyporządkowanych w danej kategorii tendencyjności, z podziałem na grupy eksperymentalne z wcześniejszymi oczekiwaniami szczerości i nieszczerości

	Oczekiwanie szczerości	Oczekiwanie nieszczerości
Niska tendencyjność (0–5 pkt)	27 (69,2%)	18 (52,9%)
Średnia tendencyjność (6–10 pkt)	8 (20,5%)	12 (35,3%)
Wysoka tendencyjność (11–15 pkt)	4 (10,3%)	4 (11,8%)
Suma	39 (100%)	34 (100%)

Źródło: oprac. na podst. D.J. Krapohl, D.W. Dutton, *Believing Is Seeing...*, op. cit., s. 100.

Uśredniony wskaźnik tendencyjności dla grupy z wcześniejszymi oczekiwaniami nieszczerości wyniósł: 5,58 (w przedziale od 0 do 15 pkt). Natomiast w grupie z wcześniejszymi oczekiwaniami prawdomówności: 4,44 (w przedziale od 0 do 13 pkt). Nie zaobserwowano przy tym istotnej korelacji między stażem praktyki a wysokością omawianego wskaźnika. Skoro jednak widzimy, że jedni poligraferzy w ogóle nie sugerowali się dodatkowymi informacjami przy ewaluacji zapisów na poligramach, a inni czynili to nawet za każdym razem – świadczy to o zróżnicowanej podatności na manipulację i uleganie subiektywnym czynnikom przy opiniowaniu. Wydaje się, że jest to wypadkowa osobistych właściwości poligrafera i jego merytorycznych kwalifikacji (z którymi – jak wykazano wyżej – nie zawsze idzie w parze doświadczenie zawodowe utożsamiane z latami praktyki).

Problem subiektywizmu i ewentualnego tendencyjnego oceniania został pilotażowo sprawdzony w populacji polskich poligraferów. W 2013 r. przygotowano eksperyment z udziałem 20 poligraferów praktykujących w instytucjach państwowych (m.in. Agencji Bezpieczeństwa Wewnętrznego, Straży Granicznej, Służby Kontrwywiadu Wojskowego, Służby Wywiadu Wojskowego, Żandarmerii Wojskowej i Policji), a także w sektorze prywatnym. Część tego eksperymentu była poświęcona ewentualnemu wpływowi informacji przekazanych w związku z analizą zapisów z badania poligraficznego. 16 badających zgodziło się rozwiązać przekazane zadania w czasie V Międzynarodowego Seminarium Poligraficznego zorganizowanego w ośrodku Żandarmerii Wojskowej w Waplewie (26 września 2013 r.). Pozostałe cztery osoby zostały dołączone w innym terminie, bez znajomości założeń projektu.

Zadania praktyczne poprzedziło badanie ankietowe, w którym pytano, z jakim rodzajem informacji poligraferzy zapoznawali się zwykle przed ekspertyzą w celach śledczych (w sprawie karnej lub wewnętrznym dochodzeniu), przy czym można było udzielić równocześnie kilku odpowiedzi. Zrobiło to 16 z 20 biegłych, a ich rozkład był następujący:

- a) ogólnikowe informacje na temat celu badania zawarte w postanowieniu o zasięgnięciu opinii biegłego lub innej formie zlecenia – 69% (11),
- b) propozycje konkretnych pytań lub zagadnień, problemów do zwerfikowania – 75% (12),
- c) akta sprawy – 44% (7),
- d) możliwe warianty wersji śledczych (bez rozróżniania zakładanego prawdopodobieństwa poszczególnych wersji) – 19% (3),
- e) sugestie na temat konkretnego podejrzenia wobec badanego przekazane ustnie przez zleceniodawcę lub zawarte w postanowieniu o przedstawieniu zarzutów / akcie oskarżenia – 63% (10),
- f) kategoryczne przekonanie zleceniodawcy co do stanu faktycznego, który badanie poligraficzne ma tylko potwierdzić przy niedostatecznej liczbie dopuszczonych formalnie innych dowodów – 6% (1)³⁶.

Można powiedzieć, że w większości wyżej opisanych sytuacji poligrafer styka się z takimi informacjami, które mogą oddziaływać na

³⁶ M. Gołaszewski, niepublikowany materiał szkoleniowy, Waplewo wrzesień 2013.

świadome lub nieświadomie oczekiwania wobec wyników. Najrzadziej badający doświadczali bardzo silnie sugestywnego oddziaływania ze strony zleceniodawcy, który wprost wyrażał swoje przekonanie w kwestii winy osoby badanej (6% przypadków). Sam fakt, że takie okoliczności mają miejsce, może jednak niepokoić i wymaga od biegłego pewnej samoświadomości i dyscypliny w zachowaniu obiektywizmu.

Następnie poligraferów podzielono na dwie równe grupy, z których każda musiała wykonać część zadań wspólnych (które zostały poruszone jeszcze osobno w części II, w rozdziale 5) i jedno, które różnicowało obie grupy. Obie oceniały dwa zestawy poligramów. Jeden zestaw składał się z zapisów reakcji fizjologicznych zarejestrowanych podczas przeprowadzania w warunkach laboratoryjnych testu typu MGQT z czterema pytaniami krytycznymi, z udziałem poligraferów Agencji Bezpieczeństwa Wewnętrznego i Żandarmerii Wojskowej. Osoba badana odpowiadała wówczas na pytania krytyczne zgodnie z prawdą – co znalazło odzwierciedlenie w zespołowej ocenie numerycznej dokonanej zgodnie z regułami Empirycznego Systemu Oceniania przez trzech certyfikowanych poligraferów, którzy nie byli probantami w eksperymencie i wydali opinię NSR (z j. ang. *no significant responses* – brak znaczących reakcji, z najniższą oceną cząstkową +2 przy pytaniu oznaczonym jako R2). Drugi zestaw był złożony z trzech poligramów zarejestrowanych podczas rzeczywistych badań, w ramach których wykorzystywano testy typu ZCT z dwoma pytaniami krytycznymi. Wykresy były zaczerpnięte z trudno wówczas dostępnej w Polsce publikacji znanych ekspertów z Izraela, którzy dołączyli do nich własne oceny metodą numeryczną w skali 7-pozycyjnej³⁷. Prawidłowa suma ocen zapisów na wykorzystanych w eksperymencie poligramach powinna prowadzić do opinii nierozstrzygującej (INC – z j. ang. *inconclusive*).

Dobrano takie poligramy, które można określić jako skomplikowane, tzn. wymagające od oceniającego pogłębionej analizy z wykorzystaniem zaawansowanej wiedzy specjalistycznej. Założono, że gdyby zapisy były zbyt proste i oczywiste, moglibyśmy nie uchwycić wpływu interesujących nas kontekstów na dokonywane przez poligraferów oceny.

Do jednego z tych zestawów dołączono krótki opis sprawy z fikcyjnymi danymi, ale wzorowanymi na jednym z postępowań karnych,

³⁷ T. Shurany, I. Ravid, *Evaluation of Polygraph Charts...*, *op. cit.*

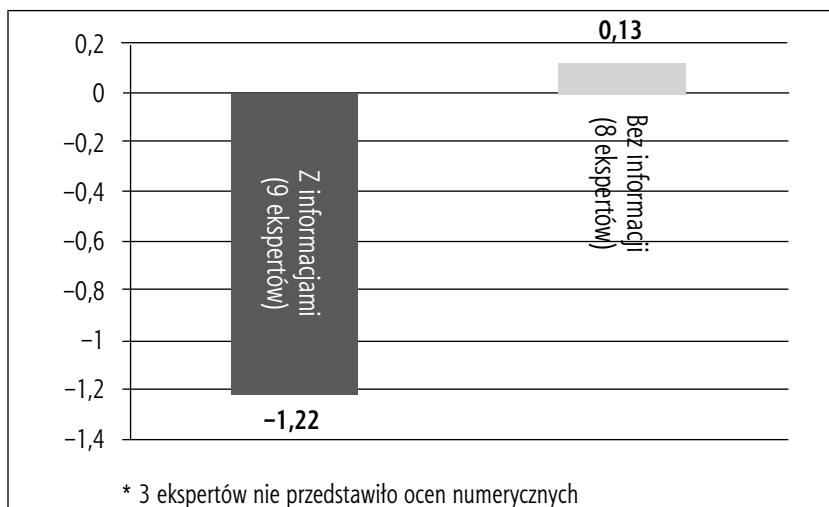
w którym autor niniejszej monografii został zaangażowany osobiście w charakterze biegłego. Chodziło przede wszystkim o wybrane informacje na temat biografii badanego (dane o karalności za inne czyny w przeszłości), opinie biegłych innych specjalności (w tym przypadku opinia rodzinnego ośrodka diagnostyczno-konsultacyjnego o wiarygodności wersji dziewczynki, która miała być rzekomo molestowana seksualnie) i zeznania niektórych świadków (np. ze szkoły – wskazujące na to, że dziewczynka bała się ojca, który odbierał ją po lekcjach). Wszystkie te informacje stawiały badanego na starcie w dość niekorzystnym świetle, tworzyły więc kontekst w kierunku oczekiwania wyniku testu poligraficznego typowego dla osoby odpowiadającej nieszczercze.

W obu grupach wykorzystano te same dane na temat sprawy, lecz w zestawieniu z różnymi testami (zestawami poligramów). W ten sposób każdy test (MGQT – prawidłowo rozstrzygnięty jako NSR oraz ZCT – prawidłowo nierozstrzygnięty) był oceniany zarówno przez poligraferów, którzy uprzednio zapoznali się ze sprawą, jak i tych, którzy nie brali pod uwagę żadnych dodatkowych informacji i skupili się wyłącznie na wykresach poligraficznych.

Analizy danych testowych należało dokonać za pomocą potwierdzonej naukowo metody numerycznej – dowolnej w przypadku testu MGQT i 7-pozycyjnej w przypadku testu ZCT. Nadmienić należy, że w przypadku testu MGQT decydująca o jego ogólnym wyniku jest najniższa ocena cząstkowa (dla najniższej ocenionego pytania relewantnego), natomiast w teście typu ZCT obowiązuje zasada dwuetapowej analizy – najpierw bierze się pod uwagę sumę całkowitą, a w razie nieosiągnięcia punktu odcięcia od zbioru wyników wykraczających poza akceptowany poziom tolerancji błędów – obowiązuje dodatkowo sprawdzenie wyników cząstkowych, dla których istnieje odrębny próg decyzyjny.

W teście MGQT probanci – podobnie jak sędziowie kompetentni z grupy kontrolnej – najniżej ocenili pytanie relewantne oznaczone jako R2. Jednakże w tej grupie, która dla tego testu otrzymała wyżej opisane dane dotyczące sprawy, ocenianie przebiegało w przewidywanym kierunku, czyli ujemnym. Średnia ocen dla pytania R2 wyniosła tutaj $-1,22$. W grupie, która dokonała klasycznej ślepej oceny, była wyższa: $+0,13$ (zob. ryc. 2).

Ryc. 2. Średnia ocen numerycznych dla najniżej ocenionego pytania krytycznego w teście wieloaspektowym typu MGQT, analizowanym przez dwie eksperymentalne grupy poligraferów (tych, którym przekazano informacje dotyczące sprawy i osoby badanej oraz tych, którzy dokonywali ślepo interpretacji)

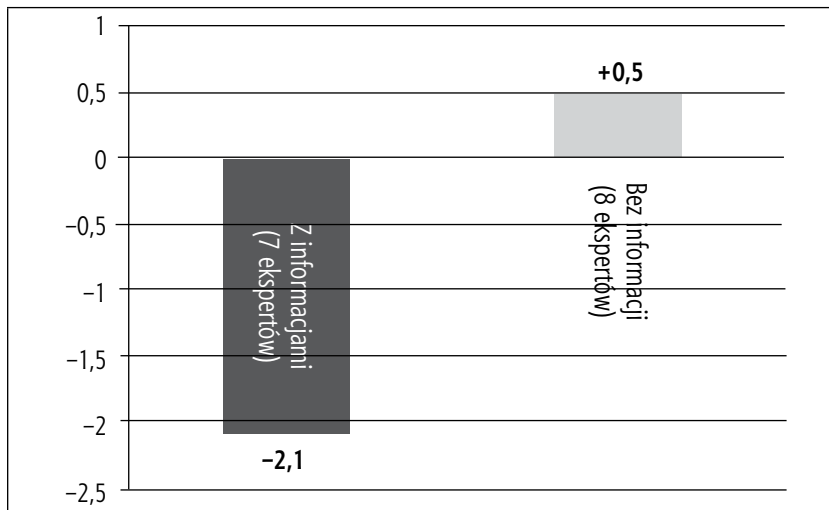


Źródło: opracowanie własne autora.

Gdyby spojrzeć tylko na te średnie – wynik badania rozstrzygniętego zmieniłby się na nierozstrzygnięty, przy czym w grupie z kontekstem winy badanego zauważalna była tendencja w stronę wyniku fałszywego pozytywnego. W przełożeniu na końcowe opinie w grupie oceniających test MGQT, którzy nie zapoznawali się z dodatkowymi informacjami, wydano jedną nietrafną opinię (SR – z j. ang. *significant responses* – znaczące reakcje) – co stanowi 10% wszystkich opinii wydanych w tej grupie (w zbieżności z ogólną wartością diagnostyczną technik badań poligraficznych). Natomiast w grupie, która zapoznała się ze szczegółami sprawy w zestawieniu z tym testem, wydano trzy nietrafne opinie (SR), co stanowi 30% wszystkich opinii wydanych w tej grupie.

Jeżeli chodzi o drugi z testów – ZCT – średnia ocen numerycznych dla pytań relewantnych R1 i R2 (traktowanych łącznie ze względu na jednoprobemowy charakter testu) wyniosła w grupie ślepo oceniających +0,5. Natomiast u tych poligraferów, którzy wraz z treścią zadania mieli przedstawione okoliczności sprawy, średni wynik był niższy – nie dodatni, lecz ujemny: -2,1 (zob. ryc. 3).

Ryc. 3. Średnia ocen numerycznych dla pytań krytycznych w teście jednoproblemowym typu ZCT, analizowanym przez dwie eksperymentalne grupy poligraferów (tych, którym przekazano informacje dotyczące sprawy i osoby badanej oraz tych, którzy dokonywali ślepej interpretacji)



Źródło: opracowanie własne autora.

Na podstawie powyższego można stwierdzić, że widać tendencję, że te same informacje wpływają w jednakowym kierunku na analizę danych fizjologicznych z zupełnie niepowiązanych ze sobą testów. W grupie z kontekstem winy badanego dwóch oceniających wydało opinię DI (z j. ang. *deception indicated* – wskazanie na wprowadzanie w błąd), pozostali (80%) INC. Nikt nie zdecydował się na opinię o reakcjach typowych dla osoby, która w błąd nie wprowadza. Inaczej było w grupie „ślepo oceniających” – tam wydano aż pięć opinii NDI, trzy – INC i dwie – DI. Tendencja związana z wcześniejszymi informacjami uwidoczniła się w tych wynikach, a jednocześnie poziom skomplikowania zapisów zapewne przyczynił się też do obniżenia zgodności między oceniającymi. Można ponadto sądzić, że w sytuacji eksperymentalnej znaczna część poligraferów nie przywiązywała do zadań takiej wagi i nie poświęcała tyle czasu, jak miałyby to miejsce, gdyby ich opinia miała potencjalne konsekwencje dla losów osoby badanej, bezpieczeństwa wewnętrznego danej instytucji czy wymiaru sprawiedliwości.

Dla porządku należy uzupełnić, że 20 probantów przedstawiło swoje konkluzje po testach, natomiast nie wszyscy opatrzyli je wynikami liczbowymi. Przy teście MGQT zrobiło to 17 oceniających, a ZCT – 16. Liczba zaangażowanych poligraferów może wydawać się nieduża, ale wcale nie odbiega od innych eksperymentów tego typu, które prowadzono na świecie. W Polsce problemem jest przede wszystkim niewielka liczba poligraferów (kilkadziesiąt osób w skali kraju) i ich dostępność (większość zatrudniona jest w służbach mundurowych, w tym specjalnych, a więc w środowiskach hermetycznych). Trzeba też zauważyć, że zebranie większej próby badawczej pośród poligraferów było i jest zadaniem trudnym także dlatego, że jak zauważają I. Dror i R. Rosenthal – eksperci oraz ich organizacje są bojaźliwe w kwestii poddawania badaniom, szczególnie zaś kiedy przedmiotem analiz są źródła błędów, rzetelność i subiektywizm. Budzi to więc naturalny opór, a na szczycie tego wszystkiego znajdują się ewentualne konsekwencje prawne związane z ujawnieniem błędów w takich dziedzinach jak medycyna czy wymiar sprawiedliwości³⁸. W Polsce znaczna część biegłych wolałaby wykazać ostrożność i nie chciałaby zostać zrecenzowana, nawet przy zadeklarowaniu anonimowości. Całkowita szacunkowa liczebność tej populacji w naszym kraju (mniej niż 100 osób) na pewno nie sprzyja prowadzeniu badań naukowych, choć trzeba też zauważyć, że rośnie wraz z dostępnością szkoleń specjalistycznych, m.in. realizowanych przez biegłych z Polskiego Towarzystwa Badań Poligraficznych.

Ogółem wyniki powyższych badań były zgodne z oczekiwaniami z uwagi na znane w psychologii społecznej zjawisko prymowania, które polega na aktywizacji jakiejś struktury wiedzy, a następnie wykonywaniu przez badanych właściwego zadania – np. dokonywania ocen. Prymowanie wpływa na przetwarzanie informacji i zachowanie szczególnie – bądź wyłącznie – wtedy, kiedy sytuacja zachowuje pewien stopień wieloznaczności³⁹. Dlatego sensowny był dobór w eksperymentach przede wszystkim właśnie takich zestawów poligramów, które nie były z marszu oczywiste i wymagały bardziej skrupulatnych analiz.

Opisane eksperymenty wykazały, że biegli z zakresu badań poligraficznych sugerują się informacjami przekazanymi im przed

³⁸ I.E. Dror, R. Rosenthal, *Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts*, „Journal of Forensic Sciences”, 53(4), 2008, s. 901.

³⁹ B. Wojciszke, *Psychologia społeczna*, op. cit., s. 78–79

przystąpieniem do analizy danych testowych i to bez względu na doświadczenie zawodowe czy pochodzenie poligrafera (USA, Meksyk, Izrael, Bułgaria, Polska). W zależności od ich wydzźwięku (w kierunku winy lub niewinności) występowały znaczące różnice w ocenach numerycznych dla zapisów przy poszczególnych relewantnych pytaniach testowych, ale błąd konfirmacji miał znaczenie wtedy, gdy zapisy z badań były skomplikowane w ocenie – nierozstrzygnięte lub na granicy rozstrzygnięcia. Jeśli poligraferzy nie mogli znaleźć na poligramach danych potwierdzających hipotezę wynikającą z wcześniejszych oczekiwań, kierowali się wyłącznie parametrami fizjologicznymi. Współczesne procedury testów, optymalizacja punktowych progów decyzyjnych skutecznie ograniczają liczbę nietrafnych końcowych identyfikacji, choć są i takie propozycje, aby poszerzyć jeszcze zakres wyników uznawanych za nierozstrzygnięte⁴⁰. Zakres ten ma niejako absorbować efekty związane z czynnikami losowymi, różnicami indywidualnymi w sposobie reagowania na bodźce, zakłóceniami zewnętrznymi i kontaminacją spowodowaną aspektami pozadyskursywnymi.

W omawianych eksperymentach błędne opinie – owszem – zdarzały się, ale nie była to skala istotna statystycznie. Co więcej A. Ginton⁴¹ dokonał nowej analizy tych samych danych, które pochodziły z eksperymentu E. Elaada *et al.* z początku lat 90. Z jego ustaleń wynika, że w realnych warunkach konsekwencje wcześniejszych oczekiwań nie są aż tak duże, jak chętnie powołują się na to przeciwnicy badań poligraficznych⁴². Uznał, że dla praktyki najistotniejsze jest, ile takich *nierozstrzygniętych* testów może niepoprawnie przekształcić się w rozstrzygnięte pod wpływem wcześniejszych oczekiwań badających. Przyjął przy tym odsetek wyników nierozstrzygniętych do 20% (i słusznie – ponieważ przy aktualnych wymogach walidacyjnych dla testów rekomendowanych do różnych typów badań jest to maksymalna dopuszczalna liczba, a metaanalitycznie średni udział takich wyników został

⁴⁰ Zob. G. Ben-Shakhar, M. Bar-Hillel, I. Liebllich, *Trial by Polygraph. Scientific and Juridical Issues in Lie Detection*, „Behavioral Sciences”, 4(4), 1986, s. 459–479.

⁴¹ Pułkownik w st. spoczynku izraelskiej policji, prezydent Izraelskiego Stowarzyszenia Poligraferów (IPA).

⁴² A. Ginton, *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, 2019, s. 1–6.

oszacowany na 12,7%⁴³). A. Ginton oszacował, że błąd confirmacji dotyczy realnie ok. 3% ogółu badań w obszarze praktyki, a nie wszystkie przecież kończą się niedokładną identyfikacją, ponieważ część wcześniejszych oczekiwań pokrywa się ze stanem faktycznym i nie idą one tylko w jednym kierunku, obciążającym lub oczyszczającym badanego z podejrzeń.

Amerykańska National Research Council, będąca częścią National Academy of Sciences (NAS), w raporcie *The Polygraph and Lie Detection* z 2003 r. postulowała więcej badań empirycznych, aby sprawdzić, czy wcześniejsze oczekiwania badających i badanych rzeczywiście istotnie rzutują na trafność metody⁴⁴. Kwestia subiektywizmu i tendencyjnego oceniania nie dotyczy, rzecz jasna, tylko badań poligraficznych. Tak jest również w innych badaniach kryminalistycznych, szerzej stosowanych i cieszących się wyższym uznaniem, jak np. daktyloskopia. W innym raporcie NAS – w którym analizowano kondycję badań kryminalistycznych w USA – uznano, że w wielu przypadkach poprawy wymaga standaryzacja, rzetelność i dokładność różnych badań. Dostrzeżono też problem tendencyjności związanej z kontekstem⁴⁵.

Gdy w 2005 r. I. Dror *et al.* zbadali biegłych z zakresu badań daktyloskopijnych z kilkunastoletnim doświadczeniem, okazało się, że pod wpływem udzielonych informacji czterech na pięciu wydało inną opinię, niż sami wydali pięć lat wcześniej⁴⁶. Naukowcy krytycznie stwierdzili, że jeśli biegły nie jest spójny sam ze sobą w swoich opiniach, to jego profesjonalizm albo podstawa oceniania mogą budzić wątpliwości⁴⁷. W każdym razie wskaźnik rzetelności wewnętrznej (z j. ang. *intra-rater reliability*) pozostawia w takim wypadku wiele do życzenia. I. Dror i jego partnerzy naukowci kontynuowali swoje badania.

⁴³ M. Gougler, R. Nelson, M. Handler, D. Krapohl, P. Shaw, L. Bierman, *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), 2011, s. 194–305.

⁴⁴ National Research Council – National Academies Press, *The Polygraph and Lie Detection*, Washington 2003, s. 90.

⁴⁵ National Academy of Sciences, *Strengthening Forensic Science in the United States. A Path Forward*, Washington 2009.

⁴⁶ I.E. Dror, D. Charlton, A.E. Pèron, *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, „Forensic Science International”, 156, 2006, s. 74–78.

⁴⁷ I.E. Dror, R. Rosenthal, *Meta-analytically Quantifying...*, *op. cit.*, s. 900.

Ze względu na niewielkie próby badawcze (w sumie 11 sędziów kompetentnych) stosowali metaanalizę danych oraz odpowiednie narzędzia statystyczne mierzące wielkość efektu (wskaźnik $r_{equivalent}$). Łączne dane prowadziły do podobnych wniosków, jakie płynęły z pierwszego wspomnianego eksperymentu⁴⁸.

Warto jeszcze dodać, że na pracę biegłego oddziałują nie tylko okoliczności, z którymi jest zapoznawany przed ekspertyzą, ale również ewentualna interakcja ze zleceniodawcą i dopływ informacji z alternatywnych źródeł w trakcie realizacji badań, a nawet po ich realizacji. Trudno biegłemu przejść obojętnie nad nowymi ustaleniami w danym postępowaniu, które nie są bezpośrednio związane z jego badaniami, a zostały dokonane m.in. dzięki znalezieniu nowych dowodów rzeczowych, результатам innych ekspertyz czy przyznaniu się podejrzanego. Wiadomo na przykład, jak informacja o przyznaniu się do dokonania jakiegoś czynu przez jedną z typowanych osób może wpłynąć na decyzje świadka w związku z czynnością procesową okazania. Abstrahując generalnie od wad okazania i niepewności zeznań świadka naoczego⁴⁹, warto zasygnalizować, że w eksperymencie L. Hasel i S. Kassina aż 61% świadków zmieniło swój typ pod wpływem informacji, że do czynu przyznała się inna osoba z szeregu⁵⁰. Wyobraźmy sobie też inną sytuację – laboratorium przekazuje zleceniodawcy, że dostarczony materiał nie daje podstaw do pozytywnej identyfikacji, a oficer śledczy zwraca się o ponowne przeprowadzanie badań w nadziei – nawet artykułując to biegłemu – że następne dadzą może lepsze wyniki. W konsekwencji ryzyko tendencyjnego podejścia automatycznie rośnie⁵¹.

Zachowanie poligrafera wobec osoby badanej uwarunkowane jest podejściem, jakie stosuje przy prowadzeniu badań. Zasadniczo wyróżnia się dwa – diagnostyczne i konfrontacyjne. W pierwszym biegły koncentruje się wyłącznie na przeprowadzeniu efektywnego testu,

⁴⁸ Zob. *ibidem*, s. 900–903.

⁴⁹ Por. G.L. Wells, A. Memon, S.D. Penrod, *Eyewitness Evidence. Improving its Probative Value*, „Psychological Science in the Public Interest”, 7(2), 2006, s. 45–75.

⁵⁰ L.E. Hasel, S.M. Kassina, *On the Presumption of Evidentiary Independence. Can Confessions Corrupt Eyewitness Identifications?*, „Psychological Science”, 20, 2009, s. 122–126.

⁵¹ Szerzej na ten temat: D.M. Risinger, M.J. Saks, W.C. Thompson, R. Rosenthal, *The Daubert / Kumho Implications of Observer Effects in Forensic Science. Hidden Problems of Expectation and Suggestions*, „California Law Review”, 90(1), 2002, s. 35–42.

który ma prowadzić do jak najdokładniejszego rezultatu. W drugim poligraf służy jako taktyczne narzędzie przesłuchania, a poligrafer z góry przystępuje do czynności z nastawieniem na uzyskanie przyznania się badanego do winy. W konsekwencji może wykazywać postawę agresywną, wyrażać nieufność już na etapie wywiadu przedtestowego, nie przejmować się pytaniami porównawczymi i zachowaniem odpowiedniego balansu w stosunku do pytań krytycznych. Wreszcie podczas analizy danych testowych może część zapisów lekceważyć, a do innych przykładać zbyt dużą wagę⁵². W przypadku osób prawdomównych to podejście zwiększa ryzyko wyników fałszywych pozytywnych, m.in. z silnej obawy o błąd badającego. W Polsce i w Europie dominuje podejście czysto naukowe, kliniczne – zarówno z powodu utartej praktyki, jak i ograniczeń w przepisach prawa (art. 171 § 5 k.p.k. – zakaz dowodowy łączenia badania poligraficznego z przesłuchaniem). W USA przed laty królowało podejście konfrontacyjne. Dziś już w mniejszym stopniu, choć naturalne (oprócz USA, także w Izraelu), pozostaje przechodzenie do fazy przesłuchania na koniec badania w przypadku wyników wskazujących na nieszczerłość osoby badanej.

Przykładem fatalnej pomyłki sądowej, do której przyczyniło się badanie poligraficzne przeprowadzone w konfrontacyjny sposób, jest sprawa *Deskovic v. City of Peekskill et al.*⁵³ z 1991 r. Oskarżony o gwałt Jeffrey Deskovic przyznał się do przestępstwa bezpośrednio po badaniu poligraficznym i został skazany na karę dożywotniego pozbawienia wolności z możliwością warunkowego zwolnienia po 15 latach. W 2006 r. ujawniono jednak dowód z badań DNA, który obciążał kogoś innego. Deskovic złożył pozew przeciw miastu i innym podmiotom, ostatecznie zawierając ugodę w kwestii odszkodowania na kwotę 6,6 mln dolarów. W procesie cywilnym wyszło na jaw, że poligrafer o nazwisku Stephens przez 6–8 godzin (nie było nagrania – stąd różne postrzeganie czasu przez strony) prowadził badanie m.in. w oskarżającym tonie, podnosił głos, nie umożliwił badanemu spożycia posiłku, stosował przestarzałą i zawodną technikę Arthera pomimo przeszkolenia w bardziej uznanej technice Backstera. Badanie zakończył słowami: „Poprzez wyniki badania poligraficznego powiedziałeś mi, że to zrobiłeś. Wszystko, czego chcemy od ciebie – to zwerbalizowanie

⁵² Zob. T.T. Amsel, *Examiner Approach and its Impact on Polygraph Results*, „European Polygraph”, 10, 1(35), 2016, s. 7–10.

⁵³ *Deskovic v. City of Peekskill*, 894 F. Supp. 2d 443 (S.D.N.Y. 2012), casetext.com/case/deskovic-v-city-of-peekskill-3 [dostęp: 8.06.2020].

tego⁵⁴. Potem poligrafera zastąpił śledczy, który rzekomo miał powstrzymać innych policjantów od zrobienia podejrzanemu krzywdy, obiecał łagodne podejście i uzyskał przyznanie się – jak się okazało – do przestępstwa, którego Deskovic nie dopuścił się. Stephens miał w swoim wydziale renomę oficera bardzo skutecznego w przesłuchaniach i poprowadził feralne badanie w taki sposób, aby potwierdzić wcześniejsze oczekiwania odnośnie do winy badanego.

Znaczenie wpływu informacji udzielanych biegłemu przed ekspertyzą na opiniowanie skłania do refleksji o tym, z czym należy biegłego zapoznać. Problem subiektywizmu i błędów atrybucyjnych dostrzegł polski ustawodawca, decydując się na zmiany w art. 198 § 1 k.p.k. przy nowelizacji z dnia 19 lipca 2019 r.⁵⁵ Do tej pory przepis mówił, że „w miarę potrzeby udostępnia się biegłemu akta sprawy w zakresie niezbędnym do wydania opinii [...]”. Po zmianie przepisowi nadano brzmienie:

Jeżeli jest to niezbędne do wydania opinii, sąd lub prokurator udostępnia biegłemu poszczególne dokumenty z akt sprawy lub uwierzytelnione kopie tych dokumentów. [...] Biegłemu powołanemu z tego względu, że wydana przez innego biegłego opinia jest niepełna lub niejasna albo gdy zachodzi sprzeczność w niej samej lub między różnymi innymi opiniami w tej samej sprawie, przed wydaniem opinii nie udostępnia się tej innej opinii lub tych innych opinii. Inną opinię lub inne opinie można udostępnić biegłemu, w niezbędnym zakresie, tylko w wyjątkowym, szczególnie uzasadnionym wypadku, gdy przedmiot opinii powołanego biegłego bezpośrednio dotyczy treści tej innej opinii lub tych innych opinii.

Sprawa budzi jednak kontrowersje. Z jednej strony ma to swoje uzasadnienie, z drugiej – sama selekcja po stronie zlecającego też może być obciążona subiektywizmem, a biegły nie otrzyma tego, co pozwoli mu w pełni przygotować się do ekspertyzy. W wielu badaniach pełne akta sprawy w ogóle nie są przydatne, np. w ekspertyzie z zakresu pisma ręcznego, daktyloskopii czy osmologii, ale w przypadku badań poligraficznych znajomość jak największej liczby szczegółów dotyczących rozpatrywanych zdarzeń wydatnie wpływa na trafniejsze formułowanie pytań testowych.

⁵⁴ *Ibidem*.

⁵⁵ Ustawa z dnia 19 lipca 2019 r. o zmianie ustawy – Kodeks postępowania karnego oraz niektórych innych ustaw (Dz.U. z 2019 r., poz. 1694).

2. Niedokładność i błędy pomiaru poligrafu

2.1. Istota pomiaru

Według K. Jaegermanna ekspertyza biegłego składa się z działań pomiarowo-opisowych, interpretacyjnych i wnioskowania⁵⁶. Pomiar mówi nam o właściwościach jakiegoś przedmiotu, zjawiska: „obejmuje m.in. użycie narzędzi i przyrządów pomiarowych oraz wykonanie obliczeń. Sposób porównywania wartości wielkości, wykorzystany przy pomiarze, nazywa się metodą pomiarową”⁵⁷.

Każdy pomiar obarczony jest błędem, niepewnością. Skala tej niepewności może wpływać na wnioski z badań i wiarygodność opinii biegłego. Błąd jest różnicą między mierzoną wartością a *prawdziwą wartością* mierzonej rzeczy. Z kolei niepewność jest miarą wątpliwości dotyczących rezultatu pomiaru¹⁾. Ta ostatnia pozwala odbiorcy zrozumieć oraz racjonalnie zważyć przedstawiony przez badacza rezultat. Niepewność i błędy pomiarów mogą wynikać z rozmaitych okoliczności: niedoskonałości przyrządu, wpływu warunków panujących w środowisku badań, niestabilności przedmiotu pomiaru i sposobu jego przeprowadzenia, umiejętności badacza czy problemów z reprezentatywnością próby badawczej.

Przez dokładność pomiaru (z j. ang. *accuracy*) należy rozumieć stopień zgodności wyniku pomiaru z wartością rzeczywistą wielkości mierzonej. Natomiast, potocznie podobnie rozumiana, precyzja pomiaru – to inaczej powtarzalność (z j. ang. *repeatability*), czyli stopień zgodności wyników kolejnych pomiarów tej samej wielkości mierzonej, wykonywanych w tych samych warunkach pomiarowych²⁾. Z naszego punktu widzenia istotniejsze będzie to pierwsze z pojęć. Znaczenie dokładności pomiaru należy rozpatrywać przede wszystkim pod względem skutków ewentualnych błędów. Sportowiec poddany kontroli antydopingowej może zostać niesłusznie oskarżony, stracić szansę dalszej kariery i pieniądze, nawet jeśli sprawa ostatecznie się wyjaśni i skończy się tylko na czasowym zawieszeniu w prawach zawodnika³⁾. Opinii publicznej znane są kontrowersje dotyczące np. pomiarów prędkości przez policyjne radary i gminne fotoradary. Nieprawidłowe pomiary glukometrem

⁵⁶ K. Jaegermann, *Opiniowanie sądowo-lekarskie. Eseje o teorii*, Warszawa 1991, s. 29.

⁵⁷ *Pomiar* [hasło], encyklopedia.pwn.pl/haslo/pomiar;4009595.html [dostęp: 15.05.2020].

u diabetyka tworzą ryzyko utraty kontroli nad poziomem cukru we krwi i zagrażają zdrowiu. Źle skalibrowany prywatny alkomat może sprawić, że nietrzeźwy kierowca podejmie decyzję o prowadzeniu pojazdu i narazi siebie na odpowiedzialność karną, a innych uczestników ruchu drogowego – na niebezpieczeństwo itd.

Świadomość różnic w dokładności badań sprawia, że od czasu do czasu sprawdzana jest zbieżność wyników jednakowych typów badań w różnych laboratoriach, choćby – trzymając się powyższych przykładów – na zawartość alkoholu i cukru w surowicy krwi. Poszczególne laboratoria prowadzą wówczas badania w odniesieniu do tej samej próbki wzorcowej albo porównuje się między takimi placówkami ich rozrzut wyników w całym zbiorze badań w danym okresie. Gdy sięgniemy do analiz Instytutu Ekspertyz Sądowych z 1966 r., zauważymy, że rozpiętość między wynikami dotyczącymi jednej z próbek zawierającej w rzeczywistości 0,23 promila alkoholu mieściła się między 0,0 promila a 0,36 promila⁴⁾ (a więc różnica ogromna, która oznaczałaby zupełnie inne konsekwencje prawne dla kierowcy). Z kolei zestawienie zbiorów badań glukozy w 21 polskich laboratoriach w 2007 r. z jednej strony pozwoliło na wyodrębnienie trzech grup, w których nie było istotnych statystycznie różnic, z drugiej natomiast okazało się, że występują znaczące różnice w rozrzucie wyników między tymi trzema grupami. Implikacją praktyczną takiego stanu rzeczy był brak możliwości przyjęcia jednolitego zakresu referencyjnego na obszarze całego kraju⁵⁾. Efekty opisanych badań porównawczych dowodzą, że problem dokładności pomiarów był i nadal pozostaje aktualny. Dlatego tak ważne jest doskonalenie urzędów, weryfikacja ich sprawności oraz prawidłowej realizacji procedur przez podmioty, które dokonują pomiarów, np. w laboratoriach kryminalistycznych czy medycznych.

¹⁾ S. Bell, *A Beginner's Guide to Uncertainty of Measurement*, Teddington 2001.

²⁾ S. Fita, *Słownik metrologiczny*, www.metrologia.pwr.edu.pl/pliki/slownik_metrologiczny.pdf [dostęp: 18.05.2020].

³⁾ Na marginesie rozważań, temat dopingu jest ciekawy z punktu widzenia możliwości wykorzystania badań poligraficznych w dochodzeniu świadomości podejrzanego sportowca o znanych lub nieznanach śledczym okolicznościach dostania się zabronionej substancji do jego organizmu. Autorowi znane są przypadki, kiedy Panel Dyscyplinarny przy Polskiej Agencji Antydopingowej dopuszczał tego typu dowody w rozpatrywanych sprawach.

⁴⁾ J. Markiewicz, J. Nedoma, H. Serda, *Z badań nad dokładnością oznaczeń alkoholu we krwi*, „Problemy Kryminalistyki”, 61–62, 1996, s. 472–478.

⁵⁾ Zob. J. Janecki, *Rozrzut rocznych wyników badania glukozy z 21 laboratoriów polskich i wynikające z niego wnioski dotyczące wdrożenia ogólnokrajowych norm glukozy we krwi*, „Diagnostyka Laboratoryjna”, 47(1), 2011, s. 77–83.

2.2. Warunki w środowisku badawczym

Jeżeli chodzi o warunki, jakie należy spełnić w środowisku badań poligraficznych – najważniejsze jest, aby zminimalizować liczbę bodźców zewnętrznych. Standardy praktyki ogłoszone przez American Polygraph Association (APA) oraz organizację normalizacyjną ASTM International mówią, że to środowisko powinno być *możliwie wolne od zakłóceń*⁵⁸. Chodzi o takie zakłócenia, które „mogłyby kolidować ze zdolnością do właściwego skupienia uwagi na poruszanych zagadnieniach”⁵⁹. Wymienia się przy tym głównie hałas, dystraktory wizualne i obecność osób trzecich w pomieszczeniu badań. Zasadniczo w pomieszczeniu do badań należy utrzymywać temperaturę typową dla pomieszczeń biurowych (20–26°C). Meble, sprzęt i rekwizyty należy tak rozmieścić, by badający znajdował się bliżej drzwi i miał swobodny dostęp do wyjścia w razie nieprzewidzianej agresji badanego⁶⁰.

Standard badań poligraficznych w sprawach karnych, opracowany w 2004 r. przez nieistniejące już Stowarzyszenie Poligraferów Polskich, wskazywał ponadto, że nie powinno się badać w pomieszczeniach, w których badany był przesłuchiwany lub uczestniczył w innych czynnościach procesowych, takich jak okazanie, pobieranie odcisków itp.⁶¹

Dopuszcza się obecność tłumacza, jeżeli jest to niezbędne dla zapewnienia prawidłowej komunikacji⁶². Choć spisane standardy o tym

⁵⁸ APA [American Polygraph Association], *Standards of Practice*, 23.08.2019, www.polygraph.org/assets/docs/Misc.Docs/standard/adopted%208-23-19%20APA%20STANDARDS%20OF%20PRACTICE%20.pdf [dostęp: 9.06.2020].

⁵⁹ APA, *Model Policy for law enforcement/public-service pre-employment polygraph screening examinations*, apoa.memberclicks.net/assets/docs/le%20screening%20model.pdf [dostęp: 9.06.2020]; ASTM International, *Standard Guide for PDD Examination Standards of Practice. E2062-07*, www.astm.org/Standards/E2062.htm [dostęp: 9.06.2020].

⁶⁰ M. Gołaszewski, M. Widacki, *Aktualny standard badań poligraficznych a praktyka polska*, [w:] J. Widacki (red.), *Badania poligraficzne w Polsce*, Kraków 2014, s. 173.

⁶¹ *Standard badań poligraficznych w sprawach karnych* – uchwalony na nadzwyczajnym zjeździe Stowarzyszenia Poligraferów Polskich w dn. 6 stycznia 2004 r. w Warszawie. Tekst opublikowany w języku angielskim: *Standard for Polygraphic Examinations in Criminal Cases*, „European Polygraph”, 1(1), 2007, s. 65–77.

⁶² Zob. pkt 9 – APA, *Model Policy for Release and Management of Polygraph Reports and Polygraph Data*, apoa.memberclicks.net/assets/docs/records%20and%20data%20model%20policy.pdf [dostęp: 9.06.2020];

nie mówią, uzasadniony jest także ewentualny udział psychologa lub opiekuna prawnego w przypadku badania osoby małoletniej (poza tym członkowie rodziny nie mogą być przy badaniu z uwagi na poruszane zwykle kwestie wrażliwe) czy konwojentów osoby zatrzymanej lub osadzonej w zakładzie karnym w celu zapewnienia bezpieczeństwa i zapobieżenia ucieczce. W porozumieniu ze zleceniodawcą biegły musi to jednak zorganizować w taki sposób, by prowadzenie ekspertyzy nie było utrudnione. Z praktyki biegłych z zakresu badań poligraficznych wiadomo, że zdarzają się okoliczności, w których strona postępowania wyraża stanowisko wskazujące na konieczność udziału w badaniu lekarza specjalisty, np. psychiatry, argumentując to zażywaniem przez osobę badaną określonych leków (np. przeciwdepresyjnych). Jest to stanowisko bezzasadne. O ile biegły z zakresu badań poligraficznych są szkoleni w zakresie ewentualnego wpływu środków farmakologicznych na procesy fizjologiczne, biegły psychiatra nie posiada wiadomości specjalnych w zakresie badań poligraficznych. Co zatem miałby wnieść udział biegłego psychiatry do badania poligraficznego? Byłoby to nieprzydatne, a wręcz szkodliwe.

Badanie poligraficzne jest ekspertyzą, z której sporządza się opinię, a nie czynnością procesową, z której sporządza się protokół. Zatem udział innych osób (np. pełnomocnika strony czy prokuratora) nie jest zagwarantowany przez prawo. Zważywszy na to, że każda dodatkowa osoba mogłaby ujemnie wpływać na przebieg badania, biegły mógłby wyrazić zgodę na obserwację, ale z wykorzystaniem środków technicznych, które to umożliwiają spoza pomieszczenia, w którym znajduje się badany (np. na monitorze lub przez lustro weneckie). Każde badanie powinno być rejestrowane audiowizualnie, więc do dyspozycji zainteresowanych i upoważnionych osób pozostaje też nagranie. Po rozpoczęciu wywiadu przedtestowego nie powinno być już żadnej interakcji osoby badanej z innymi osobami niż biegły⁶³.

Charakter badań pokazowych musi być wyraźnie zakomunikowany publiczności⁶⁴, przy czym zakazane jest wydawanie opinii odnoszącej się do prawdomówności, ponieważ są to warunki, w których ryzyko błędu pomiaru spowodowane czynnikami środowiskowymi jest ogromne.

por.: R.E. Godby, *Polygraph Examinations With Interpreters*, „Polygraph”, 21(2), 1992, s. 83–91; J.A. Matte, *Forensic Psychophysiology...*, *op. cit.*, s. 549–553; M. Gołaszewski (red.), *Współczesne standardy...*, *op. cit.*, s. 110–114.

⁶³ Zob. pkt 10 – APA, *Model Policy for Release...*, *op. cit.*

⁶⁴ APA, *Standards of Practice...*, *op. cit.*

Trzeba przy tym zaznaczyć, że występowanie pseudobiegłych w telewizyjnych programach rozrywkowych przynosi wiele szkody badaniom poligraficznym i sprowadza się zwykle do zaprezentowania *maszynki do wykrywania kłamstw*.

2.3. Oprzyrządowanie wykorzystywane w badaniach poligraficznych

Kolejnym aspektem rzutującym na poprawność pomiaru w badaniach poligraficznych jest stan przyrządu, czyli poligrafu (analogowego – przestarzałego już lub komputerowego). Urządzenie to powinno rejestrować co najmniej: dwie krzywe pneumograficzne (piersiową i brzusznią); aktywność elektrodermalną, odzwierciedlającą zmiany w przewodności elektrycznej lub oporności naskórka; aktywność sercowo-naczyniową przez zmiany tętna i względnego ciśnienia krwi oraz aktywność motoryczną (ruchy ciała, napięcie mięśni) osoby badanej. Fakultatywnie wykorzystuje się fotopletysmograf, mierzący zmiany w amplitudzie pulsu. Pomocniczo mogą być stosowane także inne czujniki, ale nie wolno ich wykorzystywać przy wydawaniu opinii odnoszącej się do prawdomówności. Warunkiem uwzględnienia dodatkowych parametrów w analizie danych testowych jest ich naukowa walidacja i publikacje na ten temat⁶⁵.

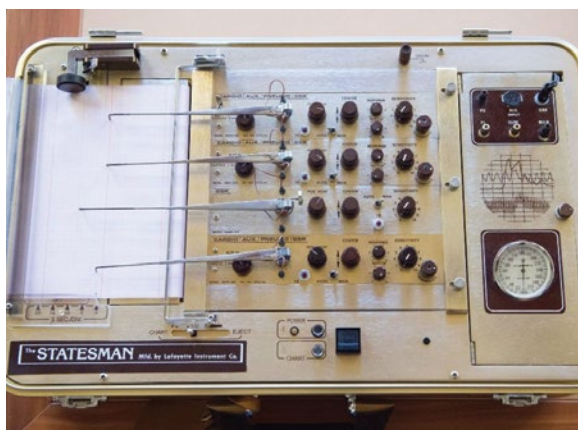
Sprawdzenie poprawności działania urządzenia powinno się odbywać zgodnie z zaleceniami producenta. Poligraf nie podlega prawnemu obowiązkowi homologacji. Jeśli producent nie wskazuje częstotliwości i okoliczności warunkujących konieczność takiego sprawdzenia, należy to robić przy każdym podejrzeniu awarii, a normalnie – raz na pół roku i zachować wykres potwierdzający sprawność sprzętu przez minimum rok⁶⁶. W przypadku poligrafów analogowych czynność sprawdzenia należy przeprowadzać częściej – raz w miesiącu i po każdym transporcie sprzętu. Nie wolno używać poligrafu, którego działanie okazało się wadliwe, dopóki usterka nie zostanie usunięta i potwierdzona testem sprawności. Poligraferzy są uczeni na szkoleniach, w jaki sposób

⁶⁵ ASTM International, *Standard Guide for Instrumentation, Sensors and Operating Software Used in Forensic Psychophysiological Detection of Deception (Polygraph) Examinations*. E2439-09(2016), www.astm.org/Standards/E2439.htm [dostęp: 9.06.2020]; APA, *Standards of Practice...*, *op. cit.*

⁶⁶ APA, *Model Policy for law enforcement...*, *op. cit.*; ASTM International, *Standard Practice for Calibration and Functionality Checks Used in Forensic Psychophysiological Detection of Deception (Polygraph) Examinations*. E2063-12(2017), www.astm.org/Standards/E2063.htm [dostęp: 9.06.2020].

w różnych warunkach dokonać sprawdzenia sprzętu, nawet z wykorzystaniem butelki czy krzesła. Najłatwiej jednak – i zarazem najbardziej profesjonalnie – wykorzystać do tego przeznaczone urządzenie (niegdyś nazywane *kalibratorem*, choć kalibracja dokonywana jest wyłącznie przez producenta na etapie wytwarzania sprzętu) – przykładowe obrazuje ryc. 6.

Ryc. 4. Poligraf analogowy Statesman firmy Lafayette Instrument Co.



Źródło: J. Sullum, 'Lie Detectors' Do Not Detect Lies, „Reason”, 10.03.2018, reason.com/wp-content/uploads/2019/04/15381648125406-856x630.jpg [dostęp: 9.06.2020].

Ryc. 5. Zestaw z poligrafem komputerowym LX6 (producent: Lafayette Instrument Co.)



Źródło: LX6 Polygraph System, lafayettepolygraph.com/products/lx6 [dostęp: 9.06.2020].

Ryc. 6. Urządzenie do testu sprawności poligrafu komputerowego, wytwarzane przez Lafayette Instrument Co.



Źródło: Computerized Polygraph Functionality Check Device, lafayettepolygraph.com/products/computer-polygraph-functionality-check-device [dostęp: 9.06.2020].

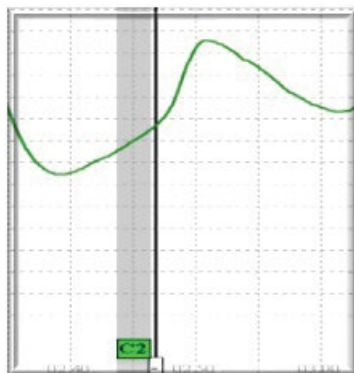
2.4. Pomiary poligrafu, przetwarzanie sygnału i rejestrowanie danych

Tak zwane dane surowe rejestrowane przez poligraf są następnie przetwarzane w programie komputerowym. Ich wizualną reprezentacją są zapisy na poligramach. Aby były czytelne, podlegają mniejszemu lub większemu filtrowaniu, które jest konieczne, aby usunąć szumy niezwiązane z mierzonym parametrem. Od decyzji poligrafera zależy rodzaj filtrowania w kanale reakcji elektrodermalnych. W tym kanale po reakcji fazowej następuje powrót krzywej do linii bazowej (aktywność toniczna), ale gdy trwa odparowywanie potu z naskórka, krzywa ta może opadać nawet poza zakres obszaru na monitorze, a później na wydruku. Ekspert może zapanować nieco nad przebiegiem krzywej przez tzw. centrowanie, czyli przywrócenie biegu krzywej do wybranego poziomu na wykresie. Szczególnie początkującym poligraferom nie jest łatwo wykonywać złożone czynności jednocześnie, dlatego wprowadzone są jeszcze dwa tryby przetwarzania sygnału z kanału EDA. Jeden nazywamy z j. ang. *detrended* (hamujący trend spadkowy krzywej), drugi – *automatic* (automatyczny). Pierwszy powoduje jedynie zatrzymanie spadku krzywej EDA na ustalonym poziomie. Poniżej takiej linii bazowej zapis już nie opadnie. Inne elementy krzywej

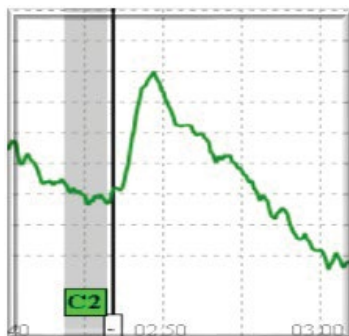
nie są już w tym trybie dodatkowo filtrowane. Z kolei tryb automatyczny polega na bardziej zaawansowanym filtrowaniu – krzywa jest smuklejsza, czytelniejsza i nie wykracza poza obszar poligramu obserwowany na monitorze.

Okazuje się, niestety, że tryb automatyczny (chętnie stosowany przez wielu praktyków) ma bardzo poważne wady. Na przełomie 2012 i 2013 r. w niektórych sytuacjach krzywa filtrowana w trybie automatycznym była tak zaprezentowana na poligramie, że wynikało z niej, iż początek reakcji fazowej nastąpił, zanim wprowadzono bodziec, tzn. wzrost amplitudy od linii bazowej widać jeszcze przed strefą zadawanego pytania (zob. na ryc. 7 – szary pionowy pas na wykresie obejmuje czas zadawania pytania oznaczonego jako C2, a zielona krzywa to zapis przebiegu zmian w przewodnictwie elektrycznym skóry). Jednocześnie reprezentacja tego samego pomiaru bez zaawansowanego filtrowania pokazuje, że reakcja fazowa nastąpiła w odpowiednim czasie i ma związek z bodźcem, nie jest przedwczesna (zob. ryc. 8). Z kolei inny rodzaj filtrowania – *detrended* – nie zniekształcił istoty przebiegu krzywej (ryc. 9).

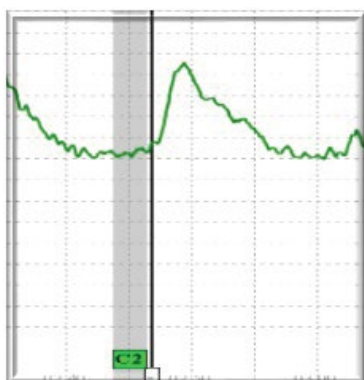
Ryc. 7. Zapis reakcji w kanale EDA w trybie *automatic* w programie LX Software v. 11.1.5



Źródło: materiał autora.

Ryc. 8. Zapis reakcji w kanale EDA w trybie *manual* w programie LX Software v. 11.1.5

Źródło: materiał autora.

Ryc. 9. Zapis reakcji w kanale EDA w trybie *detrended* w programie LX Software v. 11.1.5

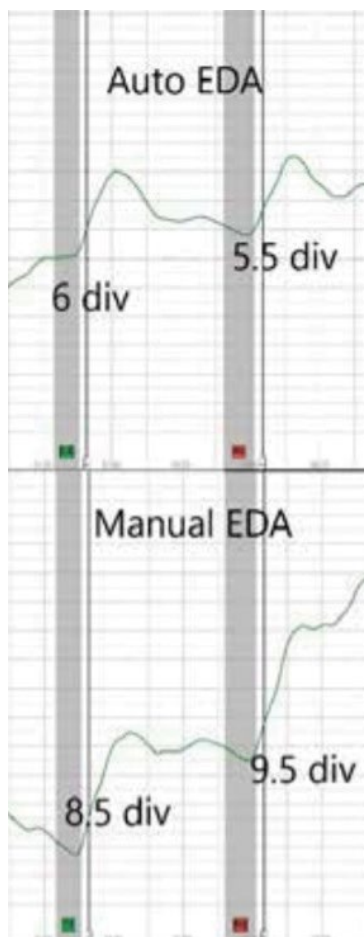
Źródło: materiał autora.

Z przeglądu piśmiennictwa wynika, że problemy związane z filtrowaniem występowały już w latach 90., ale wówczas badania poligraficzne z wykorzystaniem komputerów dopiero się rozwijały i powstawały kolejne, doskonalsze wersje oprogramowania⁶⁷. Choć opisane powyżej spostrzeżenia nastąpiły kilka lat temu, a programiści zapewniali w późniejszym czasie o właściwym skorygowaniu filtrowania w kanale EDA – wygląda na to, że problem dalej istnieje. Świadczy o tym publikacja na ten temat z 2019 r. w wydawnictwie APA, w której autorzy podnosili to samo i – co więcej – zauważyli jeszcze inny niepokojący

⁶⁷ Zob. J.A. Matte, *Examination and Cross-Examination of Experts in Forensic Psychophysiology. Using The Polygraph*, New York 2000, s. 244–246.

sposób przekształcenia krzywej – taki, w którym relatywna amplituda reakcji może ulec tak istotnej zmianie, że w testach pytań porównawczych odwróci punktację dla tego parametru (ryc. 10). Przyczyna bierze się stąd, że filtr automatyczny działa silniej na wolniejsze reakcje fazowe niż te, które zachodzą gwałtowniej, ponieważ pozostaje więcej czasu, aby filtr wywarł swój wpływ⁶⁸.

Ryc. 10. Porównanie zapisów reakcji w kanale EDA dla dwóch trybów przetwarzania: *manualnego* i *automatycznego*



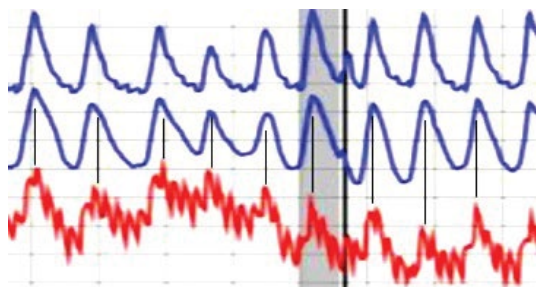
Źródło: D.J. Krapohl, K. Nix, *Evidence and Common Sense...*, op. cit., s. 58.

⁶⁸ D.J. Krapohl, K. Nix, *Evidence and Common Sense. Suggestions for Scoring Electrodermal Responses*, „APA Magazine”, 52(1), 2019, s. 58–59.

Na powyższej rycinie widać, że w trybie manualnym amplituda w kanale EDA jest wyższa przy bodźcu po prawej stronie (krytycznym) niż po lewej (porównawczym). Natomiast w trybie automatycznym zapis reakcji został tak przekształcony, że powolniejsza zmiana przy pytaniu z prawej strony wygląda na mniejszą pod względem amplitudy, a więc głównego (i w niektórych systemach analizy danych jedynego) kryterium diagnostycznego w tym parametrze. Zjawisko tym istotniejsze, że wkład kanału EDA w wartość decyzyjną na tle pozostałych parametrów jest podwójny.

Dokładność pomiaru w badaniach poligraficznych zależy również od poligrafera i jego zdolności do odpowiedniego posługiwania się aparaturą. Jednym z aspektów, który należy wskazać, jest poprawność zamocowania czujników. Jeśli nie będą założone, tak jak powinny – pomiar nie odzwierciedli tego, co ma być mierzone, dane mogą być niestabilne lub mechanicznie zakłócone, szczególnie gdy chodzi o ułożenie pneumografów. Przykładem nieprawidłowego pomiaru jest takie założenie mankietu kardio, które sprawia, że wierzchołki zapisów za każdym razem będą pokrywały się z szczytem amplitudy w kanale pneumo (ryc. 11).

Ryc. 11. Fragment poligramu ilustrujący źle założony mankiet do pomiaru ciśnienia (wierzchołki zapisu respiracji pokrywają się z wierzchołkami zapisu kardio)



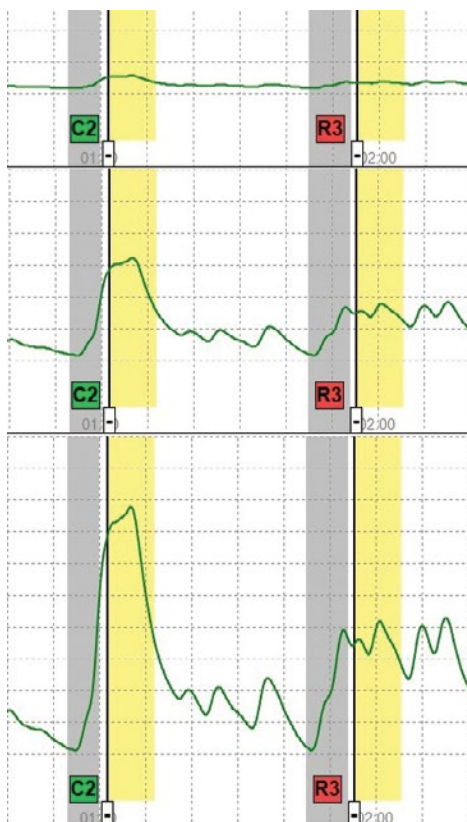
Źródło: materiał autora.

Zapisy reakcji fizjologicznych powinny być ciągłe i mieć ustawioną odpowiednią amplitudę, aby można je było bez trudu odczytać przez badającego i kontrolerów jakości⁶⁹. Minimalny rozmiar amplitudy pneumo i kardio wyznacza się między 0,5 a jednym centym (ok. 1,25 cm – 2,5 cm). W przypadku poligrafów analogowych źle dobrane

⁶⁹ APA, *Standards of Practice...*, op. cit.

czułości powodowały, że zadrukowany wykres bywał bezwartościowy. Możliwa była również miejscowa zmiana czułości w trakcie zapisu. W poligrafach komputerowych istnieje możliwość regulacji czułości również *post factum*, ale obecnie tylko w taki sposób, aby zapobiec manipulowaniu zapisami pojedynczych reakcji (i wpływać przez to na wynik badania). Jeśli zmienia się czułość – dotyczy to całej krzywej na wykresie, a proporcje między amplitudami reakcji pozostają zachowane. W dalszym ciągu jednak ostrożnego podejścia wymaga praca wyłącznie na wydrukach, ponieważ zredukowanie czułości do minimum może tak zniwelować zmiany w przebiegu krzywych, że będzie się wydawało, iż nie wystąpiły żadne istotne reakcje (zob. ryc. 12 – są to te same zapisy, ale przedstawione przy ustawieniu różnych czułości).

Ryc. 12. Zapis reakcji w kanale EDA przy ustawieniu różnych czułości (od minimalnej – 0,1 jednostki; do standardowej – 2,5 jednostki)



Źródło: materiał autora.

Rola poligrafera jest kluczowa w dokonywaniu oznaczeń na poliogramach. Dla zakwalifikowania danej zmiany w reakcjach jako spowodowanej wprowadzonym bodźcem (pytaniem) istotny jest moment, w którym ta zmiana wystąpiła. W kanałach pneumo i kardio bierze się pod uwagę te reakcje, które rozpoczęły się od momentu wprowadzenia bodźca. W przypadku kanału EDA należy uwzględnić opóźnienie trwające zaledwie 0,5 sekundy⁷⁰, a w kanale reakcji naczynioruchowych 2 sekundy. Początek bodźca oznacza badający, a nie *maszyna*. Jego refleks (dobry lub słaby) może więc rzutować na to, czy określona zmiana reakcji będzie zaliczona do oceny czy nie. Pół sekundy to przecież mgnienie oka. Kolejne istotne fragmenty poliogramów dotyczą momentów udzielania odpowiedzi przez osobę badaną, bowiem za reakcję związaną z bodźcem uznaje się zwykle tylko taką, która zaistnieje najpóźniej 5 sekund po odpowiedzi. To również oznacza poligrafer.

Różnice indywidualne w tym, jak sprawnie biegły reaguje na bodziec słuchowy, a potem naciska przycisk klawiatury komputerowej, należy więc traktować jako jedną z przyczyn różnic w dokładności pomiarów poligrafu. Producenci sprzętu starają się wychodzić naprzeciw wyzwaniom tego typu i np. w oprogramowaniu LX Software istnieje możliwość ustawienia automatycznego dodatkowego oznaczenia na wykresie wraz z wychwyceniem przez komputer odpowiedzi osoby badanej.

Na poprawność późniejszej analizy danych rzutują także przypisy na poliogramach dokonywane (lub zaniechane) przez badającego, a odnoszące się do wszelkich zjawisk, które zaistnieją w trakcie testu i nie wynikają z naturalnych, niewolitionalnych reakcji organizmu osoby badanej. Chodzi tu o konieczność pominięcia w ewaluacji, niezaliczenia do potencjalnych przesłanek diagnozowania (z wyjątkiem intencjonalnych prób zakłóceń) tzw. artefaktów. Do dokonywania przypisów służą skróty wywodzące się od angielskich słów, opisujące dane zdarzenie, np. OS (z j. ang. *outside stimulus* – bodziec zewnętrzny, np. hałas), DB (z j. ang. *deep breath* – głęboki oddech), Mv (z j. ang. *movement* – ruch), EI (z j. ang. *examiner's instruction* – instrukcja badającego), C (z j. ang. *cough* – kaszel) i wiele innych.

⁷⁰ Osobnym problemem jest to, że ten czas latencji wynoszący 0,5 sekundy wydaje się za krótki, ponieważ normalnie badany nie jest w stanie jeszcze rozpoznać znaczenia pytania, zanim usłyszy kilka pierwszych słów – co zajmuje realnie ok. 2 sekund. Zob. D.J. Krapohl, K. Nix, *Evidence and Common Sense...*, *op. cit.*, s. 59–60.

3. Zakres niezbędnych wiadomości specjalnych biegłego z zakresu badań poligraficznych

3.1. Weryfikacja wiarygodności biegłego z zakresu badań poligraficznych jako źródła dowodowego

Konieczność powołania biegłego zachodzi, gdy ustalenie pewnych faktów wymaga wiadomości specjalnych, którymi nie dysponuje organ procesowy (o czym mowa w art. 193 k.p.k.). Przez wiadomości specjalne rozumie się „okoliczności przekraczające poziom wiedzy człowieka wykształconego o odpowiednim doświadczeniu życiowym”⁷¹, przy czym nie zalicza się do nich wiedzy prawniczej, jako że organ procesowy z założenia powinien taką posiadać⁷².

Zgodnie z art. 195 k.p.k. biegłym może być każda osoba, o której wiadomo, że ma odpowiednią wiedzę w danej dziedzinie. Co więc oznacza ta *odpowiednia wiedza* w odniesieniu do badań poligraficznych? Tego już kodeks nie precyzuje, a więc ocena kwalifikacji biegłego pozostaje w gestii prezesa sądu okręgowego, wpisującego na listę biegłych, i organu procesowego, który swoim postanowieniem powołuje biegłego.

W 1972 r. Sąd dla Okręgu Południowego w Kalifornii w sprawie *United States v. DeBetham*⁷³ – rozpatrując kwestię dopuszczalności dowodu z opinii z zakresu badań poligraficznych – uznał, w ślad za J. Reidem i F. Inbaudem⁷⁴, że poligrafer powinien mieć wyższe wykształcenie; przynajmniej sześć miesięcy stażu szkoleniowego pod kierunkiem kompetentnego badającego lub badających, z uwzględnieniem wystarczającej ilości praktyki, pozwalającej na pozostawanie pod częstą superwizją w warunkach rzeczywistych spraw. Od biegłego wymaga się, aby spełnił warunek przynajmniej pięcioletniego doświadczenia jako specjalista w zakresie badań poligraficznych. Jego opinia powinna opierać się na zapisach z badania poligraficznego, które przeprowadził. Wreszcie konieczne jest jej przedstawienie w sądzie i umożliwienie weryfikacji przez strony. Poziom wyszkolenia biegłego, jego

⁷¹ K.T. Boratyńska, *Dział V. Dowody*, [w:] A. Sakowicz (red.), *Kodeks postępowania karnego. Komentarz*, wyd. 7, Warszawa 2016, s. 522.

⁷² Zob. wyrok SN z dnia 12 listopada 1973 r., II KR 285/72.

⁷³ 348 F. Supp. 1377 (S.D. Cal. 1972).

⁷⁴ J. Reid, F. Inbau, *Truth and Deception...*, *op. cit.*

wiedza i sposób przeprowadzenia badań będą pozwalały na ocenę, na ile opinia może odpowiadać stanowi faktycznemu⁷⁵.

W momencie, w którym poligrafer decyduje się zostać biegłym, musi się liczyć z tym, że jego kwalifikacje mogą być zweryfikowane wraz z wezwaniem do sądu i przesłuchaniem na okoliczność wydanej opinii. Trzeba być przygotowanym na pytania o ogólne wykształcenie, specjalistyczne przeszkolenie, doświadczenie zawodowe itp.

Na początku lat 70. przypominał o tym poligraferom C. Sevilla, kierujący w San Diego zespołem adwokackim – tym samym, który reprezentował B. DeBethama z przytoczonej wyżej sprawy¹⁾. Wieszczył przy okazji, że w niedługim czasie (a było to przecież jeszcze w okresie, gdy obowiązywała restrykcyjna zasada *general acceptance*) nastąpi przełom i sądy zaczną dopuszczać dowód z badań poligraficznych. Wtedy – w sprawie DeBethama – tak się nie stało, choć orzekający sędzia w znacznej mierze przychylnie się do tego odnosił. Ostatecznie musiał uwzględnić inne postanowienia sądów z tego samego okręgu, które poligraf odrzucały.

Natomiast równoległe zapadły już orzeczenia dopuszczające dowód z opinii biegłego z zakresu badań poligraficznych. Na przykład w sprawie *United States v. Ridling*²⁾ Sąd Okręgowy w Michigan ocenił, że dowód z przesłuchań biegłych uwzględniał podstawową teorię badań poligraficznych, zaufanie do poligrafu ze strony instytucji rządowych i sektora prywatnego, porównanie rzetelności poligrafu z innymi dowodami naukowymi – jak odciski palców i badania balistyczne, a także opinie biegłych na temat tego, czy dowód ten będzie wartościowy w kontekście ustalenia kwestii rozpatrywanych przez sąd w tej sprawie i w wymiarze sprawiedliwości. Dowód został dopuszczony pod pewnymi warunkami. Aby zapewnić obiektywizm, sąd oczekiwał od stron wspólnego uzgodnienia propozycji trzech biegłych – innych, niż zaoferował oskarżony. Sąd zastrzegł sobie decyzję o wyborze co najmniej jednego z nich. Oskarżony miał stawić się na badania w ustalonym czasie. Biegły z kolei – przedstawić wyniki badania sądowi, ławie przysięgłych i obu stronom postępowania. Ponadto sąd zapowiedział pominięcie dowodu, jeśli oskarżony zrezygnuje z udziału w badaniu lub nie będzie współpracował z biegłym, a także w przypadku nierozstrzygniętego rezultatu testu.

¹⁾ C. Sevilla, *The Polygraph Examiner as a Witness in Court*, „Polygraph”, 2(2), 1973, s. 122–132.

²⁾ *United States v. Ridling*, 350 F. Supp. 90 (E.D. Mich. 1972).

⁷⁵ Por. Ch.M. Cook, *The Role and Rights of the Expert Witness*, „Journal of Forensic Sciences”, 9(4), 1964, s. 459.

Polskie sądy, na zasadzie swobodnej oceny, sprawdzając kompetencje biegłego, nie są związane takimi kryteriami, jakie w USA wynikały z orzecznictwa tamtejszych sądów (reguła Frye'a, a potem Dauberta). W dodatku, gdy w konkretnej sprawie powołana zostaje osoba z listy biegłych, domniemuje się, że posiada wiadomości specjalne, ponieważ przeszła procedurę ustanowienia biegłym przy konkretnym sądzie okręgowym i prezes tego sądu uznał, że spełnia odpowiednie warunki. W rzeczywistości na listach biegłych można spotkać poligraferów o wątpliwych kwalifikacjach. Z drugiej strony, prezesi sądów nie mają w tych procedurach odpowiednich kryteriów i skutecznych instrumentów sprawdzenia kandydata na biegłego. Pisze o tym m.in. S. Waltoś:

Wpisanie na listę biegłych sądowych nie gwarantuje, że biegły wyda opinię zgodną z najwyższymi standardami nauki. Przed wpisaniem na listę nie przeprowadza się żadnych rozmów, nie sprawdza się kandydata. Wystarczą dokumenty poświadczające kwalifikacje. Wydawałoby się zatem, że między listą biegłych sądowych znajdującą się w sądzie okręgowym a zbiorem biegłych umieszczonych i reklamujących się w Internecie, dostępnym dzięki stosownym wyszukiwarkom, nie ma żadnej różnicy. Taki wniosek jest jednak mylny. W przeciwieństwie do Internetu na liście biegłych sądowych znajdują się nazwiska tych, których kwalifikacje zostały wstępnie sprawdzone⁷⁶.

Poniekąd wracając do punktu wyjścia, wydaje się jednak konieczne, aby w każdej sprawie z osobna przyjrzeć się biegłemu z zakresu badań poligraficznych – już nie tyle w sensie dopuszczalności samej metody, bo ta jest od dawna powszechnie akceptowana, ale na pewno jeżeli chodzi o tego, kto tą metodą się posługuje⁷⁷.

Biegły powinien w pierwszej kolejności opierać się na tym, co zostało w jego dziedzinie udowodnione naukowo i opublikowane, oraz

⁷⁶ S. Waltoś, *Opinie biegłych – na ostrym zakręcie w polskim procesie karnym*, [w:] J. Czapska, A. Okrasa (red.), *Bezpieczeństwo – policja – kryminalistyka. W poszukiwaniu wiedzy przydatnej w praktyce*, Kraków 2015, s. 33.

⁷⁷ Por. J. Widacki, *Obrońca wobec opinii biegłego w procesie karnym*, „Studia Prawnicze. Rozprawy i Materiały”, 1(16), 2015, s. 5–20; Fundacja Europejskie Centrum Inicjatyw w Naukach Sądowych, *Ocena kompetencji biegłych sądowych. Oczekiwania i rekomendacje*, Warszawa 2015, forensicwatch.pl/web/pliki/baza-wiedzy/Opracowania/Ocena-Kompetencji-Bieglych-Sadowych.pdf [dostęp: 18.05.2020].

postępować zgodnie z powszechnie uznanymi, spisаныmi, oficjalnie ogłoszonymi standardami profesjonalnymi. Nie trzeba przy tym przekreślać możliwości posiłkowania się także materiałami niepublikowanymi czy wiedzą pochodzącą od innych biegłych tej samej specjalności. Punktem wyjścia dla zapewnienia prawidłowego toku postępowania poligrafera są jego kwalifikacje merytoryczne. Powierzenie zadania biegłemu, który jest nim nie tylko z nazwy, gwarantuje odpowiednią jakość wykonania zadania. Jakość (zgodność postępowania z regułami sztuki, rzetelność i trafność opiniowania) nie zależy istotnie od sprzętu – wykrywacza kłamstw – jak się potocznie sądzi, lecz od osoby realizującej ekspertyzę. Do tego są potrzebne: wiedza, doświadczenie z praktyki zawodowej, predyspozycje (o czym szerzej w rozdziale 4) oraz odpowiedni poziom wykształcenia.

W Polsce dużym zagrożeniem jest działalność podmiotów, których wiarygodność i przygotowanie do prowadzenia badań poligraficznych budzi uzasadnione wątpliwości. Bywa, że w nazwie takiego podmiotu znajdują się człony stwarzające pozory powagi i naukowości, jak np. *instytut*, a zlecniodawcy nie zadają sobie większego trudu w potwierdzeniu kwalifikacji osoby desygnowanej do badań⁷⁸. Pierwsze poważne pełnowymiarowe przeszkolenie specjalistyczne z zakresu badań poligraficznych zostało w Polsce zorganizowane w 2008 r. Obejmowało wówczas 320 godzin zajęć i odbyło się z udziałem ekspertów z USA na potrzeby Agencji Bezpieczeństwa Wewnętrznego. Szkolenie zakończyło się egzaminem i wydaniem certyfikatów uznanych przez APA. Do tego momentu najwyżej pojedyncze osoby miały możliwość nabycia kwalifikacji zawodowych inaczej niż drogą uczeń–mistrz w ramach działalności różnych służb mundurowych.

⁷⁸ Por. J. Widacki, *Glosa do wyroku Sądu Apelacyjnego w Katowicach z 7 czerwca 2017 r., II AKA 167/17*, „Palestra”, 10, 2017, s. 74–77; *idem*, *Instytucja naukowa lub specjalistyczna w rozumieniu art. 193 § 2 k.p.k.*, „Państwo i Prawo”, 9, 2013, s. 45–53.

W 2010 r. w Centralnym Ośrodku Szkolenia Agencji Bezpieczeństwa Wewnętrznego w Emowie odbyło się pierwsze w Polsce międzynarodowe sympozjum poligraferów, które zapoczątkowało proces ujednoczenia standardów badań poligraficznych w naszym kraju. W kolejnych latach kontakty polskich poligraferów z uznanymi ekspertami z zagranicy były intensyfikowane, prowadzono regularne szkolenia doskonalenia zawodowego dla oficerów Agencji Bezpieczeństwa Wewnętrznego, a później podobny kierunek obrały również inne instytucje państwowe. Pracownie badań poligraficznych zostały wyposażone w nowoczesny sprzęt. Krajowym forum wymiany doświadczeń były cyklicznie organizowane przez Żandarmerię Wojskową seminaria międzyresortowe, a na szczeblu międzynarodowym – doroczne seminaria APA.

W 2013 r. reaktywowano formę zrzeszenia się profesjonalistów tej specjalności i naukowców w Polsce. W miejsce Stowarzyszenia Poligraferów Polskich, które zakończyło działalność, ustanowiono dwie organizacje o podobnym profilu: Polskie Towarzystwo Badań Poligraficznych i Polskie Stowarzyszenie Poligraferów. W 2016 r. PTBP uzyskało status instytucji afiliowanej przy APA. W tym czasie eksperci Towarzystwa podjęli współpracę organizacyjno-szkoleniową z Uniwersytetem Kardynała Stefana Wyszyńskiego w Warszawie i zainicjowali regularne szkolenia specjalistyczne z zakresu badań poligraficznych o charakterze otwartym, zbieżne programowo ze standardami APA.

3.2. Zdefiniowanie kompetencji poligrafera w standardach profesjonalnych, standardy kształcenia

W przypadku APA już sam fakt członkostwa w kategorii *Associate Member* i *Full Member* świadczy o formalnym przeszkoleniu poligrafera, ponieważ jest to jeden z warunków przynależności do stowarzyszenia⁷⁹. Pierwszą kategorię mogą otrzymać osoby, które ukończyły kurs podstawowy z programem spełniającym normy akredytacyjne APA oraz zapoznały się z kodeksem etyki i standardami praktyki APA. Natomiast kandydat do pełnego członkostwa powinien, poza spełnieniem powyższych kryteriów, legitymować się co najmniej rocznym stażem w tej organizacji, wykształceniem wyższym licencjackim,

⁷⁹ APA, *By-laws*, 27.08.2019, www.polygraph.org/assets/docs/Misc.Docs/BY-LAWS%20adopted%208-27-19.pdf [dostęp: 10.06.2020].

przeprowadzeniem co najmniej 200 badań z wykorzystaniem zwalidowanych technik badawczych oraz zaliczeniem 60 godzin doskonalenia zawodowego (włącznie z uczestnictwem w przynajmniej jednym rocznym seminarium APA jako *associate*). Jest jeszcze jedna istotna kategoria członkostwa w APA – *science and technology* – przewidziana dla osób, które działają na rzecz społeczności poligraferów przez badania naukowe i doskonalenie oprzyrządowania.

Standardy akredytacyjne szkoleń specjalistycznych wydane przez APA określają zarówno wymogi dla placówek szkoleniowych, jak i ramowy program zajęć, który – zgodnie ze stanem na 2022 r. – obejmował 400 godzin nauki⁸⁰. Poza egzaminem końcowym słuchacz musi zostać sprawdzony co najmniej trzykrotnie na zaliczeniach częściowych partii materiału.

W tym programie powinny się znaleźć następujące zagadnienia:

- regulacje prawne odnoszące się do badań poligraficznych oraz prawa człowieka – 8 godzin;
- standardy praktyki APA, przepisy wykonawcze i etyka – 4 godziny;
- wstęp do historii i ewolucji psychofizjologicznej detekcji nieszczerości – 8 godzin;
- naukowy wymiar testów – 8 godzin; obsługa sprzętu do badań poligraficznych – 16 godzin;
- techniki badań poligraficznych – 40 godzin;
- wywiad przedtestowy – 32 godziny;
- rozmowa po testach – 8 godzin;
- psychologia – 20 godzin;
- fizjologia – 20 godzin;
- konstrukcja pytań testowych – 32 godziny;
- analiza danych testowych – 40 godzin;
- środki zakłócające – 8 godzin;
- praktyczne zastosowanie poligrafu (badania dotyczące zainscenizowanych zdarzeń) – 80 godzin;
- inne wybrane zagadnienia – 74 godziny.

Od praktykujących poligraferów (przez co rozumie się osoby prowadzące badania, szkolenia lub sprawujące kontrolę jakości) wymaga się udziału w doskonaleniu zawodowym w wymiarze co najmniej

⁸⁰ APA, *Education Accreditation Committee. Accreditation Standards (version 2.05)*, 23.09.2019, www.apapolygraph.org/accredited-programs [dostęp: 23.04.2021].

30-godzinnym co dwa lata. Udokumentowanie spełnienia tego wymogu i przechowywanie dokumentów to potwierdzających spoczywa na poligraferze. Tematyka takiego doskonalenia została sformułowana ogólnie jako obszar związany z badaniami poligraficznymi⁸¹. Mogą to być zatem kursy, konferencje naukowe, warsztaty. Trzeba natomiast z wyprzedzeniem zwrócić się do Biura Krajowego APA o uznanie programu takiego przedsięwzięcia i zaliczenie na poczet godzin nauki.

Odrębnie potraktowano poligraferów zajmujących się specyficznym typem badań, a mianowicie dotyczącym sprawców przestępstw przeciwko wolności seksualnej (PCSOT – z j. ang. *post-conviction sex offenders testing*)⁸². Dla tej grupy przewidziano dodatkowe, 40-godzinne przeszkolenie oraz program stażowy w postaci 10 badań pod nadzorem poligrafera uprawnionego do badań typu PCSOT.

Wracając na grunt polski, warto wspomnieć, że pierwszą procedurę certyfikacji biegłych z zakresu badań poligraficznych wprowadziło Polskie Towarzystwo Badań Poligraficznych, podejmując stosowną uchwałę na zjeździe założycielskim (2013 r.)⁸³. Procedura ta została podana do publicznej wiadomości, podobnie jak standard kursów i szkoleń (2016 r.)⁸⁴. W 2017 r. wewnętrzną certyfikację wprowadziło Polskie Stowarzyszenie Poligraferów, lecz ani procedura, ani kryteria dla kandydatów nie zostały w tym wypadku ogłoszone.

⁸¹ APA, *APA Continuing Education Hours Policy*, www.polygraph.org/policy-for-apa-continuing-education-hours [dostęp: 10.06.2020]; zob. też – ASTM International, *Standard Guide for Minimum Continuing Education and Training of Individuals Involved in the Detection of Deception (PDD): E2064-00(2014)*, www.astm.org/Standards/E2064.htm [dostęp: 10.06.2020].

⁸² APA, *PCSOT Committee Operational Policy*, April 2016, www.polygraph.org/assets/docs/Misc.Docs/pcsot%20operational%20policy%20adopted%205-17-16.pdf [dostęp: 8.06.2020].

⁸³ Polskie Towarzystwo Badań Poligraficznych, *Uchwała Nr 6/2013 Zebrań Założycielskiego Polskiego Towarzystwa Badań Poligraficznych z dnia 24 listopada 2013 r. w sprawie procedury wpisu na listę rekomendowanych przez Towarzystwo ekspertów z zakresu badań poligraficznych*, ptbp.pl/wp-content/uploads/Dokumenty/Uchwała_w_sprawie_procedury_wpisu_na_liste_rekomendowanych_przez_Towarzystwo_ekspertow.pdf [dostęp: 7.06.2023].

⁸⁴ Polskie Towarzystwo Badań Poligraficznych, *Uchwała nr 7/2016 Zarządu Polskiego Towarzystwa Badań Poligraficznych z dnia 20.07.2016r.: Standard kursów i szkoleń z zakresu badań poligraficznych (z późn. zm.)*, ptbp.pl/wp-content/uploads/Dokumenty/PTBP-standard_kursow_i_szkolen_-_tekst_jednolity_2022.pdf [dostęp: 10.08.2022].

Procedura udzielania rekomendacji eksperckich przez PTBP dotyczy trzech typów badań poligraficznych: w celach procesowych, pozaprocessowych i terapeutycznych. Kandydat nie musi być członkiem stowarzyszenia. Warunki merytoryczne zostały określone przez Komitet Naukowy Towarzystwa⁸⁵. Rekomendację do badań w celach procesowych i pozaprocessowych może uzyskać osoba, która:

- a) posiada wykształcenie wyższe, magisterskie;
- b) ukończyła kurs specjalistyczny i uzyskała stosowny certyfikat autoryzowany lub uznany przez APA, albo ukończyła równorzędny kurs organizowany lub autoryzowany przez PTBP, albo posiada niekwestionowane, nabyte w innej drodze umiejętności prawidłowego przeprowadzania badania poligraficznego i interpretacji jego wyników;
- c) wykazuje nienaganną postawę moralną;
- d) zobowiązuje się do przestrzegania kodeksu etyki poligrafera obowiązującego w PTBP.

Z kolei w przypadku badań terapeutycznych należy ukończyć specjalistyczne przeszkolenie w zakresie badań poligraficznych w dziedzinie, której dotyczy terapia oraz legitymować się ukończonymi studiami psychologicznymi lub lekarskimi i posiadać uprawnienia psychoterapeuty albo przejść specjalistyczne przeszkolenie (na poczet którego może zostać uznane także zaliczenie na uczelni wyższej wykładu, konwersatorium lub ćwiczeń: z podstaw psychologii, z psychologii klinicznej lub psychologii sądowej, a nadto z psychiatrii sądowej oraz seksuologii klinicznej lub seksuologii sądowej o łącznym wymiarze 120 godzin).

W celu przedłużenia rekomendacji wymagane jest ukończenie kursu doskonalenia zawodowego autoryzowanego przez APA lub PTBP. Zamiast ukończenia takiego kursu akceptowane są: wykonanie badań eksperymentalnych, których wyniki opublikowano w czasopiśmie naukowym, a nadto czynny udział w warsztatach, konferencjach i seminariach naukowych.

⁸⁵ Polskie Towarzystwo Badań Poligraficznych, *Uchwała nr 1/2014 Komitetu Naukowego Polskiego Towarzystwa Badań Poligraficznych z dnia 17 marca 2014 r. w sprawie określenia warunków wpisu na listę rekomendowanych przez Towarzystwo ekspertów z zakresu badań poligraficznych*, ptbp.org.pl/Uchwała%20KN%20PTBP%20w%20sprawie%20warunkow%20wpisu%20na%20liste%20ekspertow.pdf [dostęp: 10.06.2020].

Standard szkolenia przyjęty przez PTBP przewiduje kilka wariantów. Pełny kurs specjalistyczny ma wymiar od 320 do 400 godzin zajęć teoretycznych i praktycznych. Tematyka jest spójna z programem APA, a w praktyce program poszerza się m.in. o kwestie związane ze sporządzaniem pisemnych opinii czy elementy statystyki. Inne warianty stwarzają warunki do rozłożenia procesu kształcenia na dwa etapy: kurs podstawowy (I stopnia) i zaawansowane szkolenie uzupełniające (oba obejmujące od 160 do 200 godzin zajęć). Przewidziano także szkolenia doskonalenia zawodowego oraz przeznaczone dla przedstawicieli organów ścigania oraz wymiaru sprawiedliwości na temat istoty, użyteczności i uwarunkowań prawnych badań poligraficznych.

Mając powyższe na uwadze, należy przyjąć, że miarodajnym potwierdzeniem kwalifikacji biegłego z zakresu badań poligraficznych są certyfikaty wydawane przez podmioty akredytowane przez APA, Polskie Towarzystwo Badań Poligraficznych⁸⁶, Polskie Stowarzyszenie Poligraferów⁸⁷, urzędy państwowe oraz (z zastrzeżeniami nie tyle natury formalnej, co praktycznej) wpis na listę biegłych sądowych. Wiadomości specjalne w zakresie badań poligraficznych powinny więc wynikać przede wszystkim z formalnego specjalistycznego przeszkolenia (w takim wymiarze czasowym i z takim programem merytorycznym, jaki wymagany jest przez wiodące zrzeszenia poligraferów lub jak stanowi prawo, o ile istniałyby stosowne przepisy). Zasadne byłoby przy tym również potwierdzenie doświadczenia z praktyki zawodowej. Szkolenie ma charakter interdyscyplinarny i nie ma przy tym wskazania na jakikolwiek konkretny kierunek wykształcenia akademickiego. Wbrew częstym obiegowym opiniom poligrafer nie musi być jednocześnie psychologiem. Ze względu na pożądane rozwinięte umiejętności interpersonalne preferuje się ogólnie raczej wykształcenie humanistyczno-społeczne niż np. techniczne. Dostrzega się także przydatność wykształcenia prawniczego ze względu na fakt bliskości ekspertyzy poligraficznej i faktu głównego⁸⁸.

⁸⁶ Polskie Towarzystwo Badań Poligraficznych, *Lista rekomendowanych ekspertów*, ptbp.pl/lista-ekspertow [dostęp: 10.06.2020].

⁸⁷ Polskie Stowarzyszenie Poligraferów, *Certyfikowani poligraferzy Polskiego Stowarzyszenia Poligraferów*, poligrafer.com/certyfikowani-poligraferzy.html [dostęp: 10.06.2020].

⁸⁸ Zob. R. Jaworski, *Opinia z ekspertyzy poligraficznej jako dowód odciążający*, Wrocław 1999, s. 92–93.

Teoretycznie, przy założeniu równego poziomu formalnego wykształcenia, biegli powinni te same dane interpretować jednakowo, a jednak wydają niekiedy opinie rozbieżne. K. Jaegermann i Z. Marek wyróżnili trzy grupy przyczyn owych rozbieżności na przykładzie opinii sądowo-lekarskich⁸⁹. Pierwsza grupa obejmuje różnice w interpretacji obserwacji i pomiarów. Druga – sprzeczną ocenę materiału dotyczącego analizowanego przypadku (w takim sensie, czy w ogóle daje podstawę do wydania opinii i ewentualnie, z jakim poziomem stanowczości). Trzecia natomiast dotyczy odmiennego stosowania zasad ustalania związku przyczynowego między skutkiem biologicznym a działaniem opisanym w aktach sprawy. Autorzy słusznie konstatują, że

dyskusja nad rozbieżnościami w opiniach jest rozważaniem, jak zrealizować równocześnie dwa postulaty: pierwszy to maksymalne wyjaśnienie wątpliwości, drugi to zakaz przekroczenia zasad, zakaz dowolności ocen. Pierwszy postulat wymaga odwagi, drugi ostrożności. Sztuka więc opiniowania to przyjęcie postawy rozważnej, łączącej odwagę z ostrożnością⁹⁰.

4. Znaczenie osobistych właściwości poligrafera

4.1. Umiejętność odpowiedniego zaprezentowania wyników badań

Profesjonalizm biegłego – rozumiany jako dostatecznie wysoki poziom przygotowania merytorycznego i ścisłe postępowanie według reguł sztuki – przekłada się niewątpliwie na jakość wykonania ekspertyzy. W połączeniu z odpowiednimi predyspozycjami psychologicznymi i osobistymi zdolnościami powinien być również odpowiedzią na zagrożenia związane z subiektywnymi aspektami procesu badawczego i wydawania opinii.

Agent specjalny FBI – F. Miller – pisząc w latach 60. na łamach prestiżowego „*Journal of Forensic Sciences*”, położył nacisk na kilka istotnych zadań, które spoczywają na biegłych⁹¹. Wśród nich wskazał dbanie o pozycję zawodową – np. przez zrzeszanie się, bycie na

⁸⁹ Zob. K. Jaegermann, Z. Marek, *Rozbieżności w opiniach sądowo-lekarskich*, „*Archiwum Medycyny Sądowej i Kryminologii*”, 29(4), 1979, s. 249–257.

⁹⁰ *Ibidem*, s. 257.

⁹¹ F.M. Miller, *The Forensic Scientist in Court*, „*Journal of Forensic Sciences*”, 8(3), 1963, s. 323–324.

bieżąc z literaturą i zaangażowanie w badania naukowe. Ponadto biegli muszą być pewni sporządzanych opinii, należyście dowodzić przedstawianych konkluzji, weryfikować przeciwne hipotezy. Bardzo ważne jest unikanie wpływów zewnętrznych na proces opiniowania, bazowanie wyłącznie na ustaleniach wydedukowanych z przeprowadzonych badań, zachowanie bezstronności. Kluczowe, by skupiać się nie tylko na pogoni za winnymi, ale ochronić osoby niewinne, z których – na szczęście – większość nigdy nie trafia na salę rozpraw. Przeważa więc praca biegłych nie w sądach, lecz laboratoriach.

Gdy opinia biegłego jest w sądzie poddawana krytycznej analizie – zwłaszcza przez stronę, która postrzega ją za niekorzystną – biegły zabezpieczy się najlepiej, jeśli przeprowadzi badania w sposób profesjonalny, wnikliwy, a określone wnioski sformułuje tylko wtedy, gdy będą możliwe do obrony, należyście uzasadnione. Niekiedy podczas przesłuchania są stosowane takie zabiegi, aby przymusić biegłego do udzielenia kategorycznej odpowiedzi *tak/nie*. Nie ma jednak żadnego wymogu, aby odpowiadać w sposób nierzetelny, który nie odzwierciedla prawidłowo stanowiska biegłego. Nie należy przy tym spekulować. Jeśli biegły nie jest czegoś pewny, to nie powinien tego zawierać w swojej opinii. Lepiej i wiarygodniej jest czasem odpowiedzieć: *nie wiem*⁹².

W przypadku biegłego zajmującego się badaniami kryminalistycznymi nie wystarczy wykonać dobrą robotę w laboratorium – powiada F. Miller⁹³. Jego ustalenia muszą być jeszcze skutecznie zaprezentowane w sądzie, gdzie będzie zwracana uwaga na staranność przeprowadzonych badań oraz to, jak stoi przy barierce i radzi sobie z krzyżowym ogniem pytań. To – rzecz jasna – wymaga stosownych predyspozycji i przygotowania. Przy skomplikowanej materii, jaką są badania poligraficzne – w odróżnieniu od wielu innych, możliwych do przynajmniej częściowego opanowania dzięki osobistemu doświadczeniu życiowemu – zdolności perswazyjne poligrafera odgrywają rolę nie do przecenienia.

Często wybiera się konkretnego biegłego nie tylko ze względu na jego profesjonalizm, ale także zważywszy na jego obycie na sali sądowej i sposób przełożenia wniosków z badań na papier pisemnej opinii oraz w trakcie ewentualnego ustnego przesłuchania. Te umiejętności są bardzo pomocne w klarownym wyjaśnieniu różnych zawiłości

⁹² Zob. Ch.M. Cook, *The Role and Rights...*, *op. cit.*, s. 457–458.

⁹³ F.M. Miller, *The Forensic Scientist...*, *op. cit.*, s. 315.

metodologicznych czy statystycznej ewaluacji wyników badań, choć sposób zaprezentowania nie wydaje się wystarczającym kryterium do oceny wiarygodności dowodu, a na pewno nie jedynym. Do tego dochodzi jeszcze, zwłaszcza w postępowaniu kontradiktoryjnym, ryzyko identyfikowania się biegłego (świadomie lub nieświadomie) z jedną ze stron postępowania. Nawet gdy tak w istocie nie jest, biegły może być w ten sposób postrzegany. Tak bywa z punktu widzenia obrony, gdy biegły wywodzi się z organów ścigania. Z kolei oskarżyciel zwykle z rezerwą podchodzi do bezstronności autora opinii prywatnej.

Przydatną biegłemu jest, z pewnością, umiejętność zachowania spokoju. Konfrontację z jedną ze stron w sądzie należy potraktować jako niemal nieuniknioną. Dzieje się tak m.in. dlatego, że istnieje złudne przekonanie, iż prawidłowe przeprowadzenie ekspertyzy za każdym razem gwarantuje dokładne wyniki i właściwe odpowiedzi biegłego. W konsekwencji jedynym sposobem, w jaki wydaje się, że można zaatakować taki dowód, jest wykazanie błędów w sztuce albo zatajenia istotnych informacji przez biegłego, podważając jego kompetencje albo uczciwość⁹⁴. Dużo zależy wówczas od postawy sędziego, który przewodniczy składowi orzekającemu. Zwykle biegły jest traktowany nieco inaczej niż świadek. Jeśli nie narazi się czymś sędziemu – ten zadba o to, by ataki strony niezadowolonej miały swoje granice i nie zawaha się np. z wykorzystaniem możliwości uchylania niektórych pytań. Niemniej jednak zachowanie nerwowej odporności może być tak samo pożądane jak posiadanie odpowiednich kwalifikacji merytorycznych czy zdolności dobrego argumentowania.

4.2. Psychologiczne predyspozycje kandydata na poligrafera i rozwój zawodowy

Obok kwalifikacji wynikających z odpowiedniego przeszkolenia kandydata na biegłego z zakresu badań poligraficznych (w wąskim rozumieniu ograniczających się do nabytej wiedzy) oraz doświadczenia zawodowego warto rozważyć kwestię potencjalnego wpływu czynników osobowościowych na jakość pracy takiego biegłego, na skuteczność i dokładność prowadzonych przez niego badań. Przydatne byłoby udzielenie odpowiedzi na pytanie, czy możliwe jest przy tym stworzenie modelu specyficznych kompetencji w procesie selekcji kandydatów

⁹⁴ Zob. B. Robertson, G.A. Vignaux, Ch. Berger, *Interpreting Evidence...*, *op. cit.*, s. 135.

na poligraferów, uwzględniającego zestaw pożądanых cech indywidualnych (w tym stylów funkcjonowania, motywacji i umiejętności). Zagadnienie to powinno być przedmiotem zainteresowania szczególnie tam, gdzie etatowo zatrudnione są osoby zajmujące się badaniami poligraficznymi.

Aktualnie znane są Autorowi trzy opublikowane prace, które odnoszą się do zagadnienia psychologicznych kompetencji w kontekście efektów pracy poligrafera. Wśród nich są opracowania DACA (Defense Academy for Credibility Assessment) z 2007 r. oraz rozprawa doktorska przedłożona na Uniwersytecie Technologicznym Nanyang w Singapurze w 2018 r.

Pierwszy z raportów DACA przedstawiał ocenę optymalnych podejść przy prowadzeniu przesłuchań na podstawie przebiegu rozmów z osobami, które nie przeszły pomyślnie testów z wykorzystaniem poligrafu w rzeczywistych sprawach karnych⁹⁵. Przez wywiady ustrukturyzowane z przesłuchującymi oraz analizy materiałów video wytypowano zachowania odróżniające poligraferów wybitnych od przeciętnych pod względem zdolności w uzyskiwaniu od osób badanych przyznania się do winy.

W badaniu wzięło udział 25 funkcjonariuszy lokalnych organów ścigania z różnych stanów. Ich doświadczenie zawodowe mieściło się w granicach od 3 do 34 lat. Konsultanci realizujący projekt w drodze analizy notatek z wywiadów podsumowali wskaźniki behawioralne, które miały pozytywny wpływ na przebieg przesłuchania (rozmowy potestowej). Opinie na temat uczestniczących w projekcie funkcjonariuszy wyrazili również ich przełożeni. Ci, którzy zostali sklasyfikowani zgodnie – zarówno przez przełożonego, jak i osobę realizującą projekt – wśród czołowych 10% przesłuchujących – utworzyli 12-osobową grupę o *ponadprzeciętnych* kompetencjach, a pozostali trafili do drugiej grupy.

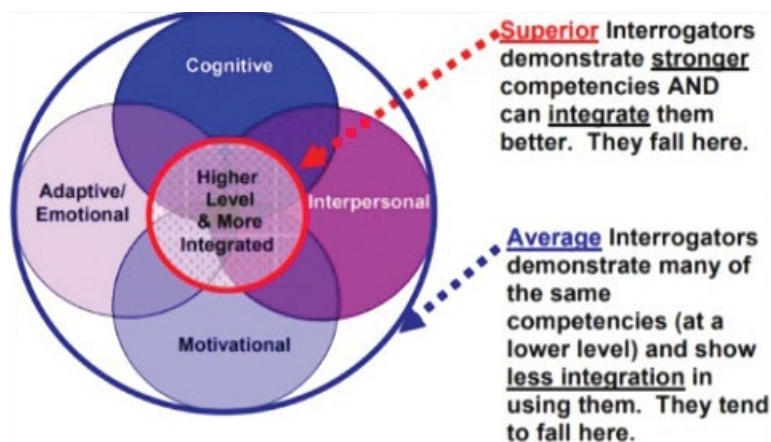
Następnie konsultanci zaprojektowali wstępny model, który został przekazany do analizy czterem sędziom kompetentnym – doświadczonym przesłuchującym z DACA. Eksperti mieli ocenić poszczególne kompetencje na skali od nieistotnych – przez ważne, bardzo ważne, a skończywszy na krytycznie istotnych. Najwyżej punktowano

⁹⁵ Defense Academy for Credibility Assessment, *Assessment of Optimal Interrogation Approaches*, 2007, apps.dtic.mil/sti/pdfs/ADA476369.pdf [dostęp: 27.06.2020].

zdolność przystosowania przesłuchania, wybiegi taktyczne, budowanie zaufania i wytrwałość. Najniżej zaś integrację wiedzy prawniczej z praktyką sądową. Konsultanci z kolei odnieśli 17 wytypowanych rodzajów kompetencji do ogólnych wskaźników sukcesu zawodowego, które bazowały na wynikach różnych badań naukowych zrealizowanych na przestrzeni kilkudziesięciu lat. Kompetencje podzielono na cztery kategorie: poznawcze (zaczynając np. od zbierania danych), interpersonalne (m.in. aktywne słuchanie), motywacyjne (w tym zarządzanie kierunkiem i tempem przesłuchania) oraz adaptacyjne (np. dostosowanie taktyki).

Ustalono, że przeciętni przesłuchujący mogą posiadać część, a nawet większość takich samych umiejętności jak wybitni śledczy, lecz tym, co ich rzeczywiście odróżnia – jest zdolność do zastosowania umiejętności w praktyce. Jednym przychodzi to z łatwością, innym trudniej. Ci lepsi wykazują silniejsze kompetencje i w sposób bardziej zintegrowany. Przeciętni robią to na niższym poziomie i w sposób mniej zintegrowany (zob. ryc. 13). Zauważono również, że niektóre kompetencje są bardzo trudne albo bardzo kosztowne w nabyciu poprzez szkolenie i należą do cech indywidualnych, przyrodzonych. Dlatego na specjalistyczne szkolenia powinno się wysyłać tylko te osoby, których predyspozycje pozwalają na rozwinięcie umiejętności, które uczynią je ponadprzeciętnymi fachowcami.

Ryc. 13. Prezentacja rozkładu kompetencji pośród wybitnych i przeciętnych przesłuchujących



Źródło: Defense Academy for Credibility Assessment, *Assessment of Optimal Interrogation Approaches*, op. cit. [dostęp: 27.06.2020].

Drugi raport przygotowany na zlecenie DACA dotyczył roli rozmowy przedtestowej w badaniu poligraficznym i umiejętności potrzebnych w jej efektywnym przeprowadzeniu⁹⁶. W tym celu konsultanci najpierw zorganizowali pogłębione wywiady z 20 wykwalifikowanymi, wzorcowymi poligraferami z amerykańskich agencji rządowych, w których wykonywane są testy w konkretnych sprawach (przedmiotem zainteresowania w mniejszym stopniu były badania przesiewowe) – przy czym uznano, że za wzorzec może służyć badający, który jest skuteczny w prowadzeniu rozmowy przedtestowej, uzyskiwaniu dokładnych rezultatów testów z niewielką liczbą wyników nierozstrzygniętych, rzadko ma problemy z kontrolą jakości, a przełożony ufa mu w powierzaniu najtrudniejszych spraw. Średni wiek uczestników projektu wyniósł 47 lat, doświadczenie w zakresie badań poligraficznych – 11 lat, a ogółem w pracy śledczej – 19 lat.

Chodziło o określenie najważniejszych celów do osiągnięcia i optymalnych podejść stosowanych w rozmowie przedtestowej, a w konsekwencji usprawnienie procesu kształcenia i doskonalenia zawodowego kadry poligraferów oraz zwiększenie wartości diagnostycznej wyników badań. W wywiadach z poligraferami pozyskano wskazówki, które trafiły do trzech kategorii obejmujących cechy indywidualne, które powinien posiadać biegły; umiejętności oraz działania ułatwiające początkującym poligraferom szybsze nabycie wyższych kwalifikacji.

Na podstawie zebranych odpowiedzi podsumowano, że głównym celem do osiągnięcia w toku rozmowy przedtestowej było odpowiednie przygotowanie osoby badanej do poddania testom z wykorzystaniem poligrafu, a w tym m.in.: uzyskanie zgody na te testy; podjęcie decyzji, czy stan psychofizyczny umożliwi kontynuowanie badania; przedstawienie istoty i procedury testowania oraz zagadnień dotyczących fizjologii człowieka; ustanowienie właściwego nastawienia psychologicznego oraz przekonanie badanego do tego, że procedura jest rzetelna i dokładna; ustalenie wspólnego rozumienia krytycznego zagadnienia. Jako zasadniczą umiejętność, którą powinien posiadać poligrafer, wskazano nawiązanie kontaktu, dobrej komunikacji z badanym – stosownie do jego pochodzenia i poziomu intelektualnego. Wymieniano także zachowanie obiektywizmu i formułowanie

⁹⁶ Defense Academy for Credibility Assessment, *Assessment of Psychophysiological Detection of Deception (PDD) Pretest*, 2007, apps.dtic.mil/dtic/tr/fulltext/u2/a483331.pdf [dostęp: 27.06.2020].

zrozumiących pytań testowych. Spośród powyższych wyzwań za najważniejsze postrzegano uzyskanie zgody na badanie, osiągnięcie jednakowego rozumienia krytycznych zagadnień, które badanie ma zwerfikować oraz ułożenie listy pytań krytycznych i porównawczych (co jest zarazem zadaniem najtrudniejszym). Poligraferzy podkreślali też, że istotniejsze jest, aby osoba badana wierzyła w skuteczność badania, aniżeli rozumiała, dlaczego tak się dzieje – tym bardziej że materia jest skomplikowana i zwięzłe omówienie w wywiadzie mechanizmów psychofizjologicznych nie gwarantuje, że będzie to w pełni zrozumiałe. Z kolei w sytuacji, gdy na etapie rozmowy przedtestowej nie ma jeszcze jasności, czy kondycja osoby badanej nie stoi na przeszkodzie w przeprowadzeniu miarodajnych testów, poligraferzy nie rezygnują pochopnie i z reguły decydują się na przejście do fazy testowania, a dopiero opierając się na rejestrowanych danych fizjologicznych podejmują dalsze decyzje.

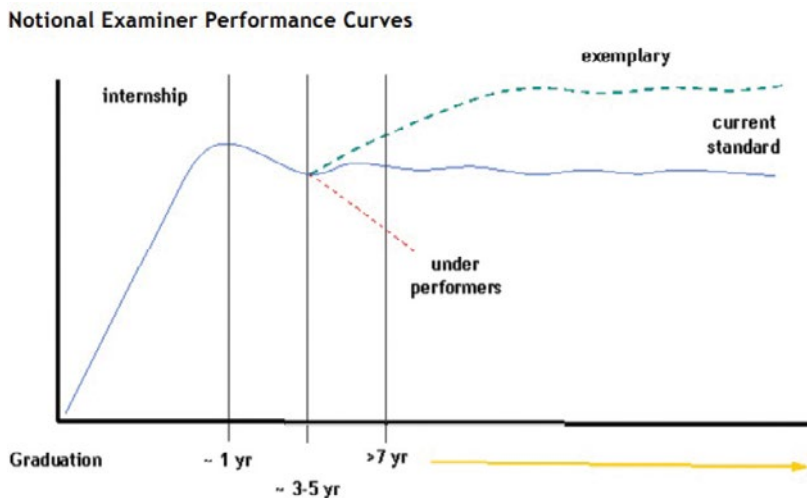
Jeżeli chodzi o indywidualne pożądane cechy poligraferów, oni sami wymieniali m.in.: pasję do tego zawodu; pewność własnych umiejętności przy jednoczesnym nieustannym doskonaleniu i czerpaniu z doświadczeń innych biegłych; zorientowanie na szczegóły; kreatywność; otwartość i bezstronność.

Analizowano ponadto, jak wygląda rozwój poligrafera wraz z upływem kolejnych lat praktyki. Z większości opinii wynika, że największy postęp następuje w okresie stażowym, w pierwszym roku po ukończeniu szkolenia podstawowego. W kolejnych kilku latach osiągnięcia nieco spadają (np. rośnie odsetek nierozstrzygniętych wyników testów). Przypuszcza się, że większość poligraferów odchodzi nieco w tym okresie od *książkowego* realizowania procedur badawczych – tak jak uczyli się tego na kursie. Kończy się ponadto indywidualna superwizja ze strony opiekuna stażu i poligraferzy niekiedy pozostają sami, bez wsparcia, a nowe problemy czy nietypowe przypadki wciąż się pojawiają. Niektórzy uznają też, że badania poligraficzne ostatecznie nie spełniają ich oczekiwań, jeśli chodzi o wybór ścieżki kariery i tracą motywację do dalszego rozwoju zawodowego.

Mniej więcej po ok. pięciu latach, przy zastosowaniu odpowiedniej interwencji o charakterze motywacyjnym i szkoleniowym, poligrafer może obrać jeden z trzech kierunków. Większość odzyska swój wcześniejszy poziom jakości wykonywanej pracy i ustabilizuje go do końca swojej kariery. Część dalej utrzyma się w trendzie spadkowym,

włącznie z całkowitym porzuceniem tej specjalności lub zwolnieniem z pracy. Wreszcie trzecia grupa zyska nowe spojrzenie i wraz z kolejnymi doświadczeniami wejdzie na wyższy poziom, stając się przykładem dla innych przedstawicieli profesji (zob. ryc. 14).

Ryc. 14. Hipotetyczna krzywa osiągnięć zawodowych poligrafera z biegiem lat stażu praktyki



Źródło: Defense Academy for Credibility Assessment, *Assessment of Psychophysiological Detection of Deception (PDD) Pretest*, op. cit.

Wyodrębniono ponadto czynniki, które pozytywnie i negatywnie wpływają na jakość pracy poligrafera. Pierwszą grupę spośród pozytywnych czynników tworzyły okoliczności powodujące wewnętrzny wzrost satysfakcji z realizowanych zadań, a wśród nich m.in. pomoc śledczym w rozwiązywaniu spraw, przyczynianie się do zapewnienia bezpieczeństwa państwa, oczyszczenia niewinnych osób z zarzutów itp. Poligraferzy cenią sobie również autonomię w pracy, niektórzy po prostu kochają to, co robią. W drugiej grupie znalazło się uznanie i szacunek ze strony innych osób – kolegów z pracy, przełożonych, zleceniodawców badań – wszystko to, co składa się na wysoką reputację przedstawicieli przedmiotowej specjalności. Wreszcie, trzecia grupa czynników obejmowała uwarunkowania instytucjonalne, po stronie pracodawcy – takie jak: zapewnienie budżetu na szkolenia, dobrego sprzętu, umożliwienie skupienia się wyłącznie na prowadzeniu badań poligraficznych – bez konieczności łączenia tego z innymi

obowiązkami. Ważna, ale nie najistotniejsza, okazała się motywacja finansowa. Szacunek ze strony bezpośredniego otoczenia górował nad tym aspektem. Część docenia również możliwość podróżowania, także za granicę, aczkolwiek zbyt częste wyjazdy i rozłąka z bliskimi negatywnie odbijają się na efektywności w pracy.

Po stronie negatywnych czynników o charakterze instytucjonalnym zwracano uwagę również na niekiedy zbyt duże obciążenie zadaniami służbowymi w konsekwencji braków kadrowych; na niezbyt kompetentnych współpracowników; niedostateczną ilość informacji przedstawianych przez zleceniodawców przed przeprowadzeniem badań i oczekiwanie z ich strony przeprowadzenia testów w sytuacjach, kiedy nie jest to odpowiednie. Ubolewano ponadto, że oficjalne rekomendacje dotyczące obowiązujących standardów badań nie nadążają za innowacjami, o których często trzeba dowiadywać się bezpośrednio od osób, które świeżo ukończyły kursy. Z kolei, jeżeli chodzi o okoliczności nieodłącznie związane ze specyfiką zawodową, eksperci narzekają m.in. na godziny nadliczbowe, krótki czas na realizację zadań, brak zrozumienia istoty badań poligraficznych przez laików.

Omawiany projekt wienńczyły konkluzje w zakresie potencjalnych działań, które mogłyby podjąć instytucja nadzorująca programy badań poligraficznych w USA – ówczesnie DACA, a obecnie National Center for Credibility Assessment (NCCA). Zaproponowano, by zastosować profilowanie osobowościowe kandydatów na poligraferów, mając na uwadze cechy, które predestynują do wysokich osiągnięć w tej profesji. Celowe byłoby też zwiększenie dostępności zaawansowanych szkoleń w okresie pierwszych kilkunastu miesięcy po ukończeniu kursu podstawowego, z uwzględnieniem większej liczby ćwiczeń praktycznych (w tym poprzez reżyserowane wywiady z osobami badanymi) i wykazaniem znaczenia fundamentów rozmowy przedtestowej dla późniejszej rozmowy po testach. Tematyka szkoleń doskonalenia zawodowego powinna koncentrować się na zdobywaniu umiejętności efektywnego nawiązywania relacji i przesłuchiwania, aktualizowaniu wiedzy o metodach interpretacji danych testowych. Wartościowe byłoby utworzenie odpowiednio zabezpieczonego internetowego forum wymiany doświadczeń między poligraferami, a bezpośrednie spotkania seminaryjno-warsztatowe należałoby organizować nie rzadziej niż raz na dwa lata. Dla osób, które weszły na ścieżkę kariery w zawodzie poligrafera, po okresie stażowym można by rozważyć programy mentorskie.

Dostrzeżono także zasadność szkolenia oficerów śledczych w zakresie korzystania z pomocy biegłych z zakresu badań poligraficznych.

V. Kopparumsolan na potrzeby swojej dysertacji doktorskiej analizował zespół cech poligrafera pod względem skuteczności detekcji nieuczciwości w wymiarze behawioralnych wskaźników werbalnych i niewerbalnych oraz przez analizę poligramów z zarejestrowanymi reakcjami fizjologicznymi osób badanych⁹⁷. Najpierw przez wywiady kwestionariuszowe z poligraferami z amerykańskich agencji rządowych pozyskał listę cech różnicujących osoby o ponadprzeciętnych zdolnościach w zakresie detekcji nieuczciwości od pozostałych. Następnie zajął się analizą pożądaných cech osobowości poligraferów w połączeniu z rezultatami szkoleń.

Rozpatrywane były takie konstrukty jak inteligencja emocjonalna, style poznawcze czy motywacje, które mogą odpowiadać za różnice indywidualne w zdolnościach poligraferów rzutujących na jakość ich pracy. Inteligencja emocjonalna bywa różnie definiowana. Odnosi się m.in. do zdolności rozpoznawania i rozumienia emocji oraz zarządzania nimi. Na przykład P. Salovey, D. Caruso i J. Mayer określili, że jest to zdolność do trafnej percepcji, oceny i ekspresji emocji; zdolność do asymilowania emocji w procesy poznawcze; zdolność do rozumienia i analizowania emocji oraz wykorzystywania wiedzy emocjonalnej, a także zdolność do regulacji emocji i kontroli emocjonalnej⁹⁸.

Pośród kilku wymiarów inteligencji emocjonalnej w detekcji nieuczciwości najistotniejsze wydaje się trafne odczytywanie sygnałów emocjonalnych (w szczególności tzw. przecieków świadczących o rzeczywistych przeżyciach względem jakiejś sytuacji, zagadnienia). Z tym na ogół lepiej radzą sobie kobiety. Podobnie – z wyrażaniem własnych emocji oraz analizowaniem i regulowaniem emocji innych ludzi. Z kolei mężczyźni mają z reguły wyższe zdolności rozumienia i kontroli własnych emocji⁹⁹. To drugie jest równie ważne jak interpretacja cudzych emocji, ponieważ zachowania badającego są bodźcami, które oddziałują

⁹⁷ V. Kopparumsolan, *The Effects of Examiner Personality Variables and Training on the Accuracy of Detection of Deception*, Singapore 2018, dr.ntu.edu.sg/bitstream/10356/73352/1/FINAL%20.pdf [dostęp: 28.06.2020].

⁹⁸ P. Salovey, D. Caruso, J.D. Mayer, *Emotional Intelligence Meets Traditional Standards for an Intelligence*, „Intelligence”, 27, 1999, s. 267–298.

⁹⁹ K.A. Knopp, *Czy kobiety naprawdę są bardziej inteligentne emocjonalnie niż mężczyźni? O różnicach międzyplciowych w zakresie zdolności emocjonalnych*, „Fides et Ratio”, 4(12), 2012, s. 95–112.

na badanego i jego reakcje. Mogą przedwcześnie wskazywać na kierunek, w jakim zmierzają rezultaty testów, albo ujawniać rzeczywisty stosunek badającego wynikający z oceny moralnej postępowania osoby poddawanej badaniu (np. podejrzanego o przestępstwa przeciwko wolności seksualnej, działającego w warunkach recydywy). W takich okolicznościach nie tylko badany stara się maskować swoje emocje, ale zdarza się, że i biegły również musi to czynić. Osoby o wysoko rozwiniętej inteligencji emocjonalnej z pewnością lepiej adaptują się do sytuacyjnych wymogów związanych z danym stanowiskiem pracy. Blisko związana z inteligencją emocjonalną jest tzw. obserwacyjna samokontrola. Badający, których cechuje wysoki stopień obserwacyjnej samokontroli, powinni łatwiej nawiązywać odpowiednie relacje ze swoimi interlokutorami, dopasowywać sposób komunikacji do ich profilu, a przez to pozyskiwać więcej istotnych informacji¹⁰⁰. Ogólnie rzecz biorąc – stabilność emocjonalna jest czymś, co sprzyja lepszemu radzeniu sobie ze stresem, a niewątpliwie sytuacja stresująca dotyczy nie tylko osoby poddawanej badaniu, ale i badającego – obciążonego poznawczo przy równoległym wykonywaniu wielu skomplikowanych czynności i dźwigającego ciężar odpowiedzialności za decyzje podejmowane w toku ekspertyzy.

Zwykle uważa się, że bardziej czule na sygnały niewerbalne (zwłaszcza ekspresję mimiczną) są kobiety i w ten sposób trafniej rozpoznają emocje. Część badań naukowych wskazuje jednak, że przewaga kobiet w tym aspekcie dotyczy jedynie rozpoznawania emocji pozytywnych, natomiast mężczyźni niekiedy lepiej odczytują negatywne emocje, włącznie z gniewem¹⁰¹. Kłamstwu może towarzyszyć nawet radość z oszukiwania, ale na pewno częściej strach, złość czy smutek, a więc emocje o pejoratywnym zabarwieniu.

Mając na uwadze, że u poligrafera pożądane są takie cechy jak: podejście mocno analityczne, zorientowanie na zadanie, względna odporność na czynniki behawioralne ze strony osoby badanej (pochwały, próby spoufalania, wzbudzania współczucia, rozmycia istoty sprawy

¹⁰⁰ Por. V. Koppurumsolan, *The Effects of Examiner Personality...*, op. cit., s. 46. Szerzej o samokontroli obserwacyjnej – zob. M. Snyder, *Self-monitoring of Expressive Behavior*, „Journal of Personality and Social Psychology”, 30(4), 1974, s. 526–537.

¹⁰¹ Zob. H.A. Elfenbein, A.A. Marsh, N. Ambady, *Emotional Intelligence and the Recognition of Emotion from Facial Expressions*, [w:] L.F. Barrett, P. Salovey (eds), *The Wisdom in Feeling. Psychological Processes in Emotional Intelligence*, New York 2002, s. 37–59.

przez nadmierne przekazywanie zbędnych informacji, gra przejawianymi emocjami itp.), dystraktory zewnętrzne podczas przeprowadzania badania i analizy danych – to kolejną przydatną kategorią określającą dobre psychologiczne predyspozycje do wykonywania tego zajęcia jest styl poznawczy, będący przejawem w sferze poznawczej szerszych wymiarów funkcjonowania indywidualnego człowieka¹⁰². To preferowany sposób funkcjonowania poznawczego, odpowiadający indywidualnym potrzebom jednostki¹⁰³. Należy go traktować jako stałą cechę w zakresie czynności percepcyjnych i intelektualnych. Style poznawcze rozpatruje się najczęściej w następujących wymiarach: impulsywność–refleksyjność; zależność–niezależność od pola percepcji; abstrakcyjność–konkretność.

Jeżeli impulsywność oznacza tendencję do szybkiego udzielania odpowiedzi oraz popełniania wielu błędów, a refleksyjność to tendencja do długiego namyślenia się i popełniania niewielu błędów – to, rzecz jasna, bardziej pożądaną u poligrafera będzie styl refleksyjny. Szczególnie jeśli weźmiemy pod uwagę konieczność często skomplikowanych analiz danych zebranych w toku przeprowadzonego badania poligraficznego. Z drugiej strony poligrafer powinien też działać szybko i sprawnie operować aparaturą badawczą w czasie testów – panować nad przebiegiem zapisów, dostrajać na bieżąco czułości czujników i dokonywać stosownych oznaczeń na poligramach, np. precyzyjnie uchwycić moment udzielenia odpowiedzi przez osobę badaną, a znaczenie dla miarodajności wyniku testu może mieć przesunięcie w czasie takiego oznaczenia nawet o pół sekundy.

Abstrakcyjność–konkretność określa preferencje dotyczące poziomu ogólności stosowanych kategorii poznawczych¹⁰⁴. Abstrakcyjność wskazuje na tendencję do posługiwania się bardziej ogólnymi kategoriami poznawczymi. Sprzyja rozwiązywaniu zadań złożonych i niejednoznacznych. Konkretność natomiast sprawdza się przy zadaniach prostych. Stwierdzono pozytywny związek abstrakcyjności z nasileniem orientacji zadaniowej, z większą elastycznością myślenia, umiejętnością

¹⁰² Zob. H. Witkin, *Psychologiczne zróżnicowanie i formy patologii*, „Przegląd Psychologiczny”, 16, 1968, s. 75, za: M. Kopeć, J.F. Terelak, *Style poznawcze a style radzenia sobie ze stresem u osób z przewlekłym bólem kręgosłupa*, „Studia Psychologica”, 9, 2009, s. 59–60.

¹⁰³ A. Matczak, *Style poznawcze*, [w:] J. Strelau (red.), *Psychologia. Podręcznik akademicki. Psychologia ogólna*, t. 2, Gdańsk 2006, s. 761.

¹⁰⁴ *Ibidem*, s. 770–771.

koordynowania i wykorzystywania przy podejmowaniu decyzji wielu różnych informacji. Co więcej – osoby preferujące abstrakcyjny sposób funkcjonowania charakteryzują się m.in. większą niezależnością od autorytetów oraz mniejszą pochopnością przy formułowaniu opinii o innych ludziach (co zatem powinno sprzyjać ograniczeniu liczby błędów poznawczych, skali subiektywizmu czy podatności na sugestię). Zatem *a priori* można przyjąć, że w zawodzie poligrafera styl abstrakcyjny powinien dawać lepsze efekty pracy, trafniejsze diagnozowanie.

Style poznawcze: zależny i niezależny od pola, wyróżnił H. Witkin. Niezależność od pola – to taki sposób spostrzegania, gdzie części pola percepcyjnego są oddzielone od zorganizowanego tła. Z kolei zależny od pola sposób spostrzegania charakteryzuje się całościową organizacją pola percepcji, jego *części doświadczane są jako stopione z całością*¹⁰⁵. Pierwszy z wymienionych stylów określa się też jako analityczny, drugi zaś – globalny. Osoby o stylu niezależnym charakteryzują się m.in. lepszymi zdolnościami wzrokowo-przestrzennymi, skuteczniej radzą sobie z zadaniami wymagającymi restrukturalizacji danych (co ma istotne znaczenie np. przy analizie zapisów na poligramach), przejawiają więcej zachowań asertywnych. Niezależność nasila się wraz z wiekiem, szczególnie między 8. a 24. rokiem życia, a u ludzi w wieku 60 lat i starszych spada. Stwierdza się w tym zakresie różnice międzypłciowe. Kobiety są bardziej zależne od pola, mężczyźni przeciwnie¹⁰⁶. Z badań, które przeprowadził D.R. Goodenough, wynika, że to właśnie osoby o stylu niezależnym preferują zawody, które wymagają obiektywnych analiz informacji¹⁰⁷. Ci z poligraferów, którzy wykazują tendencję w kierunku stylu niezależnego, powinni być bardziej odporni na manipulacyjne zagrywki ze strony osób poddawanych badaniom, a także wszelkie bodźce zewnętrzne, zakłócające procesy analityczne. Przykładowo – A. Vrij *et al.* wykonali eksperyment, który potwierdził, że policjanci o stylu niezależnym od pola percepcji podejmowali trafniejsze decyzje o oddaniu strzału z broni palnej, skuteczniej opierając się dystraktorom zewnętrznym¹⁰⁸.

¹⁰⁵ H. Witkin, *Psychologiczne zróżnicowanie...*, *op. cit.*, s. 76.

¹⁰⁶ Zob. A. Matczak, *Style poznawcze...*, *op. cit.*, s. 768.

¹⁰⁷ D.R. Goodenough, *Field Dependence*, [w:] H. London, J.E. Exner (eds), *Dimensions of Personality*, New York 1978, s. 165–216.

¹⁰⁸ A. Vrij, J. Van der Steen, L. Koppelaar, *The Effects of Street Noise and Field Independence on Police Officers' Shooting Behavior*, „Journal of Applied Social Psychology”, 25(19), 1995, s. 1714–1725.

Pośród innych aspektów struktury osobowości (z tzw. wielkiej piątki) – a pozytywnie skorelowanych ze specyfiką pracy biegłych z zakresu badań poligraficznych – intuicyjnie można wymienić skrupulatność, uważność. Są to cechy, które generalnie wydają się przydatne, również w innych zawodach. W przypadku poligrafera chodzi przede wszystkim o sprawne wychwytywanie różnych wskazówek werbalnych i niewerbalnych dotyczących zachowania osoby badanej, wnikliwe analizowane poligramów, umiejętność dostrzeganie różnych niuansów.

F. Enos *et al.* – zajmując się empirycznym badaniem zależności między głównymi cechami osobowości a detekcją nieszczeroci – wykazali, że zdolności w tym zakresie wzmocniają otwartość (stopień, w jakim człowiek jest dostępny na nowe doświadczenia i zdolny do dostosowania własnego punktu widzenia; koreluje również z poziomem inteligencji) oraz ugodowość, wrażliwość¹⁰⁹. Charakter bezpośredniej interakcji badającego z osobą badaną czy ekspozycja siły autorytetu biegłego świadczą z kolei o tym, że lepiej poradzą sobie z tym ekstrawertycy¹¹⁰. W gruncie rzeczy powyższe cechy występują u ludzi o towarzyskim usposobieniu, rozmownych, asertywnych, ambitnych, przejawiających inicjatywę. Człowiek otwarty bardziej polega na danych, które na bieżąco otrzymuje, aniżeli na z góry wyrobionych sądach. Ugodowość zaś (z wyjątkiem skrajnej postaci prowadzącej do zaburzenia osobowości zależnej) ułatwi zadanie zapewnienia atmosfery profesjonalnej współpracy i zaprezentowanie wobec badanego pozytywnego, pomocnego nastawienia (badanie poligraficzne zasadniczo różni się przecież od przesłuchania z elementami konfrontacji).

Wyróżnić można ponadto podstawowe, wewnętrzne, pozostające z reguły poza świadomym doświadczeniem, potrzeby i motywy towarzyszące pracy poligrafera. Wśród nich zarysowuje się potrzeba uznania, która materializuje się w demonstrowaniu autorytetu, kontroli nad przebiegiem procedury badawczej i wydawaniu instrukcji osobie poddawanej badaniu. Następnie motyw przynależności – właściwy

¹⁰⁹ F. Enos, S. Benus, R.L. Cautin, M. Graciarena, J.B. Hirschberg, E. Shriberg, *Personality Factors in Human Deception Detection. Comparing Human to Machine Performance*, „Interspeech”, 2006, s. 1–4.

¹¹⁰ Ekstrawersja jest pozytywnie skorelowana np. z pełnieniem służby w Policji (a wydaje się, że między śledczymi a poligraferami można tu szukać pewnej analogii) – zob. M.R. Barrick, M.K. Mount, *The Big Five Personality Dimensions and Job Performance. A Meta-Analysis*, „Personnel Psychology”, 44, 1991, s. 1–26.

ludziom raczej wrażliwym, zwracającym uwagę na różne sygnały (werbalne i niewerbalne) płynące ze strony innych ludzi. Wreszcie potrzeba osiągnięć, charakterystyczna dla ludzi chętnie podejmujących działania wymagające większej samodzielności i autonomii w podejmowaniu decyzji. Zważywszy na to, że detekcja nieszczeroci należy do wyzwań trudnych – można oczekiwać, że osoby z rozwiniętą potrzebą osiągnięć, prestiżu będą czerpały satysfakcję z tego typu zajęcia.

Panel ekspertów o co najmniej 10-letnim doświadczeniu zawodowym, wytypowany przez APA, najwyżej punktował – jako pożądane u poligraferów – takie cechy osobowości jak: potrzeby osiągnięć i uznania, uważność, niezależność od pola percepcji oraz otwartość¹¹¹. Próba empirycznej weryfikacji założeń wynikających z osobistych doświadczeń i przekonań poligraferów, a także z wnioskowań teoretycznych, której podjął się V. Kopparumsolan, potwierdziła tylko część z nich. Przy skomplikowanych sprawach osoby charakteryzujące się stylem poznawczym niezależnym od pola percepcji faktycznie wypadły lepiej niż osoby o stylu niezależnym. Ponadto motyw osiągnięć wraz z predyspozycją do większej skrupulatności również różnicował uczestników eksperymentów pod względem dokładności w detekcji nieszczeroci. Badacz zaobserwował pozytywną korelację między motywem przynależności a potwierdzaniem wersji osób prawdopodobnych. Z kolei w wykrywaniu osób nieszczerych najlepiej radziły sobie osoby z silną potrzebą przynależności i wysoko rozwiniętą inteligencją emocjonalną, ale jednak nisko umiejscowione pod względem ugodowości¹¹².

Opisanych powyżej kryteriów osobowościowych nie powinno się traktować jako *listy życzeń*. Wykaz predyspozycji psychologicznych kandydata na poligrafera powstał na podstawie badań empirycznych, prowadzonych nie tylko przez cytowanego doktoranta z Singapuru, ale również Departament Obrony USA. Taka lista może być przydatna np. przy selekcji kandydatów na stanowisko poligrafera w instytucjach, gdzie badanie psychologiczne jest elementem postępowania kwalifikacyjnego. Psycholog w takich postępowaniach zwykle wskazuje zresztą, do jakich grup stanowisk dany kandydat się nadaje, a do których nie. Jeżeli do doboru kadr mamy podchodzić patriotycznie i profesjonalnie – nie lekceważmy tych ustaleń.

¹¹¹ Zob. V. Kopparumsolan, *The Effects of Examiner Personality...*, *op. cit.*, s. 56.

¹¹² *Ibidem*, s. 167–177.

Były dyrektor laboratorium kryminalistycznego chicagowskiej Policji, a później profesor prawa na Northwestern University – F. Inbau – podkreślał, że poligrafer powinien wykazywać intensywne zainteresowanie pracą samą w sobie, dobre zrozumienie natury ludzkiej i posiadać odpowiednie cechy osobowości, które mogą jasno wynikać z generalnej zdolności do dobrego dogadywania się z ludźmi oraz bycia lubianym wśród przyjaciół i kolegów z pracy. F. Inbau stanowczo pozbawiał złudzeń osoby, które nie odpowiadają powyższej charakterystyce. Uważał, że żadna liczba szkoleń i doświadczeń nie pozwoli im na osiągnięcie satysfakcjonujących rezultatów – ani w postaci dokładnych diagnoz, ani doprowadzania do przyznania się przez osoby winne¹¹³. Takie osoby w ogóle nie powinny być przyjmowane na szkolenie. F. Inbau słusznie twierdził, że samo opanowanie obsługi poligrafu nie jest szczególnym wyzwaniem, za najważniejszą cechę uznał diagnozowanie na podstawie reakcji organizmu zarejestrowanych przez poligraf oraz umiejętne prowadzenie przesłuchania (wywiadu). Dostrzegł też potrzebę ścisłej specjalizacji. Jego zdaniem np. chemik, który jest dobry przy mikroskopie, z reguły nie będzie dobrym poligraferem i na odwrót¹¹⁴. Zapewne na bazie własnych doświadczeń zawodowych pisał, że to błąd, jeśli w Policji zakłada się, że w ramach laboratorium personel można łatwo przenieść z jednego wydziału do tego, który zajmuje się badaniami poligraficznymi. Praca poligrafera wymaga niepodzielnej uwagi i zaangażowania.

¹¹³ F. Inbau, *Some Avoidable Lie-Detector Mistakes*, „Journal of Criminal Law and Criminology”, 40(6), 1950, s. 793.

¹¹⁴ Skądinąd w praktyce polskiej Policji często łączy się dwie specjalności w laboratorium kryminalistycznym.

II. Dokładność identyfikacji i zgodność między poligraferami przy różnych metodach interpretacji zapisów na poligramach

1. Właściwości testów poligraficznych i ewolucja metod analitycznych

Biegłym, którzy wykonują badania kryminalistyczne, z pewnością powinno zależeć na tym, aby korzystać z takich narzędzi i metod badawczych, które pozwalają na osiągnięcie zamierzonego efektu, z minimalnym ryzykiem pomyłki. Pojęcie trafności określa, czy test mierzy to, do czego został zaprojektowany – jaka jest korelacja między wynikiem testu a tym, co miał ustalić, czyli kryterium¹. W przypadku badań poligraficznych chodzi więc o to, w jakim zakresie szczerza albo nieszczerza osoba zostanie poprawnie zdiagnozowana w konsekwencji rezultatu testu. Trafności nie powinno się mylić z wartością diagnostyczną testu poligraficznego, rozumianą jako uzyskana empirycznie (w badaniach naukowych dotyczących potwierdzonych przypadków) średnia dokładność – czyli odsetek osób poprawnie zdiagnozowanych spośród ogółu osób badanych, których wyniki testów były rozstrzygnięte (wyłącza się z reguły wyniki nierozstrzygające, skoro te ostatnie – z założenia – nie powinny skutkować żadną decyzją – ani na korzyść, ani na niekorzyść osoby badanej).

Z kolei rzetelność odnosi się do tego, czy dany test daje w tych samych warunkach powtarzalne wyniki. W aspekcie zewnętrznym rzetelności chodzi o zgodność między różnymi biegłymi oceniającymi tożsamy materiał badawczy. Natomiast rzetelność wewnętrzna dotyczy jednego biegłego, który ponownie analizuje te same dane testowe

¹ Zob. G.H. Barland, *The Reliability of Polygraph Chart Evaluations*, „Polygraph”, 1(4), 1972, s. 193.

w jakimś odstępie czasu. Metoda jest dobra wtedy, gdy jest zarówno trafna, jak i rzetelna. Stopień zgodności między oceniającymi sygnalizuje, z jakim zakresem subiektywizmu możemy mieć do czynienia.

Metody analizy danych testowych w badaniach poligraficznych historycznie kształtowały się, począwszy od metody jakościowej (tzw. globalnej analizy wykresów), wysoce subiektywnej, a przez to narażonej na rozbieżności między poligraferami – po częściowo zobiektywizowane metody ilościowo-jakościowe (numeryczne), wspomagane współcześnie również algorytmami komputerowymi. Od czasów J. Larsona (lata 20.) zasadniczą metodą była tzw. metoda globalna, jakościowa. We wczesnej fazie rozwoju badań poligraficznych odbywała się właściwie wyłącznie na podstawie osobistych doświadczeń z praktyki poligraferów, a w późniejszych latach, przynajmniej częściowo, z uwzględnieniem określonych cech diagnostycznych (wzorcowych reakcji), których katalog ulegał stopniowemu zawężeniu wraz z postępującymi badaniami naukowymi w tym zakresie (szerzej na ten temat w części III, rozdziale 1). Poza interpretacją samych zapisów na poligramach dużą wagę przywiązywano do innych, również jakościowych ocen, które odnosiły się do symptomów behawioralnych.

Pierwsze próby ilościowego szacowania istotności (częstotliwości i intensywności) zmian w reakcjach fizjologicznych przy poszczególnych bodźcach (pytaniach) testowych były podejmowane już w latach 30. J. Winter, który posługiwał się dwukanałowym poligrafem, składającym się z kanałów kardio i pneumo (*cardio-pneumo-psychograph*), oznaczał znakiem X obecność zmiany w reakcjach osoby badanej i oceniał jej znaczenie, wielkość. Wykazał przy tym, badając dziewczyny z akademika w sprawie wyreżyserowanej kradzieży, że test z wykorzystaniem poligrafu jest bardziej niezawodny od psychologicznego testu asocjacji słownej², który miał do tej pory zastosowanie w ocenie wiarygodności³. Na próżno jednak poszukiwać w literaturze

² Więcej o teście asocjacji słownej zob. J. Widacki, *Historia badań...*, *op. cit.*, s. 65–67.

³ Zob. J.E. Winter, *A Comparison of the Cardio-pneumo-psychograph and Association Methods in the Detection of Lying in Cases of theft Among College Students*, „*Journal of Applied Psychology*”, 20(2), 1936, s. 243–248. Kilka lat wcześniej także L. Keeler wykazał, że wykorzystanie dwukanałowego poligrafu daje dokładniejsze diagnozy wiarygodności niż test asocjacji słownej – por. *idem*, *A Method for Detecting Deception*, „*American Journal of Police Science*”, 1(1), 1930, s. 38–51.

przedmiotu, aby podobne działania, jeżeli chodzi o metodę analizy zapisów na poligramach, były powielane przez innych badaczy w najbliższym czasie.

Właściwie dopiero w latach 50., za sprawą D. Lykkena, pojawia się sposób oceniania reakcji w postaci rang liczbowych. Dotyczyło to wyłącznie jednego kanału zapisu – GSR, który D. Lykken brał pod uwagę w zaprojektowanym przez siebie teście GKT (CIT). Łączny wynik testu można było wykorzystać do obliczenia prawdopodobieństwa rozpoznania przez osobę badaną szczegółów związanych z weryfikowanym zdarzeniem – takich, które powinien znać wyłącznie jego uczestnik (sprawca przestępstwa albo naoczny świadek).

Lata 60. przyniosły nowe metody – numerycznej analizy danych testowych. Zaczęło się od J. Kubisa, który prowadził eksperymenty z udziałem studentów psychologii. Aby ułatwić im interpretację zapisów na poligramach, wprowadził 4-punktową skalę ocen (od 3 pkt do 0). Na 3 pkt oceniał reakcję bardzo znaczącą. Znaczącej przypisywał 2 pkt. Wątpliwie istotnej – 1 pkt, a nieistotnym reakcjom – 0 pkt⁴. Pod uwagę brano była tylko wielkość, znaczenie zmian w reakcjach, jeszcze bez bezpośredniego porównywania z innymi reakcjami. Analizę numeryczną, jaką znamy współcześnie (czyli taką, gdzie ocenia się nie samą wielkość reakcji, ale różnicę w reakcjach na określone pytania krytyczne i kontrolne), otrzymano dzięki C. Backsterowi w 1963 r. Wprowadzenie numerycznej oceny zapisów pozwoliło na osiągnięcie wyższych wskaźników dokładności diagnoz oraz rzetelności testów poligraficznych. Dowiedziono tego już w latach 70. m.in. dzięki grantom naukowym realizowanym na Uniwersytecie Utah⁵. Na gruncie metody Backstera powstały następnie systemy analizy danych testowych opracowane przez amerykańską szkołę żandarmerii wojskowej (United States Army Military Police School) oraz wspomniany Uniwersytet Utah. Także w praktyce Katedry Kryminalistyki Uniwersytetu Śląskiego w Katowicach stosowano metody numeryczne wzorowane na regułach Backstera.

⁴ J.F. Kubis, *Studies in Lie-detection Computer Feasibility Considerations*, 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; za: J.A. Matte, *Forensic Psychophysiology...*, *op. cit.*, s. 45–46.

⁵ D.C. Raskin, G.H. Barland, J.A. Podlesny, *Validity and Reliability of Detection of Deception*, „Polygraph”, 6(1), 1977, s. 1–39.

W latach 70. powróciła też koncepcja klasyfikowania istotności reakcji rangami. Japońscy naukowcy zaproponowali wówczas szeregowanie reakcji według ich wielkości (amplitudy) w parametrze GSR, który wydaje się najłatwiejszy do analizy i oszacowania bez stosowania narzędzi pomiarowych⁶. Największa amplituda otrzymywała rangę 1., następna – 2. i tak do 6. (tyle było bodźców w sekwencji testów szczytowego napięcia, które były w istocie testami demonstracyjnymi pochodzącymi z 30 badań w sprawach kryminalnych). Kiedy dwie reakcje miały jednakowe amplitudy, ich ranga była szacowana przez wyciągnięcie średniej z dwóch kolejnych rang, które przypadały tym reakcjom. Przykładowo – jeśli były to dwie największe reakcje, średnia z rangi pierwszej i drugiej wynosiła 1,5. Z kolei, gdy na wykresie nie zaistniały żadne istotne zmiany w kanale GSR (amplituda nie przekraczała 2 mm), każdy bodziec otrzymywał średnią z sumy sześciu rang, czyli $(1 + 2 + 3 + 4 + 5 + 6)/6 = 3,5$. Dalej sumowano rangi dla każdego bodźca testowego z czterech wykresów i obliczano średnią. Wszystkie bodźce, których średnia ranga wynosiła poniżej 1,99 – były krytycznymi. Spośród bodźców o najniższej randze po każdym teście faktycznie krytycznych okazało się łącznie 76,6%. Porównanie opisanej metody z wyłącznie subiektywną, globalną interpretacją zapisów (przeprowadzoną w sytuacji eksperymentalnej przez 26 poligraferów), nie wykazało istotnych różnic w trafności wskazań. Japońscy badacze sądzili przez to, że u oceniających wykresy globalnie, subiektywnie, proces myślowy przebiegał podobnie do obiektywnego wnioskowania na podstawie przypisanych rang.

Podobny tok myślenia do Japończyków przyjął w latach 80. XX wieku amerykański autor H. Timm, który podjął próbę zastosowania zobiektywizowanego systemu rangowania zmian reakcji w kanale pneumograficznym⁷. W odróżnieniu od kanału GSR – w tym kanale o istotniejszych reakcjach świadczą najmniejsze amplitudy, najkrótsze długości krzywej oddychania, a zatem najwyższą rangę przypisywano „najmniejszej” reakcji i dalej odpowiednio „większym” (o dłuższej linii oddychania). H. Timm sprawdzał przy tym, czy skuteczność identyfikacji „winnych” i „niewinnych” w przypadku oddychania piersiowego

⁶ A. Suzuki, S. Watanabe, K. Ohnishi, N. Arusuma, *The Objective Analysis of GSR in the Detection of Deception. An Analysis of GSR Amplitudes in Terms of Rank Scores*, „Polygraph”, 8(1), 1979, s. 53–63.

⁷ H.W. Timm, *Analyzing Deception from Respiration Patterns*, „Journal of Police Science & Administration”, 10(1), 1982, s. 47–51.

i przeponowego będzie w jakimś stopniu różniła się od siebie. Otrzymane wyniki nie odbiegały od siebie istotnie.

W czasie, gdy do praktyki poligraferów przebijały się już systemy ocen numerycznych (najpierw Backstera, a niedługo potem USAMPS i Utah), dwie pary badaczy niezależnie od siebie wysunęły propozycje systemów rangowania reakcji fizjologicznych we wszystkich podstawowych kanałach zapisów (pneumo, GSR i kardio). Były to: Horyzontalny System Oceniania (HSS – z j. ang. *horizontal scoring system*) N. Gordona i C. Cochetti⁸ oraz system oceniania według porządku rang (ROSS – z j. ang. *rank order scoring system*), autorstwa C. Hontsa i L. Driscolla⁹. Co ważne – reakcje w tych systemach nie były oceniane wyłącznie gołym okiem poligrafera, ale podlegały konkretnym pomiarom.

Choć ze współczesnej perspektywy można powiedzieć, że metody rangowania reakcji nie przyjęły się w powszechnej praktyce – to na pewno przyczyniły się do obiektywizacji sposobów analizy danych testowych i zapewne są stosowane przez część poligraferów, przynajmniej pomocniczo. Ustąpiły miejsca systemom analizy numerycznej – takim, gdzie różnicom w reakcjach na sąsiadujące ze sobą pytania krytyczne i kontrolne przypisuje się oceny w określonych skalach (3- lub 7-pozycyjnych). Wśród nich znajdują się osiągnięcia XXI wieku – dynamicznie zyskujący na popularności Empiryczny System Oceniania, którego głównym twórcą jest R. Nelson, oraz narzędzia komputerowe wykorzystujące algorytmy do analizy danych, z których najważniejszy wydaje się program OSS-3. Ewolucja różnych metod ewaluacji zapisów na poligramach była uwarunkowana dążeniem do tego, by spełniały one standardy naukowe, pozwalały na uzyskiwanie jak najdokładniejszych wyników testów w sposób możliwie najbardziej powtarzalny i obiektywny. W kolejnych rozdziałach omówiono szerzej wybrane metody.

2. Główne metody interpretacji danych w badaniach poligraficznych

Ewaluacja jest zadaniem służącym realizacji jednostkowych lub grupowych celów, przygotowaniem do podjęcia określonych decyzji.

⁸ N. Gordon, C. Cochetti, *The Horizontal Scoring System*, „Polygraph”, 16(2), 1987, s. 116–125.

⁹ C.R. Honts, L.N. Driscoll, *An Evaluation of the Reliability and Validity of Rank Order and Standard Numerical Scoring of Polygraph Charts*, „Polygraph”, 16(4), 1987, s. 241–257.

Ekspert dokonujący oceny występuje w roli narzędzia określającego nasycenie ocenianego przedmiotu wybraną wartością. [...] W wyniku oceny uzyskiwana jest większa kontrola nad funkcjonowaniem systemu, pozwalająca na podtrzymywanie stanów rzeczy uważanych za korzystne i eliminowanie lub ograniczenie stanów niekorzystnych¹⁰.

Wartości przypisywane ocenianym obiektom są z reguły względne, mają charakter instrumentalny w ramach hierarchii innych wartości i mogą pochodzić z różnych źródeł – zarówno od samego biegłego, jak i osób zainteresowanych jego oceną. Skoro tak – mogą na tym tle powstawać rozbieżności między uznawanymi przez różne strony wartościami i wynikające z tego konflikty. Zapobiegać nim ma swaista umowa w zakresie podstawowych wartości (np. bezpieczeństwo, sprawiedliwość), wspólna zgoda na pewne kryteria i konsekwencje oceniania, przełożenie na decyzje – tak jak czyni to osoba poddająca się badaniu poligraficznemu.

Ewaluacja jest procesem fazowym, w którym najpierw odbywa się wybór kryteriów – selektywne gromadzenie informacji o istotnych wymiarach zbioru alternatyw, którym przypisuje się określoną wagę. Następnie wymiary te są integrowane w celu otrzymania syntetycznego sądu o wartości każdej alternatywy¹¹. Różne strategie przetwarzania informacji mogą prowadzić do odmiennych ocen tego samego obiektu. W przykładowym modelu decyzyjnym w postępowaniu karnym, autorstwa N. Pennigton i R. Hastie, sędziowie muszą najpierw odkodować treść przypadku i ustalić kategorię przestępstwa. Dalej dokonują selekcji dowodów. Odtwarzają przebieg zdarzeń i oceniają wiarygodność dowodów. Porównują każdy z elementów przestępstwa z kategorią, do jakiej się zalicza. Biorą przy tym pod uwagę plany, cele i motywy głównych uczestników. Wreszcie następuje sformułowanie werdyktu¹².

Wymienieni badacze porównali użyteczność kilku modeli podejmowania decyzji – w tym: Andersona (każdemu dowodowi przypisuje się wagę *winny/niewinny* i uśrednia się to w ocenę globalną), Bayesa (ocenie podlega prawdopodobieństwo winy, testuje się hipotezy z uwzględnieniem kolejnych dowodów w sprawie) oraz Poissona

¹⁰ M. Materska, *Procesy ewaluacji*, [w:] M. Materska, T. Tyszka (red.), *Psychologia...*, *op. cit.*, s. 234.

¹¹ *Ibidem*, s. 237.

¹² Zob. N. Pennigton, R. Hastie, *Juror Decision-Making Models. The Generalization Gap*, „Psychological Bulletin”, 89(2), 1981, s. 246–287.

(z jednej strony mamy skumulowaną wagę dowodów przemawiających za określonym rozstrzygnięciem, zmieniającą się z każdym kolejnym etapem procesu, a z drugiej – kryterium decyzyjne, wskazujące, w którym momencie werdykt może zostać osiągnięty). Przy okazji tych badań, gdzie studenci wcielali się w role sędziów, ujawniły się takie efekty jak: wpływ informacji o osobowości oskarżonego na surowość i pewność werdyktu; wpływ stopnia surowości sędziów na sposób ważenia dowodów oraz efekt uprzedzenia w postaci przełożenia przekonania o winie oskarżonego na strategię postępowania. Ponadto waga przypisywana różnym faktom dowodowym różniła się między oceniającymi, a własne pojmowanie dobra i zła stawiali oni ponad to, co wynikało z regulowanych przez prawo stosunków między dowodami i winą. Ponownie więc ujawnia się subiektywizm i to nawet pomimo – wydawałoby się – obiektywnych (przynajmniej częściowo, np. wynikających z przepisów prawa) kryteriów oceny. Podobnych okoliczności, jakie zaistniały przy przyjęciu roli sędziego, należy się zapewne spodziewać także w przypadku odgrywania roli biegłego.

Aby zgłębić, czy biegły z zakresu badań poligraficznych w procesie analizowania danych testowych ma pozostawioną swobodę interpretacji, a jeśli tak – to jaki to może być margines – warto najpierw zapoznać się z poszczególnymi rodzajami ewaluacji i ich zasadniczymi założeniami. Metody te można sklasyfikować w sposób następujący:

- a) ze względu na sposób analizy zmiennych, brane pod uwagę przesłanki diagnozowania, sposób oszacowania relatywnego znaczenia zmian reakcji fizjologicznych oraz reguły decyzyjne – *metody jakościowe* oraz *metody jakościowo-ilościowe*;
- b) pod względem skali ocen numerycznych wyróżniamy *systemy 3- i 7-pozycyjne*;
- c) pod względem stopnia zaznajomienia z przedmiotem i podmiotem badania: samodzielną analizę poligrafera przeprowadzającego badanie oraz tzw. ślepą interpretację;
- d) w zależności od stopnia automatyzacji czynności: analizę manualną i komputerową.

2.1. Metoda jakościowa

Analiza danych z badania poligraficznego metodą jakościową polega na wizualnej interpretacji zapisów reakcji na poligramach –

w poszukiwaniu znaczących i konsekwentnych zmian miejscowych oraz ogólnych tendencji w sposobie reagowania osoby badanej w trakcie testu. Nazywana jest *globalną*, ponieważ na wykres patrzy się całościowo i obserwuje się, jak zmieniają się reakcje osoby badanej na przestrzeni testu. W tego typu analizie reakcje przy pytaniach krytycznych nie mają wyznaczonego konkretnego punktu odniesienia ani minimalnych wielkości, które pozwalałyby na uznanie jakiejś zmiany w reakcjach za znaczącą. Nie podlegają również fizycznym pomiarom. Biegły przeprowadzający analizę bazuje na własnych doświadczeniach z praktyki i zapamiętanych z literatury reakcjach wzorcowych.

Analiza poligramu składa się z kilku etapów. Wymienia je m.in. A. Krzyścin – najpierw wstępna analiza zmierzająca do ustalenia, czy wykres jest odpowiedniej jakości. Dalej – szczegółowa – zinterpretowanie odchyłeń od indywidualnej normy właściwej dla danego badanego, określenie, czym mogły zostać wzbudzone owe odchylenia, czy są pochodzenia psychologicznego, fizjologiczne, a może efektem jakiegoś czynnika zakłócającego. Każdy człowiek ma własny dominujący typ reaktywności fizjologicznej. Wspomniana „norma fizjologiczna” oznacza wzór krzywych zarejestrowanych przed wprowadzeniem bodźca testowego. Krzyścin postuluje, aby z ostrożnością podchodzić do podręcznikowych wzorów reakcji specyficznych, ponieważ zmiany reakcji zarejestrowane w czasie badania poligraficznego powinny być oceniane indywidualnie¹³.

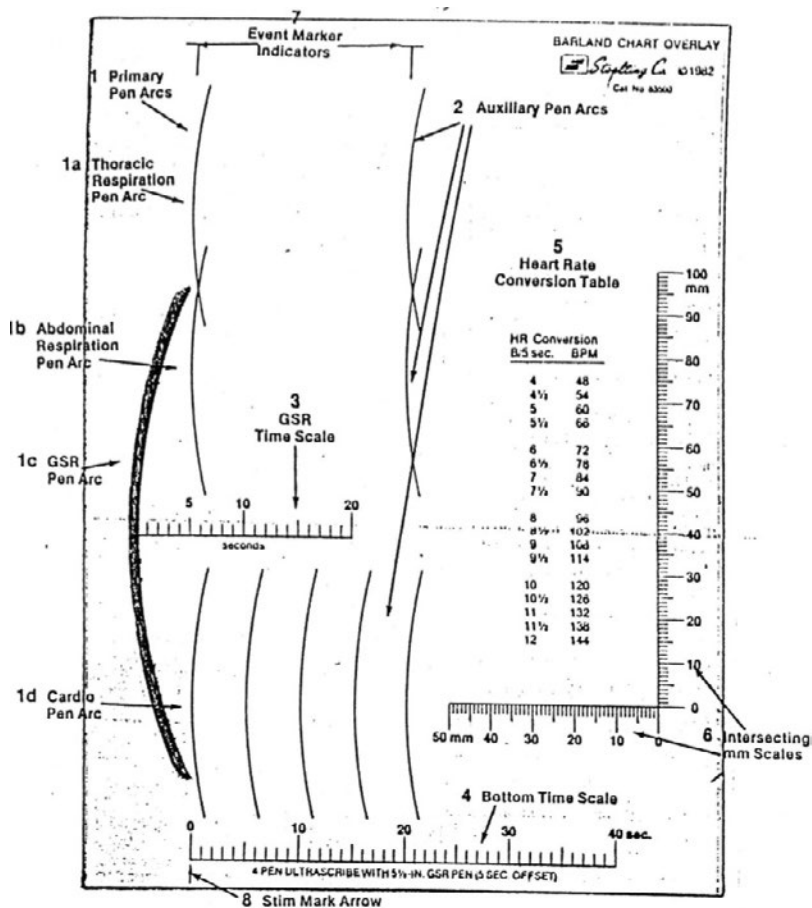
Wykres poligraficzny może być podzielony na kilka fragmentów. Według klasyfikacji J. Matte są to fragmenty przeciętnych zapisów, fazy reakcji, ulgi oraz zakłóceń. Przeciętne zapisy dotyczą poziomu tonicznego, kiedy nie ma jakichkolwiek istotnych zmian reakcji. Faza reakcji oznacza te zapisy, które świadczą o tym, że sympatyczny układ nerwowy został pobudzony w związku ze zmianą stanu emocjonalnego osoby badanej. Ulgę, odprężenie sygnalizują te zapisy, które pokazują działanie parasympatycznej części układu nerwowego, powrót do stanu równowagi. Z kolei zniekształcenia zapisów (artefakty) powstają, gdy odbiegają one od przeciętnego poziomu emocjonalnego osoby badanej, a jednocześnie nie ma fizjologicznych oznak, aby było to spowodowane czynnikami psychologicznymi¹⁴.

¹³ A. Krzyścin, *Ocena wyników...*, *op. cit.*

¹⁴ Zob. J.A. Matte, *Forensic Psychophysiology...*, *op. cit.*, s. 372–373.

Wizualna interpretacja zapisów była znacznie trudniejsza w dobie poligrafów analogowych. Wiele zależało wówczas od jakości wykresu. Po dokonaniu zapisu nie można było już skorygować czułości czujników, umiejscowienia krzywych czy dostosować skali – tak jak to umożliwia oprogramowanie komputerowe. W celu ułatwienia analizy i poprawienia jej precyzyjności tworzono w latach 80. XX wieku specjalne przezroczyste szablony, które nakładało się na poligram (zob. przykład – ryc. 15).

Ryc. 15. Szablon Barlanda, wspomagający analizę poligramów rejestrowanych przez poligrafy analogowe, produkowany przez amerykańską firmę Stoelting w latach 80. XX wieku



Źródło: A. Krzyżcin, *Ocena wyników badania poligraficznego* [niepublikowany skrypt].

Analiza globalna jest również częścią numerycznej interpretacji zapisów, ale jedynie we wstępnej fazie, kiedy sprawdza się poprawność zarejestrowanych danych i oznacza ewentualne artefakty. Poza tym ma zastosowanie do takich testów jak: R/I (pytań związanych i niezwiązanych), POT (szczytowego napięcia) czy praktycznie już przestarzałe – testy Reida i Arthera. Przy tym rodzaju analizy – dla uznania reakcji za istotne – ważne jest, aby były specyficzne, wyraźne i konsekwentne. Według wytycznych Instytutu Badań Poligraficznych Departamentu Obrony USA (Department of Defense Polygraph Institute, DoDPI) zmiany muszą wystąpić w co najmniej dwóch kanałach zapisu, dla dwóch z trzech powtórzeń danego pytania krytycznego¹⁵.

Analiza globalna bywa niekiedy mylona z tzw. klinicznym podejściem w ocenie wiarygodności, postulowanym przez szkołę J. Reida, które odnosi się do całego procesu badania, a nie tylko oceny reakcji osoby badanej. Ocena zapisów reakcji jest tylko jednym z czterech elementów tego całościowego podejścia. Pozostałe to: analiza faktów związanych z rozpatrywaną sprawą, analiza behawioralna oraz przesłuchanie po testach¹⁶. J. Reid i R. Arther opracowali procedurę wywiadu przedtestowego, która miała pomagać wydobyć wskazówki behawioralne przydatne w ocenie wiarygodności w ramach badania z wykorzystaniem poligrafu¹⁷. W technice Reida wskazówki te były uwzględniane przy podejmowaniu decyzji o końcowym wyniku badania, aczkolwiek – choć dowiedziono, że zachowanie osób prawdomównych i nieszczerých różni się – to jest to w praktyce bardzo trudne do trafnego uchwycenia¹⁸. Tam, gdzie mamy podejście kliniczne – zawsze występuje również jakościowa analiza zapisów, bo wskazanie testu nie zależy sztywno od obliczeń matematycznych, tylko od subiektywnej

¹⁵ Department of Defense Polygraph Institute, *DoDPI General Questions Test (GQT)*, 1995.

¹⁶ S.M. Slowik, *Global Evaluation. An Inductive Approach to Case Resolution*, „Polygraph”, 11(3), 1982, s. 215–224.

¹⁷ J. Reid, R. Arther, *Behavior Symptoms of Lie-Detector Subjects*, „Journal of Criminal Law and Criminology and Police Science”, 44, 1953, s. 104–110.

¹⁸ Por. F. Horvath, *Verbal and Non-verbal Clues to Truth and Deception during Polygraph Examinations* [praca magisterska obroniona na Michigan State University], 1972, d.lib.msu.edu/etd/15283/datastream/OBJ/download/Verbal_and_nonverbal_clues_to_truth_and_deception_during_polygraph_examinations.pdf [dostęp: 11.01.2021]. Zob. też artykuł o tym samym tytule w: „Journal of Police Science and Administration”, 1(2), 1973, s. 138–152.

oceny biegłego. Zresztą niezależne od sposobu oceny reakcji – jak pisze A. Krzyścin –

do postawienia diagnozy nie wystarcza sama analiza krzywych, bowiem należy w procesie interpretacji uwzględnić szereg zmiennych, które mogą być przyczyną odchyień w zapisie, takich jak wyniki obserwacji behawioralnej, introspekcyjne relacje osoby badanej itp.¹⁹

J. Matte zauważa z kolei, że subiektywizm w metodologii nie przeszkadzał wykorzystującym podejście kliniczne cieszyć się wysokim poziomem dokładności diagnoz, przekraczającym niekiedy 90%, kiedy czynili to poligraferzy odpowiednio wyszkoleni i doświadczeni²⁰. Badanie poligraficzne z podejściem klinicznym jest jednak bardziej sztuką, gdzie większą rolę odgrywają indywidualne predyspozycje i umiejętności biegłego, a nie procedura badawcza.

Przy metodzie jakościowej nie jest możliwe ilościowe oszacowanie siły dowodu. Sposób przedstawienia konkluzji ogranicza się wyłącznie do słownego opisu, a wyrażenie prawdopodobieństwa, stopnia pewności co do wyników badania może być zobrazowane i stopniowane poprzez takie wyrażenia jak: *z prawdopodobieństwem granicznym z pewnością, z wysokim prawdopodobieństwem, prawdopodobnie, zapisy zmian reakcji fizjologicznych badanego silnie wspierają hipotezę* itp.

Gdy biegły P. Horoszowski wydawał pierwsze opinie z zakresu badań poligraficznych na potrzeby postępowania karnego w Polsce, używał następujących sformułowań:

- [...] nie stwierdza się u Aleksandra P. przy zadawaniu pytań krytycznych związanych z zabójstwem Edwarda P. istotnych zmian fizjologicznych, odzwierciedlanych w przebiegu krzywych. Biorąc pod uwagę wyraźnie występujące u tego badanego zmiany w przebiegu tych krzywych przy eksperymentalnym teście cyfrowym, dochodzę do wniosku, że przy pytaniach krytycznych nie występują cechy, które pozwoliłyby przyjąć, że Aleksander P. w czasie badania odpowiadał nieszczerze.
- w odniesieniu do drugiego z podejrzanych: w przebiegu zarejestrowanych krzywych występują znaczne reakcje po pytaniach krytycznych [...]. Tego rodzaju reakcje uważa się w świetle dotychczasowych badań doświadczalnych za charakterystyczne dla osób, które przeżywają silny stan emocjonalny – w szczególności w związku

¹⁹ A. Krzyścin, *Ocena wyników...*, op. cit.

²⁰ J.A. Matte, *Forensic Psychophysiology...*, op. cit., s. 450.

z ujawnieniem lub grożącą realnie możliwością ujawnienia istotnych faktów, które chce się zataić przed innymi ludźmi. W każdym razie można ustalić, że omówione tu wykresy, stanowiące rezultat badania wariograficznego, nie są symptomatyczne dla osoby odpowiadającej szczerze na wszystkie krytyczne pytania, które dotyczą zabójstwa Edwarda P.²¹

Komentarza wymaga podejście metodologiczne zastosowane przy powyższych ekspertyzach. Widać na przykładzie wniosków z pierwszego badania, że test zapoznawczy (*eksperymentalny test cyfrowy* – jak to nazwał wówczas biegły) został potraktowany jako realny punkt odniesienia dla reakcji na pytania krytyczne w teście właściwym. Współcześnie spełnia wyłącznie cele demonstracyjne i nie służy ustaleniu żadnej reakcji wzorcowej (co najwyżej orientacyjnie). Podczas przesłuchania na rozprawie biegły zeznał m.in.: „Bardzo znamienne reakcja na pytanie krytyczne wystąpiła wówczas, gdy zadałem je w sposób niespodziewany”²². Nie wchodząc w ocenę trafności ogólnych ustaleń P. Horoszowskiego, trzeba zaznaczyć, że dziś taka interpretacja narażona byłaby na zarzut błędu w sztuce. Nie wolno bowiem opierać swojej diagnozy na pytaniach z zaskoczenia. W ogóle zadawanie pytań w ten sposób należy do sytuacji wyjątkowych i służy innym celom (sprawdzeniu, czy w aktualnym stanie psychofizycznym badany zachowuje zdolność reagowania na bodźce testowe). Z natury rzeczy zaskoczenie wywołuje istotne zmiany w reakcjach fizjologicznych, których nie należy łączyć z treścią pytania, a z samym faktem zaskoczenia.

Z kolei we wspomnianej wcześniej sprawie z Michigan (1972 r. – *United States v. Ridling*) amerykański sąd wymienił trzy rodzaje wniosków, które można wyprowadzić z badania poligraficznego:

- badany rozmyślnie nie mówi prawdy;
- badany przedstawia prawdę tak, jak ją postrzega;
- test jest nierozstrzygnięty, np. badający nie może stwierdzić, czy badany mówi prawdę.

Współcześnie uważa się, że po testach, z których dane analizowane są metodą jakościową, biegły nie powinien wyciągać kategoriycznych, daleko idących wniosków na temat wiarygodności osoby badanej.

²¹ Zob. fragmenty opinii biegłego P. Horoszowskiego w: M. Lipka, *Pierwszy w Polsce przypadek zastosowania „wykrywacza kłamstw” w postępowaniu karnym*, „Problemy Kryminalistyki”, 48, 1964, s. 252–254.

²² *Ibidem*.

Powinien ograniczyć się do wskazania, czy w czasie testu zaistniały przy pytaniach krytycznych znaczące zmiany reakcji fizjologicznych, czy też takich reakcji nie zarejestrowano. Niedookreśloność przedstawianego w ten sposób wyniku badania jest niewątpliwym mankamentem metody.

2.2. Metody numeryczne (ilościowo-jakościowe)

Jak już sygnalizowano wyżej, numeryczna analiza danych testowych została wprowadzona przez C. Backstera. Z początku miało to być ułatwienie dla uczniów jego szkoły, ale wkrótce okazało się, że koncepcja ta była przełomowa w rozwoju badań poligraficznych i zaczęła być wdrażana przez praktykujących poligraferów. Wydatnie pomogła w standaryzacji i obiektywizacji metod oceny zapisów na poligramach.

Obszary na wykresie, gdzie rejestrowano reakcje na pytania, C. Backster nazywał strefami o trzech kolorach: strefa czerwona obejmowała pytanie krytyczne, zielona – pytanie kontrolne (PLC), a czarna – pozostałe pytania, które dotyczyły problemów zewnętrznych (poza zakresem tematycznym pytań krytycznych i kontrolnych, np. obaw o zadanie pytania związanego z jakimś innym zdarzeniem ukrywanym przez badanego).

Różnice w reakcjach były ustalane odrębnie dla każdego rejestrowanego przez poligraf parametru. O tym, które pytanie kontrolne należało wytypować do porównania z pytaniem krytycznym, decydowała reguła *albo-albo* (z j. ang. *either-or*). Oznaczała ona, że gdy przy pytaniu krytycznym nie ma wyraźnej zmiany w reakcjach, dobierało się to z sąsiadujących pytań kontrolnych, przy którym była większa zmiana, zachodząca w odpowiednim czasie. Natomiast jeśli reakcja na pytanie krytyczne była znacząca – porównanie następowało z tym pytaniem, przy którym nie było reakcji lub była słabsza. Od razu przy tym warto zaznaczyć, że obecnie ta reguła nie jest akceptowana i w sytuacji, gdy dane pytanie krytyczne jest otoczone przez dwa pytania kontrolne²³ – do porównania reakcji w poszczególnych parametrach wybiera się, w zależności od techniki badawczej, to pytanie kontrolne, gdzie zmiany są istotniejsze (czyli odwrotnie niż u Backstera), ewentualnie – to pytanie kontrolne, które poprzedza pytanie krytyczne (bez

²³ Tak jest np. w przypadku formatu często stosowanego testu diagnostycznego – You-Phase / Bi-Zone.

względu na to, jakie reakcje występują przy pytaniu następującym po krytycznym). Krytycy omawianej reguły Backstera podnosili, że jest ona bardzo niekorzystna dla osób prawdomównych, ale ma też zwolenników – np. często idący pod prąd dominującym trendom w badaniach poligraficznych J. Matte dowodził teoretycznie i empirycznie jej słuszności²⁴.

W systemie Backstera ocenę dodatnią należało przypisać, gdy zmiany reakcji przy pytaniu krytycznym były mniej istotne niż przy porównywanym pytaniu kontrolnym. Ocenę ujemną, gdy to pytanie krytyczne wzbudzało bardziej znaczące zmiany reakcji – odzwierciedlając, że to na pytaniu tej kategorii badany skupia swoją uwagę. W przypadku braku dostrzegalnej różnicy w reakcjach na porównywane pytania – przyznawało się wartość zerową.

Poza samym stwierdzeniem różnicy w reakcjach należało ocenić jej wielkość. C. Backster zaproponował w tym celu skalę 7-pozycyjną (+3, +2, +1, 0, -1, -2, -3). Maksymalną wartością (3) oznaczał różnicę bardzo wyraźną, 2 – wyraźną, a 1 – niewielką²⁵. Następnie wyniki uzyskane przy analizie każdego kanału zapisu należało zsumować, co w skrajnych przypadkach, gdy poligraf rejestrował trzy podstawowe zmienne (pneumo, EDA i kardio), mogło dać +/- 9 pkt na jedno pytanie krytyczne. W teście, gdzie zadaje się trzy pytania krytyczne, po trzech seriach maksymalny wynik mógł wynieść +81, a minimalny -81. Próg decyzyjny C. Backster ustalił na poziomie +/-19. Wyniki dodatnie lub ujemne, które przekraczały ten próg, można było zakwalifikować jako wskazujące na szczerłość (NDI) lub wprowadzanie przez badanego w błąd (DI). Progi decyzyjne były zresztą zmienne, w zależności od liczby pytań relewantnych (dwa lub trzy) oraz powtórzeń testu (wykresów)²⁶. Jeśli test zawierał dwa pytania relewantne, a nie trzy jak w powyższym przykładzie – próg po trzech wykresach

²⁴ J.A. Matte, *A Field Study of the Backster Zone Comparison Technique's Either-Or Rule and Scoring System Versus Two Other Scoring Systems When Relevant Question Elicits Strong Response*, „European Polygraph”, 4, 2(12), 2010, s. 53–69. Por. E. Meiron, D. Krapohl, J.C. Kircher, *An Assessment of the Backster „Either-Or” Rule in Polygraph Scoring*, „Polygraph”, 37(4), 2008, s. 240–249.

²⁵ C. Backster, *Standardized Polygraph Notepack and Technique Guide*, New York 1969.

²⁶ Tylko u Backstera liczba wykresów wpływa na wartości progów decyzyjnych. W każdym innym znanym systemie analizy tego typu, który powstał później, progi ustanawiano na sztywnym poziomie.

wynosił ± 13 . Gdyby był jeden wykres mniej – to ± 9 . W latach 80. XX wieku C. Backster zdecydował się wprowadzić korekty w tym zakresie. Dodatni próg wymagany dla potwierdzenia prawdziwości był zbyt restrykcyjny i zwiększał liczbę wyników nierozstrzygniętych. Po modyfikacji w teście złożonym z trzech wykresów progi decyzyjne dla dwóch pytań krytycznych zostały ustanowione na: +7 dla NDI oraz -13 dla DI, zaś przy trzech pytaniach relewantnych: +10 dla NDI oraz -19 dla DI²⁷.

Metoda Backstera w wersji oryginalnej nie jest już współcześnie stosowana. Główne kierunki jego rozumowania zostały jednak zachowane, np. w zakresie stopniowania znaczenia zmian w reakcjach poprzez zastosowanie określonej skali ocen. Wśród nich znajdują się 3- i 7-pozycyjne skale wykorzystywane przy ewaluacji zapisów na poligramach z testów pytań kontrolnych, w tym:

- 7-pozycyjna „federalna” (czyli wypracowaną przez instytucje federalne w USA): [-3, -2, -1, 0, +1, +2, +3],
- 7-pozycyjna Utah: [-3, -2, -1, 0, +1, +2, +3],
- 3-pozycyjna *federalna* [-1, 0, +1],
- 3-pozycyjna w Empirycznym Systemie Oceniania (ESS): [-1, 0, +1], z wyjątkiem parametru EDA, gdzie przypisuje się wartości: [-2, 0, +2]. Podwójne ważenie tego parametru jest konsekwencją ustaleń naukowych, które dowodzą, że dane z tego parametru stanowią połowę wkładu w ostateczną diagnozę i najlepiej korelują z kryterium *winy*;

W skalach 3-pozycyjnych obowiązuje prosta zasada: *większe jest lepsze* (z j. ang. *bigger is better*). Natomiast skale 7-pozycyjne przewidują więcej wariantów oceny (na podstawie opisowych kryteriów – jak *bardzo znacząca reakcja* lub konkretne proporcje liczbowe wyrażające różnice w porównywanych reakcjach).

Numeryczne metody analizy danych testowych są obecnie wymagane we wszystkich testach rekomendowanych przez APA do badań dowodowych oraz dochodzeniowo-śledczych. Wśród tych testów są wyłącznie testy pytań porównawczych (CQT) oraz CIT (przykładowe formaty przedstawia tab. 3). Testy te muszą spełniać określone standardy w zakresie ich naukowej walidacji, dokładności oraz maksymalnego odsetka wyników nierozstrzygniętych (np. w badaniach

²⁷ Zob. R.S. Weaver, *Effects of Differing Numerical Chart Evaluation Systems on Polygraph Examination Results*, „Polygraph”, 14(1), 1985, s. 40–41.

dowodowych – co najmniej 90-procentowa dokładność oraz najwyżej 20% wyników bez rozstrzygnięcia). Jedynie na potrzeby przesiewowe dopuszcza się stosowanie szerszego katalogu technik (także takich, gdzie nie obowiązują oceny numeryczne, np. R/I) – są to wszelkie techniki, które zostały potwierdzone naukowo, o dokładności istotnie wyższej od statystycznej szansy, a wyniki badań empirycznych związanych z ich walidacją zostały opublikowane co najmniej dwukrotnie (w tym badania pierwotne oraz replikacyjne). Dodatkowym zabezpieczeniem jest w takim wypadku wymóg przeprowadzenia dodatkowych uznanych testów, gdy wynik testu przesiewowego jest niekorzystny dla osoby badanej²⁸.

Tab. 3. Wybrane formaty testów wykorzystywanych w badaniach poligraficznych

Nazwa testu	Sekwencja pytań
You-Phase (Bi-Zone)	N1, ScR, S1, C1, R1, C2, R2, C3, S2
Backster Tri-Zone Comparison Technique	N1, S1, ScR, C1, R1, C2, R2, C3, R3, S2
Federal ZCT	N1, ScR, S1, C1, R1, C2, R2, S2, C3, R3
Utah ZCT	I seria: I, ScR, N1, C1, R1, N2, C2, R2, N3, C3, R3 II seria: I, ScR, N2, C3, R2, N3, C1, R3, N1, C2, R1 III seria: I, ScR, N3, C2, R3, N1, C3, R1, N2, C1, R2
USAF MGQT – v. 1 (2–4 pytań relevantnych)	N1, ScR, C1, R1, C2, R2, C3, (R3, C4, R4)
USAF MGQT – v. 2 (2–5 pytań relevantnych)	N1, ScR, C1, R1, R2, C2, (R3, C3, R4, R5, C4)
Utah MGQT	I, ScR, N1, C1, R1, R2, C2, N2 (opcjonalnie), R3, R4, C3, N3
DLST	N1, N2, ScR, C1, R1, R2, C2, R1, R2, C1, R1, R2, C2
R/I	N1, T1, R1, R2, R3, N2, R4, R5, N1, R?, R? (dwa ostatnie są powtórzonymi pytaniami spośród R1–R5)
MQTZCT	N, ScR, S1, C1, R1, C2, R2, F, H, S2
IZCT	1N, 2I, 3ScR, 4N, 5C[ex], 6R, 7N, 8C[in], 9R, 10N, 11C, 12R, 13Cm

Legenda:

I (*introductory* – wprowadzające) – „Czy rozumie Pan/Pani, że w tym teście zadam tylko te pytania, które omówiliśmy?”

N (*neutral* – neutralne) – np. „Czy obecnie mamy rok 2021?”

ScR (*sacrificed relevant* – relevantne poświęcone) – np. „Jeżeli chodzi o kradzież biżuterii w minioną niedzielę ze sklepu jubilerskiego przy ul. Marszałkowskiej w Warszawie – czy zamierza Pan/Pani odpowiadać zgodnie z prawdą na każde pytanie?”

²⁸ Zob. APA, *Standards of Practice...*, *op. cit.*

S (*symptomatic* – symptomatyczne) – „Czy jest Pan całkowicie przekonany, że zadam w tym teście tylko te pytania, które omówiliśmy?”; „Czy obawia się Pan, że w tym teście zapytam o coś jeszcze – mimo obietnicy, że tego nie zrobię?”

C (*comparison* – porównawcze) – np. „Czy kiedykolwiek okłamał/a Pan/Pani bliską osobę?”

C[ex] (*comparison exclusive* – porównawcze wyłączające) – takie pytanie porównawcze, które nie obejmuje swoim zakresem czasowym lub tematycznym zagadnienia relewantnego

C[in] (*comparison inclusive*) – pytanie porównawcze, które obejmuje swoim zakresem czyn relewantny

R (*relevant* – relewantne) – np. „Czy zabrał/a Pan/Pani jakąkolwiek część biżuterii z tego sklepu?”

T (*overall truth* – ogólna prawdomówność) – np. „Czy zamierza Pan/Pani odpowiadać zgodnie z prawdą na każde z pytań tego testu?”

F (*fear of error* – obawa o błąd) – spełnia funkcję pytania porównawczego – „Czy obawia się Pan/Pani, że nastąpi błąd w tym teście w kwestii kradzieży biżuterii?”

H (*hope of error* – nadzieja na błąd) – traktowane jak krytyczne, w brzmieniu: „Czy ma Pan/Pani nadzieję na błąd w tym teście w sprawie kradzieży biżuterii?”

Cm (*countermeasures* – środki zakłócające) – „Czy robi Pan/Pani coś celowego, aby zmienić prawidłowy wynik testu?”

Źródło: opracowanie własne autora.

Systemy numerycznej analizy zapisów na poligramach – jeśli nie wyeliminowały całkowicie – to na pewno ograniczyły kierowanie się przez badających innymi przesłankami diagnozowania niż wynikające bezpośrednio z danych rejestrowanych przez poligraf, w tym jakościowymi ocenami zachowania osoby badanej. Metoda numeryczna nazywana jest niekiedy *półobiektywną* – z tego względu, że element subiektywny ma miejsce do momentu uznania przez biegłego określonych fragmentów zapisów na poligramach za symptomy znaczących zmian w reakcjach fizjologicznych i przypisania im określonych wartości liczbowych. Dalej natomiast wskazanie testu i diagnoza w kwestii wiarygodności osoby badanej pozostają już zobiektywizowane, ponieważ biorą się wprost z rezultatu liczbowego oraz sformalizowanych reguł decyzyjnych. Nie ma tu miejsca na zmianę wyniku przez poligrafera na podstawie czynników subiektywnych. Biegły może natomiast wzmocnić lub osłabić wydźwięk tego wyniku, formułując wnioski w swojej opinii. Wzmocni, gdy np. podkreśli jednoznaczność wskazania, powoła się na rezultat analizy statystycznej wskazujący na niskie prawdopodobieństwo błędu czy zapewni, że w toku badania nie dostrzegł istotnych czynników kontaminujących. Osłabi natomiast, jeśli zaznaczy, że wskazanie testu należy potraktować z ostrożnością, np. ze względu na liczne artefakty na poligramach albo stan psychofizyczny

osoby badanej – sprawiający, że osoba ta znajdowała się na granicy dopuszczalności do poddania testom z wykorzystaniem poligrafu. Dokonywana przez biegłego samodzielna ocena stopnia pewności co do wyników badania – choć może opierać się na obiektywnych przesłankach (jak wspomniany stan psychofizyczny osoby badanej) – zawsze pozostanie subiektywna. Kluczowe znaczenie dla oceny dowodu z opinii biegłego z zakresu badań poligraficznych ma w takim wypadku uzasadnienie stanowiska, jakie biegły wyraził w swoich konkluzjach.

Eksperti podporządkowani wytycznym APA, czy obowiązującym np. federalne instytucje w USA, zobligowani są przy rozstrzyganiu o wynikach testów z wykorzystaniem poligrafu do postępowania ściśle według kryteriów przewidzianych dla danej techniki badawczej i obranej metody analizy danych testowych. Osobliwym przypadkiem jest praktyka poligraferów w służbach izraelskich, gdzie biegłym pozostawia się większy margines swobody przy interpretacji wyników testów. Tamtejsi poligraferzy mogą wydać opinię kategoryczną albo bardziej powściągliwą, z poczynionymi zastrzeżeniami w kwestii osiągniętego rezultatu testu – podobnie jak wskazano w przykładach wyżej, lecz o ile czynienie takich zastrzeżeń należy do sytuacji raczej wyjątkowych w powszechnej praktyce – w Izraelu jest to podejście standardowe. Co więcej – eksperci w Izraelu nie poruszają się w sztywnych ramach punktowych progów decyzyjnych czy przyjętej tolerancji błędu (alfa). Matematycznie nierozstrzygnięty rezultat numerycznej analizy danych uzyskanych w toku danego testu nie musi przekładać się na brak konkluzywnej opinii, a więc pomimo nieosiągnięcia jednego z ustalonych progów decyzyjnych biegły może zaznaczyć, że zaobserwował tendencję w wynikach testu w kierunku wskazania na wprowadzanie w błąd lub przeciwnym²⁹.

2.2.1. Klasyfikowanie zmian w reakcjach metodą przypisywania rang (HSS i ROSS)

N. Gordon i P. Cochetti zauważyli, że u słuchaczy, których nauczali badań poligraficznych, występują spore rozbieżności przy analizie danych zarejestrowanych na poligramach. Przypuszczali, że wynikało to z jednej strony z czynników subiektywnych przy decydowaniu o tym, które zapisy reakcji uznać za istotne oraz jakie wartości liczbowe należy im przypisać w skali 7-pozycyjnej dla oszacowania różnic w wielkości

²⁹ Zob. A. Ginton, *The Importance of the Consistency Factor in CQT and Other Polygraph Tests*, „Polygraph”, 42(3), 2013, s. 153.

reakcji. Z drugiej strony wspomniane wyżej systemy: Backstera, amerykańskiej armii i Utah, dostarczały innych wytycznych w kwestii ewaluacji reakcji (w szczególności – z którym pytaniem kontrolnym w teście należało porównywać reakcje na pytanie krytyczne)³⁰.

N. Gordon i P. Cochetti uznali, że przestrzeń na subiektywną interpretację powinna zostać ograniczona. Dlatego zaproponowali własny system, który nazwali *horyzontalnym*. W tym systemie porównywane ze sobą są reakcje na wszystkie pytania krytyczne i kontrolne, zadane w czasie jednej serii testu (jednego zarejestrowanego wykresu). Każdy kanał zapisu analizowany jest odrębnie. Najwyższą rangę przyznaje się najsilniejszej reakcji (przy czym tutaj liczbowo oznacza się rangi malejąco, czyli jeśli mamy np. trzy pytania krytyczne i trzy kontrolne – najwyższą będzie ranga 6., a najsłabsza reakcja zostanie oznaczona rangą 1.). Następnie do rang zostają przyporządkowane oceny numeryczne, odpowiadające rangom – z tym, że dopisuje się jeszcze znak dodatni lub ujemny, w zależności od tego, czy dana ocena dotyczy reakcji na pytanie kontrolne czy krytyczne. Ostatecznie oceny podlegają zsumowaniu. Próg decyzyjny ustala się w powiązaniu z liczbą pytań krytycznych. Na każde pytanie krytyczne przypada wartość $+/-2$. Zatem, jeśli test składa się z trzech serii, w których znajdują się po trzy pytania krytyczne, będzie obowiązywał próg $+/-18$. Analogicznie – w teście z dwoma pytaniami krytycznymi, przy trzech zarejestrowanych wykresach, próg wyniesie $+/-12$. Choć w dalszym ciągu istnieje pewien zakres subiektywizmu, gdy poligrafer decyduje o tym, które zmiany w reakcjach fizjologicznych są istotne – od momentu ustanowienia ich hierarchii na wykresie dalsze postępowanie jest w pełni obiektywne i nie ma potrzeby głowić się, które z pytań kontrolnych wybrać dla porównania z danym pytaniem krytycznym i jaką ocenę numeryczną przypisać w 7-pozycyjnej skali.

Horyzontalny system oceniania był postrzegany przez niektórych uczonych – np. C. Hontsa i N. Driscolla – z dozą nadziei, wskazując, że wiele decyzji o przypisaniu reakcjom fizjologicznym określonych rang można podjąć na podstawie prostych obiektywnych pomiarów, bez konieczności skrupulatnego porównywania i szacowania różnic w wielkości poszczególnych reakcji, eliminując w ten sposób z procesu analizy czynniki subiektywne³¹. Uważali, że system rang będzie mniej

³⁰ N. Gordon, C. Cochetti, *The Horizontal...*, *op. cit.*, s. 116–125.

³¹ C.R. Honts, N.L. Driscoll, *An Evaluation of the Reliability...*, *op. cit.*

skomplikowany i łatwiejszy do przekazania podczas szkoleń poligraferów. Ci sami naukowcy, zainspirowani systemem horyzontalnym, dostrzegali jednak pewne jego wady i uznali, że warto zaprojektować system, który będzie jeszcze bardziej obiektywny i prostszy. Założenia metody (nazwanej w skrócie ROSS) i wyniki badań empirycznych opublikowali w drugiej połowie lat 80. XX wieku³².

W systemie ROSS porządkowanie reakcji według rang odbywa się na takiej samej zasadzie jak w systemie *horyzontalnym*. Jeżeli chodzi o różnice między tymi systemami – w ROSS oceniane są nie trzy, a cztery kanały zapisu, włącznie z pletysmograficznym. Próg decyzyjny ustalono na poziomie ± 13 (został obliczony na podstawie danych z testu Utah ZCT, zawierającego trzy pytania krytyczne). Jasno wskazane są także cechy reakcji, które należy mierzyć, aby je porównać ze sobą. W kanale EDA mierzy się amplitudę reakcji, a jeśli są jednakowe – bierze się pod uwagę dodatkowo złożoność reakcji i ich czas trwania. Jeżeli chodzi o względne ciśnienie krwi – mierzy się wzrost od linii bazowej po stronie rozkurczowej. Gdy dane te są równe, za istotniejszą reakcję uznaje się tę, która trwa dłużej. Jeżeli chodzi o amplitudę pulsu (kanał PLE), pomiarowi podlega zwężenie naczyń krwionośnych, a dodatkową cechą – podobnie jak w kanale kardio – jest czas reakcji. O ile szacując wielkości powyższych rodzajów reakcji, należało posłużyć się jakimś narzędziem, np. linijką czy skorzystać z podziałki na wykresie – to w przypadku kanału oddechowego takiego pomiaru nie wykonywano. Kryterium diagnostycznym uznawanym w tym kanale była długość linii oddechu w odpowiednim odcinku analizowanym na wykresie (im krótsza, tym reakcja bardziej istotna). Krzywa odzwierciedlająca zmiany w cyklu oddechowym podlega ciągłym fluktuacjom przy wdechach i wydechach. Dziś, gdy eksperci mają do dyspozycji poligrafy komputerowe, są w stanie w razie potrzeby dokładnie zmierzyć linię oddechu. Wówczas, gdy projektowano system ROSS, dominowały jeszcze poligrafy analogowe. Raz zarejestrowany zapis na poligramie nie mógł być poddany żadnej obróbce. Rzeczoną długość linii oddechu należało więc sobie wyobrazić i w ten sposób porównać zapisy na wykresie, co oczywiście było obciążone większym ryzykiem błędnej interpretacji. Z drugiej strony – nawet obecnie, pomimo doświadczeń związanych z komputeryzacją, większość analiz zapisów na poligramach odbywa się bez korzystania z narzędzi pomiarowych.

³² *Ibidem*.

Różnice w reakcjach – aby zostały uznane za istotne – powinny być na tyle wyraźne, aby móc je dostrzec gołym okiem.

Przy wykorzystaniu metody ROSS zweryfikowana empirycznie dokładność diagnoz odnoszących się do wiarygodności osób badanych została oszacowana na podobnym poziomie jak przy 7-pozytywnej numerycznej analizie danych testowych według Uniwersytetu Utah, a więc nieco ponad 90%. Zatem pomimo że metoda ROSS nie okazała się pod tym względem jakoś znacząco lepsza, wśród jej zalet autorzy dostrzegali przede wszystkim trafność i rzetelność, prostotę oraz zwiększony obiektywizm. Spójne rezultaty przyniosła replikacja badań empirycznych dotyczących systemu ROSS, dokonana przez innych badaczy, z rozszerzeniem analizowanych danych na te, które pochodziły z 300 rzeczywistych spraw – po równo, jeśli chodzi o osoby szczerze i nieszczerze (C. Honts i N. Driscoll bazowali wyłącznie na danych laboratoryjnych)³³.

Podane wyżej punktowe progi decyzyjne – zarówno w systemie HSS, jak i ROSS – odnosiły się wyłącznie do testów jednowątkowych typu ZCT i bez dodatkowych badań empirycznych nie wiadomo było, czy identyczne wartości są odpowiednie dla testów wieloproblemowych, gdzie każde pytanie krytyczne ocenia się niezależnie od pozostałych (ze względu na brak logicznego wzajemnego powiązania ich treści). Naukowcy z Uniwersytetu Utah przeprowadzili więc badania uzupełniające, aby ustalić, jak ROSS sprawdzi się w przypadku testu wieloproblemowych typu MGQT³⁴. Optymalny próg decyzyjny dla pojedynczego pytania relewantnego wyznaczono na poziomie $+/-2$. W sprawach, w których osoby badane odpowiadały na wszystkie relewantne pytania testowe jednolicie (czyli albo wyłącznie zgodnie z prawdą, albo nieszczerze), rezultaty były bardzo dobre, zbliżone do tych osiągniętych z wykorzystaniem numerycznej analizy Utah. Mniej obiecujące wyniki uzyskano w przypadku tych testów, podczas których były odpowiedzi w części szczerze, a pozostałe przeciwne. Istotnie więcej otrzymano też fałszywych wskazań na nieszczerze. Świadomi niższej użyteczności systemu ROSS w badaniach przesiewowych

³³ D.J. Krapohl, D.W. Dutton, A.H. Ryan, *The Rank Order Scoring System, Replication and Extension with Field Data*, „Polygraph”, 30(3), 2001, s. 172–181.

³⁴ C.R. Honts, L.N. Driscoll, *A Field Validity Study of the Rank Order Scoring System (ROSS) in Multiple Issue Control Question Tests*, „Polygraph”, 17(1), 1988, s. 1–15.

autorzy sugerowali, by mimo wszystko starać się z niego pomocniczo korzystać nawet w przypadku testu pytań związanych i niezwiązanych, gdzie nie ma żadnych pytań kontrolnych. Pomiar i porządkowanie według rang zmian reakcji na same tylko pytania krytyczne może pomóc wytypować te z nich, które podlegałyby dalszej weryfikacji w testach z pytaniami porównawczymi.

2.2.2. System Lykkena

D. Lykken pod koniec lat 50. XX wieku przeprowadzał eksperymenty, w których na podstawie zmian reakcji elektrodermalnych próbowano typować osoby *winne* czynów zaistniałych w ramach zainscenizowanych zdarzeń. Warto podkreślić, że chodziło mu o detekcję winy, a nie kłamstwa. Krytycznie podchodził do ówczesnej metodologii badań poligraficznych, w tym koncepcji zadawania pytań krytycznych i kontrolnych. Uważał, że zadawano je w sposób niewłaściwy, z dużym ryzykiem fałszywych wskazań testów.

Wobec powyższego D. Lykken wysunął własną ideę – *guilty knowledge test* (GKT – test wiedzy winnego). Test miał na celu sprawdzenie, czy osoba badana rozpoznaje istotny szczegół związany z danym zdarzeniem, umiejscowiony pośród szczegółów jedynie przybranych. Test poprzedzało wprowadzenie, np.:

Jeżeli to Ty jesteś mordercą, będziesz wiedział, że w pokoju, gdzie doszło do zabójstwa, był obecny nietypowy przedmiot... Czy był to: a) gramofon, b) sztaluga, c) bombonierka, d) zestaw szachów? Morderca schował broń w jednej z szuflad biurka. Która to była szuflada? Czy była to: a) górna lewa, b) dolna prawa, c) górna prawa, d) dolna lewa?³⁵

Osoby badane tylko słuchały tych wariantów, a zapisy psychogalwanometru wskazywały, jak reagują przy każdym z nich. D. Lykkenowi udawało się w ten sposób niemal bezbłędnie wytypować osoby winne.

W zależności od tego, ile testów tego typu uda się przeprowadzić, D. Lykken wyliczał prawdopodobieństwo, że osoba niewinna rozpoznaje właściwe szczegóły. Przykładowo – gdy przeprowadzi się dwa testy zawierające po pięć wariantów odpowiedzi – prawdopodobieństwo to wyniesie $0,2 \times 0,2 = 0,04$ i będzie mniejsze z każdym kolejnym

³⁵ D.T. Lykken, *The GSR in the Detection of Guilt*, „Journal of Applied Psychology”, 43(6), 1959, s. 386.

testem³⁶. Zaleca się, by starać się przeprowadzić przynajmniej trzy serie testu (w każdej po jednym unikalnym szczególe).

Dla analizy zapisów na poligramach D. Lykken wprowadził system, który polegał na uszeregowaniu zmian reakcji w kanale GSR według istotności (wielkości amplitudy). Pomija się reakcje na pierwszy bodziec testowy jako bufor pochłaniający odruch orientacyjny związany z początkiem testu. Jeśli najistotniejsza zmiana reakcji występuje przy krytycznym szczególe – przyznaje się dwa punkty. Jeden – gdy wielkość reakcji jest druga na przestrzeni testu. W pozostałych przypadkach – dana seria testu była oceniana na zero. Po wykonaniu wszystkich serii oceny sumuje się, a zatem wynik zawiera się w granicach od zera do podwójnej liczby serii testu. Próg decyzyjny wynosi tyle, ile wykonano serii testu. Równy lub wyższy wskazuje na rozpoznanie przez badanego szczegółów zdarzenia. Niższy – przeciwnie. Możliwe jest przy tym wskazanie prawdopodobieństwa, że badany nie rozpoznaje krytycznych szczegółów (zob. tab. 4). To prawdopodobieństwo jest funkcją liczby przeprowadzonych serii testu oraz punktowego rezultatu analizy danych. Przykładowo – gdy zostanie przeprowadzonych 5 serii testu, a wynik wyniesie 8 – prawdopodobieństwo, że badany nie rozpoznał *ukrytych w teście informacji* będzie oszacowane na 1%. Taki wynik silnie wspiera więc pozytywną identyfikację. Jeśli testowi poddawana jest osoba winna – w reakcji na bodziec krytyczny powinny wystąpić specyficzne mimowolne zmiany reakcji fizjologicznych tej osoby z powodu różnic w znaczeniu krytycznego bodźca względem bodźców kontrolnych. Daje to podstawę do generalizacji typu: **R** – „Osoba **A**, rozpoznaje zdarzenie **p**„”. Uprawnione jest zatem następujące uzasadnienie zdania **R**:

Zarejestrowano 6 wykresów testu CIT (po jednym dla każdego z podtestów). Na podstawie analizy danych testowych zgodnie z systemem Lykkena, opartym na danych z komponentu EDA, uzyskano ogólny wynik: *X/Y pkt* przy *Z* podtestach CIT – np. 7/12 pkt przy 6 podtestach CIT. Prawdopodobieństwo, że taki rezultat testu mógł zostać wytworzony przez badanego, który nie rozpoznał kluczowych faktów związanych z przedmiotowym zdarzeniem, wynosi: *P* (np. 1%)³⁷.

³⁶ *Idem, Psychology and the Lie Detector Industry*, University of Minnesota 1974, s. 9–11.

³⁷ Zob. A. Ibek, M. Gołaszewski, *Formułowanie opinii z badań poligraficznych i tworzenie argumentów dowodowych*, [w:] M. Gołaszewski (red.), *Współczesne standardy...*, *op. cit.*, s. 93.

Tab. 4. Prawdopodobieństwo nierozpoznania przez badanego szczegółów weryfikowanego zdarzenia przy przeprowadzaniu testu CIT/GKT, złożonego z dwóch do ośmiu serii (podtestów), analizowanego systemem Lykkena

CIT-y ↓	WYNIK →														
	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
2	0.12	0.04													
3	0.28	0.13	0.03	0.01											
4	0.44	0.25	0.10	0.04	0.01	0.00									
5	0.58	0.38	0.20	0.09	0.03	0.01	0.00	0.00							
6	0.69	0.50	0.31	0.17	0.08	0.03	0.01	0.00	0.00	0.00					
7	0.78	0.61	0.42	0.26	0.14	0.07	0.03	0.01	0.00	0.00	0.00	0.00			
8	0.84	0.70	0.53	0.36	0.22	0.12	0.06	0.03	0.01	0.00	0.00	0.00	0.00	0.00	

Źródło: D. Krapohl, J.B. McCloughan, S.M. Senter, *How to Use the Concealed Information Test*, „Polygraph”, 38(1), 2009, s. 43.

W późniejszym czasie nazwa testu GKT została zastąpiona CIT (z j. ang. *concealed information test* – test ukrytych informacji). W Polsce był w przeszłości nazywany testem wiedzy o czynie lub – nietrafnie – *metodą Kulickiego* (który tylko propagował to, co było już dorobkiem Lykkena). Wśród jego zalet, poza dokładnością na poziomie 82,3%³⁸, należy wymienić dobre podstawy naukowe i łatwość w zrozumieniu. Wadą jest niewielka liczba spraw, w których realnie można go zastosować (najwyżej kilkanaście procent)³⁹. Wynika to z tego, że nie jest łatwo wytypować dobre szczegóły do weryfikacji. Muszą być na tyle istotne, by sprawca przestępstwa dostrzegł je na miejscu zdarzenia i zapamiętał, a jednocześnie aby osoba niewinna nie mogła łatwo domyślić się ich w zestawieniu z wariantami przybrzanymi. Często badania poligraficzne przeprowadza się na późnym etapie postępowania, kiedy osoby postronne z różnych źródeł również mają krytyczną wiedzę, a z drugiej strony – sprawcy pamiętają coraz mniej szczegółów⁴⁰. Dlatego eksperci podkreślają:

³⁸ Zob. APA, *Meta-Analytic Survey of Criterion Accuracy of Validated Techniques*, „Polygraph”, 40(4), 2011, s. 193–305.

³⁹ J.A. Podlesny, *Is the Guilty Knowledge Polygraph Technique Applicable in Criminal Investigations? A Review of FBI Case Records*, „Crime Laboratory Digest”, 20, 1993, s. 57–61.

⁴⁰ Zob. J. Widacki, *W sprawie wyboru techniki badania poligraficznego. Czy technika oparta na testach GKT (CIT) jest lepsza od techniki opartej na testach CQT?*, „Problemy Kryminalistyki”, 273, 2011, s. 5–10.

Najlepszą praktyką dla znalezienia i ochrony tych kluczowych informacji jest włączenie badającego do śledztwa na wczesnym etapie. Kiedy ma on szansę odwiedzenia miejsca przestępstwa lub uczestniczenia w śledztwie od początku, pozwala to na konstrukcję testu CIT ze śledczymi i redukuje prawdopodobieństwo wycieku informacji. To wczesne zaangażowanie pomaga zabezpieczyć krytyczne informacje i zapewnia, że nie będą one ujawnione nikomu poza osobami odpowiedzialnymi za śledztwo. Nabiera to sensu w szerszym kontekście. Nie kończy się śledztwa i następnie wzywa specjalistów kryminalnych na miejsce przestępstwa po kilku miesiącach od zdarzenia w celu zdjęcia odcisków palców i poszukiwania innych fizycznych oraz śladowych dowodów po tym, jak nieznaną liczbą osób potencjalnie dokonała tu kontaminacji⁴¹.

2.2.3. System analizy według Rządu Federalnego USA (tzw. federalny)

W 1961 r. szkoła badań poligraficznych zarządzana przez Wydział Śledztw Kryminalnych amerykańskiej armii (United States Army Criminal Investigation Division) przyjął technikę badawczą i system analizy numerycznej Backstera, wprowadzając wkrótce własne modyfikacje⁴². Zmiany zostały potwierdzone także nową nazwą – odtąd była to technika USAMPS ZCT (która później przekształciła się w Federal ZCT). W formie testu inaczej usytuowano pytania symptomatyczne (por. tab. 3). Wprowadzono także możliwość pytania na *kompleks winy* (pozornie istotne pytanie o fikcyjne przewinienie), które służyło jako ostatnie z pytań porównawczych⁴³.

Wskazane różnice trudno nazwać zasadniczymi. Istotniejsze dotyczyły zasad analizy danych testowych. Przede wszystkim pytanie krytyczne należało porównywać z tym z bezpośrednio sąsiadujących pytań kontrolnych, gdzie reakcje były bardziej znaczące (w każdym kanale zapisu odrębnie). Nieco inny był też wykaz cech zapisów na poligramach, uważanych za diagnostyczne (porównanie w części III, rozdziale 2).

Jeżeli chodzi o skalę oceny numerycznej – najpierw była to skala 7-pozycyjna (jak u Backstera). Później, dla uproszczenia analizy, w systemie federalnym wprowadzono również skalę 3-pozycyjną [-1, 0, +1]. W szerszej +1 pkt należy przypisać subtelnej różnicy, +2 oczywistej, a +/-3 dramatycznej. Ocenę 0 przyznaje się, gdy nie ma

⁴¹ D. Krapohl, J.B. McCloughan, S.M. Senter, *How to Use...*, *op. cit.*, s. 36.

⁴² J.A. Matte, *Forensic Psychophysiology...*, *op. cit.*, s. 260.

⁴³ R.S. Weaver, *The Numerical Evaluation...*, *op. cit.*, s. 98.

istotnych różnic w reakcjach na pytania krytyczne i kontrolne lub zapisy są niediagnostyczne, zakłócone. W węższej skali obowiązuje jedna prosta zasada – *większe jest lepsze*, czyli punktowana jest każda dostrzegalna różnica⁴⁴. W przypadku kanałów pneumo i kardio nie było na początku sztywniejszych ram poza ogólnymi wytycznymi. Dla oceny zmian w reakcjach skórno-galwanicznych wyznaczono minimalne proporcje. Później pojawiły się one również dla kardio. Próg decyzyjny, niezależnie od liczby zarejestrowanych wykresów (minimum dwa), został ustalony na poziomie ± 6 (ocena całkowita w teście typu ZCT).

Jak przy każdej analizie danych najpierw należy ocenić, czy dana reakcja rozpoczęła się w odpowiednim czasie (z reguły od wprowadzenia bodźca testowego) i czy jest wolna od zakłóceń. Następnie ustala się, czy zapis nosi cechy uważane za diagnostyczne i szacuje się różnice w reakcji na pytanie krytyczne względem jednego z sąsiednich w sekwencji pytań porównawczych. J. Swinford, na przykładzie oceny zmian w reakcjach w kanale pneumo, zwraca uwagę, że po każdej reakcji następuje faza kompensacyjna (wyrównania do stanu homeostazy, linii bazowej), a porównywać ze sobą należy wyłącznie te zapisy, które znajdują się w fazie reakcji i jest to jeden z trudniejszych elementów analizy⁴⁵.

Generalnie w systemie federalnym najbardziej skomplikowana jest analiza zapisów w kanale oddechowym. Bierze się to z wielości cech diagnostycznych (12) i kryteriów przypisywania ocen numerycznych w skali 7-pozycyjnej. Dochodzi tu m.in. kryterium wielokrotności występowania różnych cech diagnostycznych, a w przypadku części z reakcji jedna zmiana automatycznie warunkuje inną zmianę, np. wzrost lub spadek amplitudy zwykle powodują również zmianę tempa oddychania i wówczas nie można tego traktować jako mnogiej liczby cech. Znacznie prościej przebiega analiza w kanale reakcji elektrodermalnych (EDA), gdzie zasadniczo bierze się pod uwagę jedną

⁴⁴ Federalna skala 3-pozycyjna nie spełnia obowiązujących standardów praktyki APA – ze względu na przekroczenie w badaniach empirycznych maksymalnego – 20% – limitu wyników nierozstrzygniętych. Skalę tę dopuszcza się do badań pod warunkiem wykorzystania skali 7-pozycyjnej przy braku rozstrzygnięcia po pierwszej ewaluacji w ramach węższej skali.

⁴⁵ Por. J. Swinford, *Manually Scoring Polygraph Charts Utilizing the Seven-Position Numerical Analysis Scale at the Department of Defense Polygraph Institute*, „Polygraph”, 28(1), 1999, s. 15.

cechę diagnostyczną – różnicę amplitud reakcji przy porównywanych pytaniach. Obowiązuje podstawowa reguła: większe jest lepsze, a dla przypisania wartości punktowych innych niż ± 1 wykorzystuje się wytyczne w postaci proporcji (3:1, 4:1). Gdy nie ma dostrzegalnej różnicy w amplitudach, pomocniczo korzysta się również z kryterium złożoności reakcji (np. tzw. *podwójne siodło*, fluktuacje zapisu) i czasu ich trwania. Z kolei w kanale kardio, spośród 8 cech diagnostycznych zapisów, podstawowe znaczenie ma fazowy wzrost krzywej i powrót do linii bazowej (przejściowy wzrost ciśnienia tętniczego i bezpośrednio po nim następujący spadek do względnie stałego stanu równowagi). Aby ujednolicić sposób przyznawania ocen numerycznych innych niż ± 1 , wprowadzono wytyczne, że jedna z dwóch porównywanych reakcji musi być większa o przynajmniej dwie lub trzy pełne podziałki na wykresie, aby można było rozważać ocenę odpowiednio ± 2 lub ± 3 . Szczegółowe zasady oceniania w federalnej skali 7-pozycyjnej przedstawia tab. 5.

Tab. 5. Zasady analizy numerycznej testów CQT (z pytaniami porównawczymi) – zgodnie z systemem Rządu Federalnego USA

	System oceniania <i>federalny</i>
Ogólne wskazówki	<ul style="list-style-type: none"> • skala 7-poz.: 0 – reakcje jednakowe bez dostrzegalnej różnicy lub brak reakcji do porównania z powodu artefaktów, +1/-1 – subtelna, niewielka różnica, +2/-2 – znaczna, wyraźna różnica, +3/-3 – dramatyczna różnica. • ocenę dodatnią dla danego pytania krytycznego przypisuje się, gdy bardziej znacząca zmiana w reakcjach fizjologicznych występuje przy porównywanym pytaniu kontrolnym (jeśli z pytaniem krytycznym sąsiadują dwa pytania kontrolne, bierze się pod uwagę to pytanie, przy którym zmiany w reakcjach są bardziej istotne – odrębnie w każdym kanale zapisu). Natomiast ocenę ujemną należy przypisać wówczas, gdy istotniejsze zmiany w reakcjach zaistniały przy pytaniu krytycznym; • od 3 do 5 wykresów (gdy po 3 i ewentualnie również po 4 wykresach uzyskamy wynik INC); • dopuszcza się skalę 3-poz. pod warunkiem zastosowania skali 7-poz. przy wyniku nierozstrzygniętym po pierwszej ewaluacji;

System oceniania <i>federalny</i>	
Pneumo	<ul style="list-style-type: none"> • 0 – gdy brak cech diagnostycznych przy porównywanych pytaniach albo występuje jedna równorzędna cecha diagnostyczna przy obu pytaniach; • zazwyczaj wartości +1/-1, gdy mamy pojedynczą cechę diagnostyczną przy jednym z porównywanych pytań, a żadna nie występuje przy drugim; albo gdy przy obu pytaniach występuje równorzędna cecha zapisu, ale jedna z tych reakcji trwa dłużej, a także gdy przy jednym pytaniu mamy pojedynczą cechę, a przy drugim jest ich więcej; • rzadko +2/-2, przy wielokrotności cech diagnostycznych przy jednym pytaniu i braku takich cech przy drugim; • wyjątkowo +3/-3, przy dramatycznej różnicy (wielokrotne cechy diagnostyczne przy wydłużonym czasie reakcji) – z reguły tylko wtedy, gdy tak znaczące zmiany występują przy pytaniu krytycznym; • w przypadku zaobserwowania na wykresie dwóch równorzędnych cech diagnostycznych mierzymy przedział czasowy dłuższej z reakcji i porównujemy długość krzywej (RLL) w takim samym przedziale dla obu pytań testowych.
EDA	<ul style="list-style-type: none"> • 0 – gdy brak cech diagnostycznych przy porównywanych pytaniach albo występuje jedna równorzędna cecha diagnostyczna przy obu pytaniach, a także gdy stosunek amplitud jest mniejszy niż 2:1, a jednocześnie reakcja z mniejszą amplitudą ma charakter złożony lub trwa istotnie dłużej; • 1 = zasada <i>większe jest lepsze</i> – dostrzegalna różnica i stosunek amplitudy < 3:1; drugorzędym kryterium jest złożoność reakcji, która ma zastosowanie przy jednakowych amplitudach; z kolei za trzeciorzędne kryterium uważa się czas trwania reakcji (przy zbliżonych amplitudach i złożoności reakcji); • 2 = stosunek amplitudy > 3:1 < 4:1; • 3 = stosunek amplitudy ≥ 4:1; • jeśli przy jednym z porównywanych pytań w ogóle nie ma reakcji, obowiązuje reguła z ilością odstępów (kratek, podziałek) na siatce poligramu: <ul style="list-style-type: none"> 1 = do 2 odstępów, 2 = od 2 do 3 odstępów, 3 = powyżej 3 odstępów.
Kardio	<ul style="list-style-type: none"> • 1 = do 2 razy większy wzrost linii bazowej reakcji; • 2 = od 2 do 3 razy większa reakcja; • 3 = przynajmniej 3 razy większa reakcja; • jeśli przy jednym z porównywanych pytań w ogóle nie ma reakcji obowiązuje reguła z ilością odstępów (kratek) na siatce poligramu: <ul style="list-style-type: none"> 1 = do 2 odstępów, 2 = od 2 do 3 odstępów, 3 = powyżej 3 odstępów.
PPG	<ul style="list-style-type: none"> • brak wytycznych

	System oceniania <i>federalny</i>
Progi decyzyjne*	<ul style="list-style-type: none"> • jednakowe progi decyzyjne dla skali 7- i 3-poz.; • dla testów jednoprotblemowych typu ZCT: <ul style="list-style-type: none"> DI – gdy suma całkowita ≤ -6 lub gdy suma cząstkowa (przy którymkolwiek z pytań krytycznych, czyli tzw. <i>spocie</i>) ≤ -3, NDI – gdy suma w każdym ze spotów $\geq +1$ i suma globalna $\geq +6$ ($\geq +4$ w skali 7-poz. <i>dowodowej</i>), INC – pozostałe wyniki. • dla testu jednoprotblemowego You-Phase (Bi-Zone): <ul style="list-style-type: none"> DI – gdy suma całkowita ≤ -4 lub gdy którakolwiek suma cząstkowa ≤ -3, NDI – gdy suma w każdym ze spotów $\geq +1$ i suma globalna $\geq +4$, INC – pozostałe wyniki. • w testach wieloaspektowych i wieloprotblemowych typu MGQT: <ul style="list-style-type: none"> SR – gdy suma cząstkowa dla któregoś z pytań krytycznych ≤ -3, NSR – gdy każda suma cząstkowa $\geq +3$, INC – pozostałe wyniki. • dla testów przesiewowych typu DLST: <ul style="list-style-type: none"> SR – gdy suma całkowita ≤ -4 lub gdy którakolwiek suma cząstkowa ≤ -3 NSR – gdy suma w każdym ze spotów $\geq +1$ i suma globalna $\geq +4$ INC – pozostałe wyniki.

* Wyniki testów oznaczane są następującymi skrótami: DI (z ang. *deception indicated*) – stwierdzono wprowadzanie w błąd; NDI (z ang. *no deception indicated*) – nie stwierdzono wprowadzania w błąd; SR (z ang. *significant responses*) – znaczące reakcje; NSR (z ang. *no significant responses*) – brak znaczących reakcji; INC (z ang. *inconclusive*) – nierozstrzygnięte. Szerzej na temat znaczenia poszczególnych wyników przy wydawaniu opinii zob. M. Gołaszewski (red.), *Współczesne standardy...*, op. cit., cz. III, rozdz. 2.

Źródło: opracowanie własne autora.

2.2.4. System UTAH

Od lat 70. XX wieku na Uniwersytecie Utah pracowano nad doskonaleniem technik badawczych i metod analizy danych z badań poligraficznych, co po 30 latach zaowocowało usystematyzowanym podejściem, nazywanym Utah – od miejsca, w którym pracowali naukowcy (przede wszystkim: D. Raskin, J. Kircher, C. Honts). Składają się na nie m.in. reguły prowadzenia wywiadu, testy pytań porównawczych (w wersji z trzema pytaniami relewantnymi – ZCT⁴⁶ i MGQT, zawierającym do czterech takich pytań), a także system oceniania zapisów

⁴⁶ Szerzej na temat testu jednoprotblemowego Utah ZCT zob.: M. Handler, R. Nelson, *Utah Approach...*, op. cit.; M. Gołaszewski et al., *Metodyka przeprowadzania testu Utah ZCT*, [w:] J. Widacki (red.), *Kierunki rozwoju...*, op. cit., s. 70–74.

na poligramach. System ten miał być mniej skomplikowany, bardziej przejrzysty względem dotychczasowych, zapewniający maksymalną dokładność diagnoz i minimalny poziom rozbieżności między poligraferami.

Względem systemów Backstera i federalnego zmniejszyła się liczba cech zapisów uznawanych, na podstawie badań empirycznych, za diagnostyczne (szerzej na ten temat w części III, rozdz. 2). Kryteria punktowania ustanowiono na tyle konserwatywnie, aby brane były pod uwagę tylko wyraźne różnice w reakcjach. Do trzech tradycyjnych kanałów zapisu (pneumo, EDA, kardio) dołączono jeszcze kanał reakcji naczynioruchowych (PLE). Wprowadzono zasadę, aby w testach typu ZCT reakcje na pytanie krytyczne za każdym razem porównywać z reakcjami na poprzedzające pytanie kontrolne (co miało ograniczać subiektywizm przy wyborze odpowiedniego pytania do porównania)⁴⁷. Natomiast w testach typu MGQT – gdzie pytanie krytyczne otoczone jest w sekwencji dwoma pytaniami kontrolnymi – wciąż korzystać z obu i brać pod uwagę ostatecznie to pytanie kontrolne, przy którym w danym parametrze zmiany reakcji są istotniejsze.

W systemie analizy Utah wykorzystuje się 7-pozycyjną skalę ocen $<-3, +3>$. Ocenę zero przypisujemy, gdy zmiany reakcji są zbliżone lub nie ma takich zapisów, które można by ze sobą porównać; $+1/-1$ – gdy występuje zauważalna różnica; $+2/-2$ – duża, wyraźna różnica; $+3/-3$ – dramatyczna różnica, zapis krzywej jest jednoznaczny i jednocześnie mamy do czynienia z najsilniejszą reakcją na wykresie. Zatem te skrajne oceny można przypisać w danym kanale zapisu tylko raz. Dla kanałów EDA i kardio określono proporcje wielkości zmian w reakcjach, które decydują o tym, jaką ocenę ze skali zastosować.

Z badań amerykańskich naukowców wynika, że oceny numeryczne o wartości zerowej występują znacznie częściej niż pozostałe – w szczególności w kanałach penumo i PLE: ponad 70% oraz w ok. połowie przypadków w kanałach EDA i kardio. Jednocześnie w kanale EDA występowało więcej ocen $+/-2$ i $+/-3$. Świadczy to o tym, że w systemie Utah ten kanał zapisu wywiera największy wpływ na końcowy rezultat testu⁴⁸.

⁴⁷ Zob. M.D. Handler, *Utah Probable Lie Comparison Test*, „Polygraph”, 35(3), 2006, s. 144.

⁴⁸ Zob. B.G. Bell, D.C. Raskin, C.R. Honts, J.C. Kircher, *The Utah Numerical Scoring System*, „Polygraph”, 28(1), 1999, s. 7.

W testach jednoprobemowych jedynym kryterium decyzyjnym jest numeryczna ocena całkowita zmian reakcji przy pytaniach krytycznych. W wieloaspektowych ocena przebiega dwuetapowo i brane są pod uwagę jeszcze oceny cząstkowe – dla pojedynczych pytań relevantnych (ze względu na fakt, że przynajmniej jedno z pytań krytycznych nie jest zależne treściowo od pozostałych i wiarygodność odpowiedzi badanego może być, w zależności od pytania, niejednakowa). Z kolei w testach wieloprobemowych tylko oceny cząstkowe decydują o wyniku testu. Szczegółowe wytyczne przedstawiono w tab. 6.

Tab. 6. Zasady analizy numerycznej testów CQT – zgodnie z systemem Uniwersytetu w Utah

	System oceniania UTAH
Ogólne wskazówki	<ul style="list-style-type: none"> • skala 7-poz.: 0 – jednakowe lub brak reakcji do porównania z powodu artefaktów, +1/-1 – zauważalna różnica, +2/-2 – silna, wyraźna różnica, +3/-3 – dramatyczna różnica, jednoznaczny zapis krzywej i jednocześnie najsilniejsza reakcja na wykresie, • ocenę dodatnią dla danego pytania krytycznego przypisuje się, gdy bardziej znacząca zmiana w reakcjach fizjologicznych występuje przy porównywanym pytaniu kontrolnym (poprzedzającym pytanie krytyczne). Natomiast ocenę ujemną należy przypisać wówczas, gdy istotniejsze zmiany w reakcjach zaistniały przy pytaniu krytycznym, • 3 lub 5 wykresów (gdy po 3 wykresach uzyskamy wynik INC).
Pneumo	<ul style="list-style-type: none"> • zazwyczaj wartości +1/-1, bardzo rzadko +2/-2, nigdy nie przypisujemy wartości +3/-3, w przypadku stwierdzenia dwóch równorzędnych cech diagnostycznych zwracamy uwagę na czas trwania reakcji (odcinek krzywej do porównania musi się znajdować w przedziale od wprowadzenia bodźca do 10 kolejnych sekund).
EDA	<ul style="list-style-type: none"> • 1 = podwójna różnica w amplitudzie, ewentualnie stosunek 1,5:1 + czas trwania lub złożoność, • 2 = potrójna różnica w amplitudzie, ewentualnie stosunek 2,5:1 + czas trwania lub złożoność, • 3 = poczwórna różnica w amplitudzie, jednoznaczny zapis krzywej i jest to najsilniejsza reakcja na wykresie.
Kardio	<ul style="list-style-type: none"> • 1 = stosunek wielkości reakcji 1,5:1, • 2 = stosunek 2:1, • 3 = stosunek 3:1 i jest to najsilniejsza reakcja na wykresie.

	System oceniania UTAH
PPG	<ul style="list-style-type: none"> • oceny 1 lub 2, które można przypisać także, gdy nie ma różnicy w wielkości redukcji amplitudy, ale istnieje wyraźna różnica w czasie trwania reakcji.
Progi decyzyjne	<ul style="list-style-type: none"> • w testach jednoprotymowych (ZCT): <ul style="list-style-type: none"> DI – gdy suma całkowita ≤ -6, NDI – gdy suma całkowita $\geq +6$, INC – pozostałe wyniki, • w testach wieloaspektowych (ZCT z trzema pytaniami relewantnymi, MGQT z czterema pytaniami relewantnymi): <ul style="list-style-type: none"> SR – gdy suma całkowita ≤ -6 i każda suma cząstkowa jest ujemna lub gdy suma cząstkowa dla któregośkolwiek z pytań relewantnych ≤ -3, NSR – gdy suma całkowita $\geq +6$ i każda suma cząstkowa jest dodatnia, INC – pozostałe wyniki. • wieloprotymowych (typu MGQT – od dwóch do czterech pytań relewantnych): <ul style="list-style-type: none"> SR – gdy suma cząstkowa dla któregośkolwiek z pytań relewantnych ≤ -3, NSR – gdy każda suma cząstkowa $\geq +3$, INC – pozostałe wyniki.

Źródło: opracowanie własne autora.

2.2.5. Empiryczny System Oceniania (ESS) i jego zaktualizowana wersja (ESS-M)

Te same cele, co twórcom systemu Utah (a więc m.in. zapewnienia maksymalnej dokładności diagnozowania, zgodności poligraferów, obiektywizmu), przyświecały innej grupie naukowców (R. Nelson, D. Krapohl, M. Handler), która opracowała metodę nazwaną Empirycznym Systemem Oceniania (z j. ang. *empirical scoring system*), a pierwsze publikacje na temat nowego systemu zaczęły się pojawiać od 2008 r.⁴⁹ Chodziło przy tym także, aby system był jak najmniej skomplikowany, przystępny dla poligraferów bez względu na poziom doświadczenia i możliwy do ewentualnego skomputeryzowania. U podstaw tej procedury analizy zapisów poligraficznych miały leżeć wyłącznie dane z badań empirycznych i publikacji naukowych, bez czynienia arbitralnych założeń.

⁴⁹ Zob. R. Nelson, D.J. Krapohl, M. Handler, *Brute Force Comparison. A Monte Carlo Study of the Objective Scoring System version 3 (OSS-3) and Human Polygraph Scorers*, „Polygraph”, 37(3), 2008, s. 185–215.

Zatem, historycznie kolejny już raz, zredukowano liczbę cech zapisów na poligramach, zweryfikowanych jako diagnostyczne. Tym razem do sześciu, choć część została sprowadzona po prostu do jednej szerszej kategorii, jak w przypadku supresji oddechowej (szczegółowo w części III, podrozdziale 2.2.5). Zdecydowano się na teoretycznie bardziej przejrzystą i obiektywną skalę ocen zapisów, bo jedynie 3-pozycyjną (-1, 0, +1), przy jednoczesnym podwójnym ważeniu parametru EDA (+/-2). Szczegółowe zasady oceniania przedstawia tab. 13.

W początkowym etapie rozwoju technik badań poligraficznych i metod interpretacji danych testowych (ograniczonych jeszcze niemal wyłącznie do metody jakościowej) wkład w wartość decyzyjną poszczególnych parametrów był raczej zbliżony, a nawet z lekkim wskazaniem na kanał oddechowy. W eksperymencie przeprowadzonym przez S. Slowika i J. Buckleya poligraferzy wydający diagnozę na podstawie jednego parametru byli w 80,5% dokładni, gdy polegali na respiracji; w 77,1%, gdy posłużyli się wyłącznie kanałem kardio, a w 80%, gdy brali pod uwagę tylko kanał GSR⁵⁰. Takie analizy prowadził też G. Barland – dokładność wskazań w kanale pneumo wyniosła 72,8%, w kanale GSR – 78,3%, a kardio – 70,2%⁵¹.

Z kolei w badaniach B. Jayne parametr oddechowy był miarodajny w 84%, GSR/EDA – w 85%, a kardio – 77%. Autor niniejszej monografii sprawdzał również, czy – biorąc pod uwagę różną korelację ocen w poszczególnych parametrach z poprawną diagnozą – ogólna dokładność testu zmieniałaby się, gdyby parametry te były odpowiednio wżone przy analizie numerycznej poprzez pomnożenie wyników o ich indywidualną dokładność. Trafność identyfikacji bez korekty i po niej okazała się niemal jednakowa⁵². Gdy w praktyce poligraferów zaczęły dominować numeryczne metody analizy danych testowych, okazało się jednak, że komponent GSR/EDA daje średnio aż ok. 50-procentowy wkład przy podejmowaniu przez poligrafera decyzji o diagnozie, co uzasadnia podwojenie jego wagi⁵³.

⁵⁰ S.M. Slowik, J.P. Buckley, *Relative Accuracy of Polygraph Examiner Diagnosis of Respiration, Blood Pressure, and GSR Recordings*, „Journal of Police Science and Administration”, 3(3), 1975, s. 306.

⁵¹ Zob. G.H. Barland, *The Reliability of Polygraph...*, *op. cit.*, s. 200.

⁵² B.C. Jayne, *Contributions of Physiological Recordings in the Polygraph Technique*, „Polygraph”, 19(2), 1990, s. 105–117.

⁵³ Por. R. Nelson, *Literature Survey of Structural Weighting of Polygraph Signals. Why Double the EDA?*, „Polygraph & Forensic Credibility Assessment”, 48(2), 2019, s. 105–112.

W podejściu ESS dane testowe interpretuje się na podstawie wizualnej reprezentacji danych testowych na poligramach poprzez rozpoznanie określonych cech diagnostycznych. Potwierdzono, że system sprawdza się zarówno, gdy jest używany przez poligraferów z dłuższym stażem, jak i tych mniej doświadczonych⁵⁴. Badania walidacyjne prowadzono na populacjach z różnych kręgów kulturowych – poza USA m.in. w Izraelu, Singapurze, Kolumbii, Kanadzie czy Meksyku⁵⁵.

Dzięki uzyskanym empirycznie danym referencyjnym zaistniała możliwość wyznaczenia statystycznej wartości p , która jest prawdopodobieństwem uzyskania określonego wyniku lub jeszcze bardziej ekstremalnego od hipotezy zerowej, która zakładałaby, że dla tego wyniku nie ma różnicy między osobami odpowiadającymi szczerze i nieszczerze na pytania krytyczne. Im niższa wartość p – tym niższe prawdopodobieństwo prawdziwości hipotezy zerowej, przy czym wartości p nie trzeba za każdym razem obliczać, ale bierze się je z dostępnych tablic (zob. tab. 7–10).

Tab. 7. Progi decyzyjne i wartości p dla testów typu ZCT z trzema pytaniami krytycznymi, zgodnie z systemem ESS

Testy jednoprotokolowe typu ZCT (wszystkie formaty z 3 pytaniami krytycznymi)			
Progi decyzyjne – prawdomówność		Progi decyzyjne – wprowadzanie w błąd	
Na podstawie dystrybucji całociowych ocen osób wprowadzających w błąd		Na podstawie dystrybucji całociowych ocen osób prawdomównych	
Próg decyzyjny	Wartość p	Próg decyzyjny	Wartość p
-1	0,159	1	0,159
0	0,130	0	0,127
1	0,106	-1	0,099
2	0,085	-2	0,077
3	0,067	-3	0,058

⁵⁴ Zob. B. Blalock, B. Cushman, R. Nelson, *A Replication and Validation Study on an Empirically Based Manual Scoring System*, „Polygraph”, 38(4), 2009, s. 281–286; R. Nelson, D. Krapohl, *Criterion Validity of the Empirical Scoring System with Experienced Examiners. Comparison with the Seven-Position Evidentiary Model Using the Federal Zone Comparison Technique*, „Polygraph”, 40(2), 2011, s. 79–85.

⁵⁵ M. Handler, R. Nelson, W. Goodson, M. Hicks, *Empirical Scoring System. A Cross-cultural Replication and Extension Study of Manual Scoring and Decision Policies*, „Polygraph”, 39(4), 2010, s. 200–215.

4	0,052	-4	0,043
5	0,040	-5	0,032
6	0,030	-6	0,023
7	0,023	-7	0,016
8	0,017	-8	0,011
9	0,012	-9*	0,008
10*	0,008	-10	0,005
11	0,006	-11	0,003
12	0,004	-12	0,002
13	0,003	-13	0,001
14	0,002	-14	< 0,001
15	0,001	-15	< 0,001
16	< 0,001	-16	< 0,001

Źródło: opracowanie własne na podst. R. Nelson, M. Handler, *Empirical Scoring System. NPC Quick Reference*, Lafayette Instrument, 2010.

Na podstawie danych z powyższej tabeli możemy przedstawić przykładowo, że gdy poligrafer podejmie decyzję o wyniku NDI (braku wskazania na wprowadzanie w błąd) po otrzymaniu sumy całkowitej ocen numerycznych (+10) – oznacza to, że zaledwie 0,8% badanych z referencyjnej populacji osób odpowiadających nieszczerze uzyskało taki lub wyższy rezultat testu. Z kolei, gdy wynik zostaje ustalony z powodu jednej z sum cząstkowych, np. (-9) – który wskazuje na wprowadzanie w błąd (DI) – zastosowanie ma poprawka Bonferoniego. Alfa wynosi wówczas 0,017. Wartość $p = 0,008$ pomnożona przez liczbę ocen cząstkowych w teście – czyli przez 3 – daje ułamek 0,024. Z tego wiemy, że jedynie 2,4% badanych z referencyjnej populacji osób odpowiadających szczerze na pytania krytyczne uzyskało w przedmiotowym teście ocenę cząstkową (-9) lub mniejszą.

Tab. 8. Progi decyzyjne i wartości p dla testów typu ZCT z dwoma pytaniami krytycznymi (You-Phase / Bi-Zone), zgodnie z systemem ESS

Test jednoprotoblowy You-Phase (Bi-Zone)			
Progi decyzyjne – prawdomówność		Progi decyzyjne – kłamstwo	
Na podstawie dystrybucji całościowych ocen osób wprowadzających w błąd		Na podstawie dystrybucji całościowych ocen osób prawdomównych	
Próg decyzyjny	Wartość p	Próg decyzyjny	Wartość p
-1	0,202	1	0,202
0	0,159	0	0,159

1	0,122	-1	0,122
2	0,091	-2	0,091
3	0,067	-3	0,067
4	0,048	-4	0,048
5	0,033	-5	0,033
6	0,023	-6	0,023
7	0,015	-7	0,015
8	0,010	-8	0,010
9	0,006	-9	0,006
10	0,004	-10	0,004
11	0,002	-11	0,002
12	0,001	-12	0,001
13	< 0,001	-13	< 0,001

Źródło: opracowanie własne na podst. R. Nelson, M. Handler, *Empirical Scoring System...*, op. cit.

Tab. 9. Progi decyzyjne i wartości p w systemie ESS dla testów przesiewowych typu MGQT, DLST przy ujemnych sumach cząstkowych

Wieloprotblemowe testy przesiewowe (testy typu MGQT, DLST) Wprowadzanie w błąd – którakolwiek ocena cząstkowa (≤ -3)	
Progi decyzyjne kłamstwa na podstawie dystrybucji ocen cząstkowych osób prawdomównych	
Próg decyzyjny	Wartość p (testy z 2, 3 lub 4 pyt. krytycznymi)
1	0,369
0	0,252
-1	0,159
-2	0,091
-3	0,048
-4	0,023
-5	0,010
-6	0,004
-7	0,001
-8	< 0,001

Źródło: opracowanie własne na podst. R. Nelson, M. Handler, *Empirical Scoring System...*, op. cit.

Tab. 10. Progi decyzyjne i wartości p w systemie ESS dla testów przesiewowych, zawierających od dwóch do czterech pytań krytycznych, przy dodatnich sumach cząstkowych

Wieloprotblemowe testy przesiewowe (typu MGQT, DLST) Prawdopodobność – wszystkie oceny cząstkowe (dodatnie) (progi decyzyjne w oparciu o dystrybucję ocen cząstkowych osób wprowadzających w błąd)		
Wartość p – 2 pytania krytyczne		
Próg decyzyjny	Nieskorygowana	Skorygowana
0	0,252	0,135
1	0,159	0,083
2	0,091	0,047
3	0,048	0,024
4*	0,023	0,011
5	0,010	0,005
6	0,004	0,002
7	0,001	0,001
8	0,000	< 0,001
Wartość p – 3 pytania krytyczne		
Próg decyzyjny	Nieskorygowana	Skorygowana
0	0,252	0,092
1	0,159	0,056
2	0,091	0,031
3	0,048	0,016
4	0,023	0,008
5	0,010	0,003
6	0,004	0,001
7	0,001	< 0,001
Wartość p – 4 pytania krytyczne		
Próg decyzyjny	Nieskorygowana	Skorygowana
0	0,252	0,070
1	0,159	0,042
2	0,091	0,024
3	0,048	0,012
4	0,023	0,006
5	0,010	0,002
6	0,004	0,001
7	0,001	< 0,001

Źródło: opracowanie własne na podst. R. Nelson, M. Handler, *Empirical Scoring System...*, op. cit.

Zakładając, że test przesiewowy typu DLST z dwoma pytaniami krytycznymi zakończył się wynikami, spośród których najniższa ocena cząstkowa dla pojedynczego pytania krytycznego wyniosła (+4), interpretuje się je jako brak wskazania na nieszczerłość. Korzystając z powyższej tabeli, można zauważyć, że skorygowana wartość $p = 0,011$. To oznacza, że zaledwie 1,1% badanych z referencyjnej populacji odpowiadających nieszczerze uzyskało taką lub wyższą ocenę cząstkową⁵⁶.

Wynik testu uznaje się za rozstrzygnięty wtedy, kiedy jest statystycznie istotny, a sprawdzimy to poprzez porównanie wartości p z przyjętym poziomem istotności, oznaczanym jako α . Wysoka wartość p (wykraczająca poza akceptowaną tolerancję błędów – α) oznacza, że hipotezy zerowej nie można odrzucić – albo dlatego, że jest prawdziwa; albo jest fałszywa, ale wyniki analityczne uzyskane w toku badania w odniesieniu do danych referencyjnych nie pozwalają na jej odrzucenie. W tym ostatnim przypadku mówimy wówczas o błędzie drugiego rodzaju.

Poziom istotności wyznacza się w zasadzie arbitralnie. Zwykle w naukach społecznych przyjmuje się, że jest to 0,05. W przypadku wyników wskazujących na reakcje typowe dla osób odpowiadających nieszczerze na pytania krytyczne tolerancja błędów wynosi właśnie 0,05. Z kolei w przypadku rezultatów wskazujących na reakcje typowe dla osoby prawdopodobnej próg ten jest z reguły wyższy i ustanowiony na poziomie $\alpha = 0,1$. Wynika to przede wszystkim z potrzeby ograniczenia liczby nierozstrzygniętych wyników testów. Ponadto dla decyzji opartych na sumach cząstkowych wprowadzono statystyczną poprawkę Bonferrońskiego w celu zmniejszenia ryzyka fałszywych wyników pozytywnych. W ten sposób przy teście typu ZCT z trzema pytaniami krytycznymi alfa dzieli się przez 3 – co daje nam skorygowaną wartość $\alpha = 0,017$, a gdy mamy dwa pytania krytyczne, alfa jest dzielona przez 2, co daje $\alpha = 0,025$.

Obranie konkretnych poziomów istotności rzutuje na punktowe progi decyzyjne przy numerycznej analizie zapisów na poligramach. Ostrzejsze kryterium zwykle przyjmuje się w przypadku wyników niekorzystnych dla osoby badanej ze względu na negatywne skutki, który taki wynik może dla niej wywołać – np. poprzez dowód obciążający w sprawie karnej. Wszystko zależy, w gruncie rzeczy, od celu badania. Gdy w grę wchodzi np. bezpieczeństwo państwa – zaakceptowanie bardziej liberalnego poziomu istotności dla wyników uznawanych za

⁵⁶ Szerzej o tym, jak korzystać z systemu ESS zob. R. Nelson *et al.*, *Using the Empirical Scoring...*, *op. cit.*

potwierdzające wiarygodność oświadczeń kandydata do służby specjalnej może być już mniej zasadne i wówczas na szali trzeba położyć z jednej strony dobro jednostki, a z drugiej interes państwa.

Przykład konkluzji, wynikającej z zastosowania systemu ESS
<p>Załóżmy, że ewaluacji poddajemy dane z testu przesiewowego typu MGQT z czterema pytaniami krytycznymi. Analiza, zgodna z regułami ESS, przyniosła następujące oceny zmian reakcji przy tych pytaniach – R1: +7; R2: +4; R3: +9; R4: –6. Opis rezultatu testu może przedstawiać się wówczas następująco:</p>
<p><i>SR – znaczące reakcje. Oznacza to, że badany reagował na pytania krytyczne w sposób, jaki zwykle obserwuje się u osób odpowiadających nie-szczerze (niezgodnie ze swoją faktyczną wiedzą – wprowadzając w błąd lub zatajając istotne okoliczności, których dotyczą pytania krytyczne).</i></p>
<p><i>Znaczenie statystyczne uzyskanego wyniku testu, na podstawie najniższej oceny cząstkowej (–6 dla pytania R4), wynosi 0,004 – czyli poniżej przyjętej przy tego rodzaju opinii (SR) tolerancji błędu ($\alpha = 0,05$). Odnosząc się więc do danych referencyjnych – identyczny lub niższy wynik numeryczny dla pojedynczego pytania krytycznego (w tego typu teście, jaki został wykorzystany) odnotowano zaledwie w 0,4% przypadków w populacji osób odpowiadających szczerze.</i></p>

Zaktualizowaną wersją systemu ESS jest ESS-M (z j. ang. *empirical scoring system-multinomial*), który wykorzystuje we wnioskowaniach statystycznych teoremat Bayesa. Jest to rodzaj wnioskowania subiektywnego, w którym prawdopodobieństwo określa

miarę naszego przekonania co do prawdziwości danego zdarzenia [...]. Zagadnienie Bayesa to problem, w którym należy obliczyć prawdopodobieństwo zajścia przyczyny na podstawie informacji o prawdziwości zdarzenia D, czyli skutku⁵⁷.

Posługujemy się tu kategorią szansy, która opisuje, ile razy bardziej prawdopodobna jest prawdziwość danej hipotezy niż jej fałszywość. Innymi słowy, mowa tu o ilorazie wiarygodności (współczynnika Bayesa), pozwalającym na porównanie, w jakim stopniu dane świadczą na rzecz dwóch wzajemnie wykluczających się hipotez – czyli ile razy

⁵⁷ Zob. P. Wolańska-Nowak, *Interpretacja wyników ekspertyzy*, [w:] J. Wójcikiewicz (red.), *Ekspertyza sądowa. Zagadnienia wybrane*, Warszawa 2007, s. 576.

hipoteza alternatywna (np. znaczące zmiany reakcji fizjologicznych wskazujące na wprowadzanie w błąd przy pytaniach krytycznych) jest bardziej prawdopodobna niż hipoteza zerowa (zarejestrowane przez poligraf dane zaistniały u osoby prawdomównej).

Z kolei szansa na rzecz określonego zdarzenia Z , uwzględniająca uzyskany dowód (np. opinię z zakresu badań poligraficznych wskazującą, że badany ma wiedzę o szczegółach rozpatrywanego zdarzenia, chociaż temu przeczył) – to tzw. szansa *a posteriori* (ryc. 16 – po lewej stronie równania). Aby to oszacować, musimy przyjąć jakieś subiektywne założenie *a priori* odnośnie do tego zdarzenia – ale nie zupełnie dowolnie, tylko – w miarę możliwości – na podstawie informacji zgromadzonych w toku danego postępowania, które są niezależne od dowodu D (ryc. 16 – szansę *a priori* obrazuje pierwszy ułamek po prawej stronie równania). Kombinujemy to z ilorzem wiarygodności dowodu D , który zależy od dwóch prawdopodobieństw warunkowych: ustalenia danej zgodności, dowodu D – przy założeniu, że zdarzenie Z rzeczywiście nastąpiło i ustalenia tego dowodu przy założeniu przypadkowego wyniku badań (ryc. 16 – ilorzec wiarygodności przedstawia drugi ułamek po prawej stronie równania).

Ryc. 16. Wyrażenie szansy *a posteriori* zgodnie z prawem Bayesa

$$\frac{P(Z|D)}{P(\text{nie}Z|D)} = \frac{P(Z)}{P(\text{nie}Z)} \times \frac{P(D|Z)}{P(D|\text{nie}Z)}$$

R. Nelson w 2017 r. opublikował wielomianowe tablice referencyjne dla systemu ESS, stanowiące matematyczne i logiczne wyrażenie hipotezy zerowej wobec podstawowej teorii badań poligraficznych, prowadzącej się w uproszczeniu do tego, że istotniejsze zmiany w reakcjach fizjologicznych osoby badanej występują przy określonej kategorii bodźców testowych (krytycznych, albo kontrolnych) jako funkcja wprowadzania w błąd lub prawdomowości przy bodźcach krytycznych⁵⁸. Te dane referencyjne dają informację o tym, jakie jest prawdopodobieństwo wystąpienia wszelkich możliwych wyników testu, jeśli teoria okazałaby się nietrafna, a zmiany reakcji w poszczególnych kanałach zapisu zachodziły w sposób przypadkowy. Obliczenia tych wyników dla testów typu CQT rozpoczynają się od dystrybucji wyników dla każdego z kanałów

⁵⁸ R. Nelson, *Multinomial Reference Distributions for the Empirical Scoring System*, „Polygraph & Forensic Credibility Assessment”, 46(2), 2017, s. 81–115.

zapisu, co jest funkcją możliwych ocen dla wszystkich prezentacji krytycznego bodźca testowego. Warto mieć na uwadze, że pojedynczy kanał zapisu może dostarczyć wielką liczbę kombinacji ocen – od $3^6 = 729$ dla trzech powtórzeń dwóch pytań krytycznych do $3^{20} = 3486784401$ dla pięciu powtórzeń czterech pytań krytycznych⁵⁹. Następnie potrzebne są bardziej skomplikowane działania matematyczne, aby otrzymać wielomianową dystrybucję wyników. Jest to później przydatne jako funkcja prawdopodobieństwa w analizie według reguł Bayesa, prowadzącej do ustalenia, w jaki sposób rezultat testu wspiera hipotezę o prawdomówności lub nieszczerości osoby badanej⁶⁰.

Tab. 11. Wielomianowe progi decyzyjne w zaktualizowanym systemie ESS dla testów jednoprotblemowych

Prior odds of deception	prior probability	Alpha (truth/deception)									
		.01/.01	.01/.05	.01/.10	.05/.01	.05/.05	.05/.10	.10/.01	.10/.05	.10/.10	
9 to 1 (9 in 10)	.90	+14/-9 (none)	+14/-6 (none)	+14/-4 (-21)	+13/-9 (none)	+13/-6 (none)	+13/-4 (-21)	+13/-9 (none)	+13/-6 (none)	+13/-4 (-21)	
8 to 1 (8 in 9)	.89	+13/-8 (none)	+13/-5 (none)	+13/-4 (-20)	+13/-8 (none)	+13/-5 (none)	+13/-4 (-20)	+12/-8 (none)	+12/-5 (none)	+12/-4 (-20)	
7 to 1 (7 in 8)	.88	+13/-7 (none)	+13/-5 (-22)	+13/-4 (-18)	+12/-7 (none)	+12/-5 (-22)	+13/-4 (-18)	+12/-7 (none)	+12/-5 (-22)	+12/-4 (-18)	
6 to 1 (6 in 7)	.86	+12/-7 (none)	+12/-4 (-20)	+12/-4 (-16)	+11/-7 (none)	+11/-4 (-20)	+11/-4 (-16)	+11/-7 (none)	+11/-4 (-20)	+11/-4 (-16)	
5 to 1 (5 in 6)	.83	+11/-6 (none)	+11/-4 (-17)	+11/-3 (-15)	+10/-6 (none)	+10/-4 (-17)	+10/-3 (-15)	+10/-6 (none)	+10/-4 (-17)	+10/-3 (-15)	
4 to 1 (4 in 5)	.80	+10/-5 (-21)	+10/-4 (-15)	+10/-3 (-14)	+10/-5 (-21)	+10/-4 (-15)	+10/-3 (-14)	+9/-5 (-21)	+9/-4 (-15)	+9/-3 (-14)	
3 to 1 (3 in 4)	.75	+9/-5 (-16)	+9/-3 (-13)	+9/-3 (-12)	+9/-5 (-16)	+8/-3 (-13)	+8/-3 (-12)	+8/-5 (-16)	+8/-3 (-13)	+8/-3 (-12)	
2 to 1 (2 in 3)	.67	+7/-4 (-13)	+7/-3 (-11)	+7/-2 (-10)	+6/-4 (-13)	+6/-3 (-11)	+6/-2 (-10)	+6/-4 (-13)	+6/-3 (-11)	+6/-2 (-10)	
1 to 1 (1 in 2)	.50	+4/-4 (-9)	+4/-3 (-7)	+4/-2 (-6)	+3/-4 (-9)	+3/-3 (-7)	+3/-2 (-6)	+2/-4 (-9)	+2/-3 (-7)	+2/-2 (-6)	
1 to 2 (1 in 3)	.33	+4/-7 (-11)	+4/-6 (-9)	+4/-6 (-8)	+3/-7 (-11)	+3/-6 (-9)	+3/-6 (-8)	+2/-7 (-11)	+2/-6 (-9)	+2/-6 (-8)	
1 to 3 (1 in 4)	.25	+5/-9 (-11)	+5/-8 (-9)	+5/-8 (-8)	+3/-9 (-11)	+3/-8 (-9)	+3/-8 (-8)	+3/-9 (-11)	+3/-8 (-9)	+3/-8 (-8)	
1 to 4 (1 in 5)	.20	+5/-10 (-12)	+5/-10 (-10)	+5/-9 (-9)	+4/-10 (-12)	+4/-10 (-10)	+4/-9 (-9)	+3/-10 (-12)	+3/-10 (-10)	+3/-9 (-9)	
1 to 5 (1 in 6)	.17	+6/-11 (-12)	+6/-10 (-10)	+6/-10 (-10)	+4/-11 (-12)	+4/-10 (-10)	+4/-10 (-10)	+3/-11 (-12)	+3/-10 (-10)	+3/-10 (-10)	
1 to 6 (1 in 7)	.14	+7/-12 (-13)	+7/-11 (-11)	+7/-11 (-10)	+4/-12 (-13)	+4/-11 (-11)	+4/-11 (-10)	+4/-12 (-13)	+4/-11 (-11)	+4/-11 (-10)	
1 to 7 (1 in 8)	.13	+7/-12 (-13)	+7/-12 (-11)	+7/-11 (-10)	+5/-12 (-13)	+5/-12 (-11)	+5/-11 (-10)	+4/-12 (-13)	+4/-12 (-11)	+4/-11 (-10)	
1 to 8 (1 in 9)	.11	+8/-13 (-13)	+7/-13 (-11)	+8/-12 (-10)	+5/-13 (-13)	+5/-13 (-11)	+5/-12 (-10)	+4/-13 (-13)	+4/-13 (-11)	+4/-12 (-10)	
1 to 9 (1 in 10)	.10	+9/-14 (-13)	+9/-13 (-12)	+9/-13 (-11)	+6/-14 (-13)	+6/-13 (-12)	+6/-13 (-11)	+4/-14 (-13)	+4/-13 (-12)	+4/-13 (-11)	

Źródło: R. Nelson, *Multinomial Cutscores...*, op. cit., s. 69.

⁵⁹ *Idem*, *Multinomial Cutscores for Bayesian Analysis with ESS and Three-Position Scores of Comparison Question Polygraph Tests*, „Polygraph & Forensic Credibility Assessment”, 49(1), 2020, s. 62.

⁶⁰ Szczegółową instrukcję korzystania z wielomianowych tablic referencyjnych w systemie ESS-M zob. R. Nelson, M. Handler, T. Coffey, R. Prado, B. Blalock, *How To. A Step-by-Step Worksheet for the Multinomial ESS*, www.researchgate.net/publication/332240538_poligrafia_polygraph_How_To_A_Step-by-Step_Worksheet_for_the_Multinomial_ESS [dostęp: 9.03.2021].

Tab. 12. Wielomianowe progi decyzyjne w zaktualizowanym systemie ESS dla testów wieloprogowych

Prior odds of deception	prior probability	Alpha (truth/deception)									
		.01/.01	.01/.05	.01/.10	.05/.01	.05/.05	.05/.10	.10/.01	.10/.05	.10/.10	
9 to 1 (9 in 10)	.90	(+8) / none	(+8) / none	(+8) / -9	(+7) / none	(+7) / none	(+7) / -9	(+7) / none	(+7) / none	(+7) / -9	
8 to 1 (8 in 9)	.89	(+7) / none	(+7) / none	(+7) / -6	(+7) / none	(+7) / none	(+7) / -6	(+7) / none	(+7) / none	(+7) / -6	
7 to 1 (7 in 8)	.88	(+7) / none	(+7) / -10	(+7) / -6	(+7) / none	(+7) / -10	(+7) / -6	(+7) / none	(+7) / -10	(+7) / -6	
6 to 1 (6 in 7)	.86	(+7) / none	(+7) / -8	(+7) / -5	(+8) / none	(+6) / -8	(+6) / -4	(+6) / none	(+6) / -8	(+6) / -5	
5 to 1 (5 in 6)	.83	(+6) / -14	(+6) / -5	(+6) / -4	(+6) / -14	(+6) / -5	(+6) / -4	(+6) / -14	(+6) / -5	(+6) / -4	
4 to 1 (4 in 5)	.80	(+6) / -9	(+6) / -5	(+6) / -4	(+5) / -9	(+5) / -5	(+5) / -4	(+5) / -9	(+5) / -5	(+5) / -4	
3 to 1 (3 in 4)	.75	(+5) / -6	(+5) / -4	(+5) / -3	(+5) / -6	(+5) / -4	(+5) / -3	(+4) / -6	(+4) / -4	(+4) / -3	
2 to 1 (2 in 3)	.67	(+4) / -5	(+4) / -3	(+4) / -3	(+3) / -5	(+3) / -3	(+3) / -3	(+3) / -5	(+3) / -3	(+3) / -3	
1 to 1 (1 in 2)	.50	(+2) / -4	(+2) / -3	(+2) / -2	(+1) / -4	(+1) / -3	+1 / -2	(+1) / -4	(+1) / -3	(+1) / -2	
1 to 2 (1 in 3)	.33	(+1) / -6	(+1) / -5	(+1) / -4	(0) / -6	(0) / -5	(0) / -4	(0) / -6	(0) / -5	(0) / -4	
1 to 3 (1 in 4)	.25	(0) / -7	(0) / -6	(0) / -6	(0) / -7	(0) / -6	(0) / -6	(0) / -7	(0) / -6	(0) / -6	
1 to 4 (1 in 5)	.20	(+1) / -7	(0) / -7	(0) / -6	(0) / -7	(0) / -7	(0) / -6	(0) / -7	(0) / -7	(0) / -6	
1 to 5 (1 in 6)	.17	(+4) / -8	(+4) / -7	(+4) / -7	(0) / -8	(0) / -7	(0) / -7	(0) / -8	(0) / -7	(0) / -7	
1 to 6 (1 in 7)	.14	(none) / -8	(none) / -8	(none) / -7	(0) / -8	(0) / -8	(0) / -7	(0) / -8	(0) / -8	(0) / -7	
1 to 7 (1 in 8)	.13	(none) / -9	(none) / -8	(none) / -8	(0) / -9	(0) / -8	(0) / -8	(0) / -9	(0) / -8	(0) / -8	
1 to 8 (1 in 9)	.11	(none) / -9	(none) / -8	(none) / -8	(none) / -9	(none) / -8	(none) / -8	(0) / -9	(0) / -8	(0) / -8	
1 to 9 (1 in 10)	.10	(none) / -9	(none) / -9	(none) / -8	(none) / -9	(none) / -9	(none) / -8	(0) / -9	(0) / -9	(0) / -8	

Źródło: R. Nelson, *Multinomial Cutscores...*, op. cit., s. 70.

Zaletą wielomianowych tablic referencyjnych jest ponadto uwzględnienie danych z kanału reakcji naczynioruchowych (PLE/PPG). Najnowsze tablice pokazują różne warianty progów decyzyjnych w zależności od przyjętej tolerancji błędu (optymalnie $\alpha = 0,05$) oraz szans wstępnych (standardowo stosuje się 1:1 – co oznacza, że przed badaniem poligraficznym szanse, że mamy do czynienia z osobą szczerą i nieszczerą, są równe) – zob. tab. 11 i 12. Nastąpiła także korekta optymalnych progów decyzyjnych, dających lepsze efekty (niż np. federalna 3-pozycyjna skala ocen) przede wszystkim poprzez zmniejszenie odsetka wyników nierozstrzygniętych⁶¹.

⁶¹ Por. R. Nelson, *Accuracy Effects for ESS and Three-Position Scores of Federal ZCT Exams Using the Grand Total Rule with Traditional/Federal and Multinomial Cutscores*, „Polygraph & Forensic Credibility Assessment”, 49(2), 2020, s. 157–175.

Przykład konkluzji, wynikającej z zastosowania systemu ESS-M

Załóżmy, że ewaluacji poddajemy dane z testu jednoprotobemowego typu ZCT z trzema pytaniami krytycznymi. Analiza, zgodna z regulacjami ESS-M, przyniosła ocenę całkowitą zmian reakcji przy tych pytaniach: +10 (przy ocenach cząstkowych – R1: +2; R2: +1; R3: +7). Opis rezultatu testu może przedstawiać się wówczas następująco:

Numeryczna ocena całkowita zmian reakcji fizjologicznych osoby badanej przy pytaniach krytycznych wyniosła (+10), przekraczając w kierunku dodatnim wymagany (w tego typu teście – jednowątkowym, z trzema pytaniami krytycznymi) minimalny punkt odcięcia (+3). Statystyczny współczynnik Bayesa dla tych danych wyniósł 7,5. Szanse¹⁾ *a posteriori* w przypadku prawdziwości oszacowano na: 7,5 do 1 – przy czym prawdopodobieństwo *a posteriori* (częstość) wynosiło tu 0,88. Z kolei dolna granica przedziału ufności Bayesa (1-alfa), zaczyna się od szans: 4,5 do 1 – przewyższając szanse wstępne (1 do 1). Wskazuje to na 90%, że szanse *a posteriori* prawdziwości przewyższają szanse wstępne (*a priori*).

Uzyskane wyniki analityczne świadczą zatem o *reakcjach obserwowanych zwykle u osób szczerze odpowiadających na relewantne (krytyczne) pytania testowe*. Ustalenie dowodu z powyższymi rezultatami jest 7,5 razy bardziej prawdopodobne, jeśli badany byłby prawdziwym przy udzielaniu odpowiedzi na pytania krytyczne niż wtedy, gdy odpowiadałby nieszczerze²⁾.

¹⁾ Szansa oznacza stosunek prawdopodobieństwa sukcesu (wystąpienia jakiegoś zjawiska) do prawdopodobieństwa porażki (tego, że prognozowane zjawisko nie wystąpi). Natomiast klasyczne prawdopodobieństwo oznacza stosunek liczby sukcesów do liczby wszystkich prób.

²⁾ To samo można również opisać nieco inaczej (podobnie jak wyraża się wyniki niektórych innych badań kryminalistycznych, np. biologicznych): „Sposób reagowania osoby badanej jest zgodny ze statystycznym profilem osoby odpowiadającej szczerze (referencyjnym zbiorem wyników testów, które zwykle występują w populacji osób właśnie tak odpowiadających na pytania krytyczne). Pewność identyfikacji, wyrażona współczynnikiem Bayesa wynosi 7,5:1 – co oznacza, że hipoteza wskazująca na prawdziwość jest 7,5 razy bardziej prawdopodobna niż hipoteza przeciwna (zakładająca, że taki sposób reagowania jest dziełem przypadku i wystąpił u osoby wprowadzającej w błąd)”. Szersze wyjaśnienie poszczególnych elementów opisu wyniku testu ustalonego po zastosowaniu systemu ESS-M zob. R. Nelson, *Explainer for the ESS-M Narrative Summary*, „APA Magazine”, 52(6), 2019, s. 30–48.

Tab. 13. Zasady analizy numerycznej testów CQT – zgodnie z systemami ESS i ESS-M

	Systemy oceniania ESS i ESS-M
Ogólne wskazówki	<ul style="list-style-type: none"> • zasada „większe jest lepsze”, czyli przypisujemy wartości punktowe przy każdej dostrzegalnej różnicy w reakcjach, • wartość 0 przy braku dostrzegalnej różnicy lub artefaktach, • tylko skala 3-poz. (+1, 0, -1), • wyjątek dla EDA: skala 3-poz., ale z wartościami: +2, 0, -2 (parametr EDA podwójnie ważony).
Pneumo	<ul style="list-style-type: none"> • bezdech brany pod uwagę bez zastrzeżeń tylko przy pytaniach relewantnych, zaś przy pytaniach porównawczych – z ostrożnością (łatwy do sztucznego wytworzenia), • reakcja vs. reakcja = 0 lub priorytet dla długości linii oddechu (RLL) przy dostrzegalnej różnicy, • w testach z pytaniami porównawczymi typu DLC dopuszcza się przypisanie za każdym razem wartości 0 (ale nie jest to obligatoryjne, można zdecydować się na pominięcie wyłącznie reakcji wątpliwych, możliwych do sztucznego wytworzenia).
EDA	• +2, 0, -2
Kardio	• +1, 0, -1
PPG	• +1, 0, -1
Progi decyzyjne* (tolerancja błędu – $\alpha = 0,05$)	<p>• w testach jednoproblemowych i wieloaspektowych typu ZCT z trzema pytaniami relewantnymi:</p> <p style="text-align: center;">ESS-M</p> <p>DI (wskazanie na wprowadzanie w błąd) – gdy suma całkowita ≤ -3 (-4 w pierwszej wersji ESS) lub gdy którakolwiek suma cząstkowa ≤ -7,</p> <p>NDI (brak wskazania na wprowadzanie w błąd) – gdy suma globalna $\geq +3$ (+2 w ESS, gdy $\alpha = 0,1$). Możliwy wyjątek w ESS: jeśli różnica między sumami cząstkowymi o różnych znakach wyniesie więcej niż 7 punktów (np. R1: -5, R2: +9) – to INC (nierozstrzygnięte),</p> <p>INC – pozostałe wyniki.</p> <p>• w teście jednoproblemowym z dwoma pytaniami relewantnymi (You-Phase):</p> <p style="text-align: center;">ESS-M</p> <p>DI (wskazanie na wprowadzanie w błąd) – gdy suma całkowita ≤ -3 (-4 w ESS) lub gdy którakolwiek suma cząstkowa ≤ -5 (-6 w ESS),</p> <p>NDI (brak wskazania na wprowadzanie w błąd) – gdy suma globalna $\geq +3$ (+2 w ESS, gdy $\alpha = 0,1$). Możliwy wyjątek w ESS: jeśli różnica międzysumami cząstkowymi o różnych znakach wyniesie więcej niż 7 punktów (np. R1: -5, R2: +9) – to INC (nierozstrzygnięte),</p> <p>INC – pozostałe wyniki.</p>

Systemy oceniania ESS i ESS-M	
	<ul style="list-style-type: none"> • w teście jednoproblemowym z czterema pytaniami relewantnymi (MGQT): <ul style="list-style-type: none"> ESS-M DI (wskazanie na wprowadzanie w błąd) – gdy suma całkowita ≤ -3 lub gdy którakolwiek suma cząstkowa ≤ -9, NDI (brak wskazania na wprowadzanie w błąd) – gdy suma globalna $\geq +3$, INC – pozostałe wyniki. • dla testów wieloproblemowych typu MGQT, DLST: <ul style="list-style-type: none"> ESS-M i ESS SR (znaczące reakcje przy pytaniach relewantnych) – gdy którakolwiek suma cząstkowa ≤ -3, NSR (brak znaczących reakcji) – gdy wszystkie sumy cząstkowe $\geq +1$, INC (nierozstrzygnięte) – pozostałe wyniki.

* Na temat danych referencyjnych oraz optymalnych progów decyzyjnych w ESS-M zob. R. Nelson, *Multinomial Reference Distributions...*, *op. cit.*, s. 115; *idem*, *Multinomial Cutscores...*, *op. cit.*, s. 61–72.

Źródło: opracowanie własne autora.

Korzystając z możliwości statystyki, biegły znacząco wzmacnia sposób prezentacji wyników przeprowadzonych testów i zarazem wartość dowodową sporządzonej opinii. Aby jednak nie ingerować w kompetencje sądu, musi zachować należyłą ostrożność w formułowaniu konkluzji. Przykładowo Angielski Sąd Apelacyjny – w sprawie *R vs. Doheny* – zaznaczył, że biegły dostarczający dowodu, który ma policzalny iloraz wiarygodności, powinien ograniczyć się do objaśnienia lub wyliczenia tego ilorazu⁶². B. Robertson, G. Vignaux i C. Berger sugerują, że najlepiej wypowiedzieć się wówczas tak:

Dowód jest R razy bardziej prawdopodobny, jeśli to oskarżony pozostawił ślad, aniżeli ktokolwiek inny. Dowód wspiera zatem tezę, że to oskarżony pozostawił ten ślad. Alternatywnie: Jakikolwiek szanse przypisuje się, że oskarżony był obecny [na miejscu zdarzenia – przyp. M.G.] na podstawie innych dowodów – mój dowód zwielokrotnia te szanse przez R⁶³.

⁶² [1996] TLR 504.

⁶³ B. Robertson *et al.*, *Interpreting Evidence...*, *op. cit.*, s. 65.

Zastosowanie teorematu Bayesa powinno ponadto, bez większych kontrowersji, pozwolić na szeroką akceptację poligraficznych technik pytań porównawczych w polskim procesie karnym (co w praktyce ma już miejsce od dłuższego czasu). Rozważania probabilistyczne w kategoriach szans i prawdopodobieństwa w kwestii prawdomówności osoby badanej; niewypowiadanie się wprost na temat faktu głównego oraz tego, czy oświadczenia osoby badanej zasługują na wiarę – zapewniają zachowanie przez biegłego bezpiecznego dystansu od granicy, za którą pozostaje wyłączna domena sądu. Zatrzymanie się na twierdzeniach hipotetycznych eliminuje też niebezpieczeństwa związane z przekształcaniem takich twierdzeń na kategoriyczne poprzez *subiektywną interpretację prawdopodobieństwa*⁶⁴. Biegli często spotykają się jednak z wyraźną presją na formułowanie kategoriycznych konkluzji. Jeśli decydują się na kategoriyczny opis wyników przeprowadzonych testów, powinni to czynić racjonalnie i zgodnie z aksjomatami rachunku prawdopodobieństwa⁶⁵.

Nie ma jednocześnie dostatecznego uzasadnienia dla postulatu 100-procentowej, absolutnej dokładności badań poligraficznych. Dyskryminuje to badania poligraficzne względem innych niekwestionowanych badań kryminalistycznych, które takiego wymogu również nie spełniają. „Jaki sens miałyby identyfikacja grupowa, albo dowód poszlakowy?” – pyta retorycznie R. Jaworski⁶⁶. Oczekiwanie matematycznej pewności prowadziłyby do tego, że żaden przestępca nie zostałby nigdy ukarany. Rzecz w tym, by ustalić akceptowalny poziom prawdopodobieństwa, który zminimalizuje liczbę fałszywych identyfikacji, obciążających osoby niewinne (błąd typu pierwszego), a przy tym nie zwiększy istotnie ryzyka uniknięcia odpowiedzialności przez winnego (błąd typu drugiego)⁶⁷. Trzeba też pamiętać o potencjalnym odciążającym walorze ekspertyzy poligraficznej, a tego typu dowód nie musi odpowiadać tak wysokim wymaganiom jak obciążający⁶⁸.

⁶⁴ T. Widła, *Ocena dowodu z opinii biegłego*, Katowice 1992, s. 69–70.

⁶⁵ Por. D. Kazimierska, *Wartość naukowa metod wykorzystywanych w identyfikacji*, „Prokuratura i Prawo”, 11, 2016, s. 126.

⁶⁶ R. Jaworski, *Opinia z ekspertyzy poligraficznej...*, *op. cit.*, s. 70.

⁶⁷ Por. R. Eggleston, *Sixth Wilfred Fullagar Memorial Lecture. Beyond Reasonable Doubt*, „Monash University Law Review”, 4(1), 1977, s. 1–2.

⁶⁸ Por. R. Jaworski, *Opinia z ekspertyzy poligraficznej...*, *op. cit.*, s. 113; J. Nelken, *Dowód poszlakowy w procesie karnym*, Warszawa 1970, s. 222 i nast.

2.2.6. Znaczenie konsekwencji, spójności w sposobie reagowania

Uzupełnieniem tradycyjnych reguł numerycznej analizy danych testowych jest uwzględnienie tendencji w sposobie reagowania osoby badanej. Wagę tego zjawiska podkreśla izraelska szkoła poligraferów, powołując się m.in. na to, że zagadnienie spójności łączy się z rzetelnością pomiarów, powtarzalnością wyników testów w podobnych warunkach. A. Ginton zauważa, że numeryczny sposób analizy zapisów na poligramach w testach pytań porównawczych łączy w sobie właściwie dwa aspekty. Pierwszy polega na dostrzeżeniu i oszacowaniu ewentualnych różnic w reakcjach na pytania relewantne i kontrolne w określonej fazie testu (strefie na poligramie). Drugi zaś – poprzez kolejne porównania reakcji na te pytania w poszczególnych seriach testu, w różnych miejscach i następnie zsumowanie ocen numerycznych – pozwala na stwierdzenie, czy sposób reagowania badanego jest konsekwentny i wskazuje to na przypisanie przez niego określonej kategorii pytań relatywnie większego znaczenia. Teoretycznie zatem problem konsekwencji w reakcjach zawiera się w już numerycznej metodzie analizy danych testowych, ale A. Ginton postrzega spójność w reakcjach za ważniejszy czynnik niż wielkość różnic w reakcjach przy pojedynczych zestawieniach pytań⁶⁹. Obrazuje to następującym przykładem. Załóżmy, że analizie zostały poddane dwa różne testy, składające się z trzech wykresów (serii tych samych pytań), zakończone wynikami zamieszczonymi w tab. 14:

Tab. 14. Przykładowe wyniki dwóch testów CQT złożonych z trzech wykresów

Wykres	Test A	Test B
I	-8	-2
II	+1	-1
III	+1	-1
suma ocen	-6	-4

Źródło: A. Ginton, *The Importance of the Consistency Factor...*, op. cit., s. 148.

Suma ocen numerycznych w teście A jest niższa, ale to test B – zdaniem A. Gintona – bardziej wskazuje na nieszczerłość, zważywszy na tendencję w sposobie reagowania⁷⁰. Formalnie oba testy wskazują

⁶⁹ A. Ginton, *The Importance of the Consistency Factor...*, op. cit., s. 146–162.

⁷⁰ *Ibidem*, s. 148.

na nieszczerość (a w niektórych systemach analizy wynik testu B byłby uznany za nierozstrzygnięty). Przy powyższym przykładzie pozostaje pytanie o przyczynę wysokiego wyniku ujemnego dla pierwszego wykresu testu A i zmianę tendencji w reakcjach (w przeciwnym kierunku – dodatnim) w pozostałych dwóch seriach. Na przestrzeni całego testu można by powziąć wątpliwość, czy w czasie pierwszej serii nie doszło do jakichś przypadkowych fluktuacji lub czynników zakłócających rejestrowane przez poligraf reakcje. Być może po pierwszym wykresie badany przekazał jakieś informacje dotyczące pytań krytycznych, które spowodowały, że w dalszej części te pytania miały dla niego już inne znaczenie? W tym kontekście rozważania A. Gintona niewątpliwie zasługują na uwagę.

Ustalenie minimalnych proporcji zmian porównywanych reakcji oraz liczbowych progów decyzyjnych ma z założenia ograniczać wpływ na wynik testu przypadkowej aktywności psychofizjologicznej oraz artefaktów. Izraelski ekspert uważa to za niewystarczające i poszedł dalej – opracował metodę analizy konsekwencji w reakcjach między wykresami (seriami testu).

Ujemna suma ocen numerycznych dla pytań krytycznych w jednej serii testu oznacza trend w kierunku wskazania na wprowadzanie w błąd, wyrażony w skrócie literą *D* (od *deceptive*). Suma dodatnia sygnalizuje trend w kierunku wskazania na szczerłość, który zostanie oznaczony literą *T* (od *truthful*). Z kolei wynik zerowy (bez jakiegokolwiek tendencji) będzie wyrażony literą *Z* (*zero*). W ten sposób możliwe są przedstawione niżej kombinacje:

Tab. 15. Możliwe kombinacje wyników testu złożonego z trzech serii (wykresów)

Oznaczenie	Kombinacja ocen numerycznych dla poszczególnych serii testu
3D	suma ocen w każdym z trzech wykresów jest ujemna
3T	suma ocen w każdym z trzech wykresów jest dodatnia
2Dz	dwa wykresy mają łączne oceny ujemne, a jeden zerową
2Tz	dwa wykresy mają łączne oceny dodatnie, a jeden zerową
2Dt	dwa wykresy mają łączne oceny ujemne, a jeden dodatnią
2Td	dwa wykresy mają łączne oceny dodatnie, a jeden ujemną
1dZ	jeden wykres ma łączną ocenę ujemną, a pozostałe – zerową
1tZ	jeden wykres ma łączną ocenę dodatnią, a pozostałe – zerową
1dtz	jeden wykres ma łączną ocenę ujemną, jeden dodatnią, a inny – zerową
3Z	suma ocen w każdym z trzech wykresów wynosi zero

Źródło: A. Ginton, *The Importance of the Consistency...*, op. cit., s. 149.

W metodzie Gintona przyjmuje się, że wynik świadczący o nie-szczerości występuje, gdy przynajmniej dwa wykresy pokazują trend z wynikami ujemnymi. Odwrotny wynik raportujemy, gdy co najmniej dwa wykresy dadzą dodatnie oceny numeryczne. W pozostałych przypadkach test uznaje się za nierozstrzygnięty. Oprócz tego metoda przewiduje dwa kryteria dodatkowe. Pierwsze z nich wprowadza się wówczas, gdy dwa z trzech wykresów przyniosą oceny zerowe, a tylko jeden wykazuje trend w określonym kierunku. Jeśli suma ocen z trzech wykresów przekracza próg ± 4 – w spójności z tym, na co wskazuje ten jeden z „rozstrzygniętych” wykresów (z oceną różną od zera) – możliwa jest konkluzja NDI/DI. W innym wypadku rozstrzygnięcia całego testu nie ma. Nie będzie go także wtedy, gdy suma ocen numerycznych dla wszystkich wykresów będzie stała w sprzeczności z trendem występującym w dwóch wykresach. Na przykład, gdy uzyskamy ujemne oceny w dwóch seriach, a suma wszystkich ocen okaże się dodatnia (wykres 1: -2 , wykres 2: $+4$, wykres 3: -1 ; łącznie: $+1$).

Skuteczność swojej metody A. Ginton potwierdził w badaniach empirycznych, jednakże – jak dotąd – nie doczekały się one replikacji, potwierdzonych publikacjami. Gdy dziesięciu doświadczonych poligraferów oceniło łącznie 1000 testów zgodnie z powyższymi zasadami, osiągnięto dokładność diagnoz na poziomie 87,81% (trafnie wytypowano 83,4% spośród nieszczerych badanych oraz 91,88% spośród prawdomównych), przy odsetku wyników nierozstrzygniętych wynoszącym 9,8%. Metoda Gintona może być potraktowana jako alternatywa dla powszechnie uznanych, potwierdzonych naukowo systemów analizy danych testowych, albo – jak wskazuje sam autor – ich uzupełnieniem. Jednym ze sposobów połączenia reguł Gintona z regułami innych metod (np. tzw. systemu federalnego) jest przyjęcie, że wynik nierozstrzygnięty testu będzie występował wyłącznie, gdy obie metody na to wskażą albo dostarczą wzajemnie sprzecznych wskazań (przy czym owa sprzeczność należy do rzadkości – w próbie badawczej Gintona taka sytuacja zaistniała jedynie w 1,75% przypadków). Kombinacja metod nie zmieniła istotnie dokładności diagnoz, lecz zmniejszyła procent wyników nierozstrzygniętych do 5,5, a więc o $1/4$ ⁷¹. Jednym z ograniczeń metody Gintona jest opracowanie reguł

⁷¹ *Ibidem*, s. 150–151.

decyzyjnych mających zastosowanie przy skompletowaniu jedynie trzech podstawowych wykresów. Nie zostały przedstawione zasady postępowania w sytuacji, gdy test składa się z czterech lub pięciu wykresów (co praktykuje się w innych systemach analizy przy braku rozstrzygnięcia po trzech seriach testu albo gdy uzyskano za mało zapisów nadających się do miarodajnej oceny).

3. Ewaluacje badających vs. ślepa interpretacja zapisów na poligramach

Pod względem zaznajomienia z przedmiotem i podmiotem badań można wyróżnić trzy sytuacje, w których poligraferzy dokonują analizy zapisów na poligramach. Pierwsza i podstawowa ma miejsce wtedy, kiedy biegły przeprowadza badanie osobiście, a potem analizuje dane testowe. Zna oczywiście szczegóły sprawy – tak z informacji przekazanych przez zleceniodawcę, udostępnionych materiałów, jak i interakcji z osobą badaną. Pozostałe dwie sytuacje mają miejsce wtedy, gdy do interpretacji zapisów angażuje się innych biegłych – przy czym albo udostępnia się te same informacje co badającemu, włącznie z tym, jak przebiegało samo badanie, albo ewaluatorzy podejmują zadanie zupełnie niezależnie – bez jakichkolwiek innych danych poza zapisami reakcji fizjologicznych. W tym ostatnim przypadku mamy do czynienia z tzw. ślepą interpretacją. Taki sposób postępowania eliminuje wszelkie subiektywne czynniki związane z okolicznościami sprawy, której dotyczy realizowana ekspertyza. Subiektywizm ogranicza się wyłącznie do osobistych preferencji poligrafera odnośnie do sposobu interpretacji danych tam, gdzie występuje pewien margines swobody.

Jak zauważa M. Orne – w rzeczywistych warunkach wynik badania poligraficznego może pozostawać pod wpływem innych dowodów zgromadzonych w sprawie oraz własnych przeświadczeń badającego o winie lub niewinności osoby badanej. Dla w pełni naukowo trafnej metody konieczne jest oddzielenie rezultatów testów od innych aspektów danego postępowania, ale w praktyce nie jest to możliwe do spełnienia⁷². M. Orne przytoczył m.in. dane z eksperymentu P. Bersha,

⁷² M.T. Orne, *Implications of Laboratory Research for the Detection of Deception*, „Polygraph”, 2(3), 1973, s. 177.

który porównywał diagnozy poligraferów z opiniami panelu trzech ekspertów ze środowiska prawniczego (pracującego na materiale dowodowym przedłożonym w sądzie, jak również na dodatkowych informacjach, które nie mogły być przekształcone w formalnie dopuszczalne dowody). Jeżeli wszyscy trzej eksperci byli jednomyślni, oceny poligrafera były z tym zgodne w 92,4% przypadków. Gdy jeden z prawników miał zdanie odrębne – wówczas zgodność między ocenami poligraferów i panelu obniżyła się do 74,6%⁷³. M. Orne przedstawił dwie hipotezy, które mogłyby tłumaczyć tę zmianę. Pierwsza jest taka, że w dalszym ciągu to wyniki badań poligraficznych były bardziej trafne niż oceny prawników spierających się ze sobą wobec niejednoznacznego materiału dowodowego. Natomiast druga wersja przedstawia możliwość, że przy pełnej zgodności panelu ekspertów – informacje, na których bazowali prawnicy, równie dobrze mogły istotnie wpływać na postępowanie poligrafera wobec osoby badanej; sposób, w jaki badany reagował, oraz finalny osąd badającego⁷⁴.

Różne badania nad dokładnością testów pytań porównawczych w rzeczywistych sprawach (poza laboratorium eksperymentalnym) sugerują, że eksperci osobiście przeprowadzający badania mogą uzyskiwać nawet lepsze rezultaty niż ci, którzy dokonują ślepej interpretacji poligramów. O ile obie grupy niemal jednakowo trafnie identyfikują osoby *winne* (nieuczestniczące), to badający – w świetle przedstawionych opracowań (zob. tab. 16 i tab. 17) – okazywali się dokładniejsi w identyfikacji osób *niewinnych* (prawdomównych). Autor niniejszej monografii przyjął nieco inną hipotezę – zakładając, że mniejsza liczba subiektywnie oddziałujących czynników u ślepo oceniających powinna jednak prowadzić do dokładniejszych wskazań właśnie w tej grupie (o weryfikacji tej hipotezy w części II, podrozdziale 5.2). Gdyby jednak hipoteza nie znalazła empirycznego potwierdzenia, oznaczałoby to, że dodatkowe subiektywne okoliczności, wynikające przede wszystkim z bezpośredniej interakcji z osobą badaną, z reguły pozytywnie wpływają na poprawność opiniowania poligraferów.

⁷³ P. Bersh, *A Validation Study of Polygraph Examiner Judgments*, „Journal of Applied Psychology”, 53, 1969, s. 399–403.

⁷⁴ M.T. Orne, *Implications of Laboratory Research...*, *op. cit.*, s. 178–179.

Tab. 16. Wyniki badań terenowych na temat dokładności testów CQT ocenianych przez niezależne osoby

Opracowanie*	Winni (%)				Niewinni (%)			
	n	Prawidłowo	Błędnie	Brak rozstrzygnięcia	n	Prawidłowo	Błędnie	Brak rozstrzygnięcia
Honts (1996)	7	100	0	0	6	83	0	17
Honts i Raskin (1988)	12	92	0	8	13	62	15	23
Patrick i Iacono (1991)	52	92	2	6	37	30	24	46
Raskin <i>et al.</i> (1988)	37	73	0	27	26	61	8	31
Średnia	27	89	1	10	82	59	12	29
Procent opinii		98	2	–		83	17	–

* C.R. Honts, *Criterion Development and Validity of the Control Question Test in Field Application*, „Journal of General Psychology”, 123, 1996, s. 309–324; C.R. Honts, D.C. Raskin, *A Field Study of the Directed Lie Control Question*, „Journal of Police Science and Administration”, 16(1), 1988, s. 56–61; C.J. Patrick, W.G. Iacono, *Validity of the Control Question Polygraph Test. The Problem of Sampling Bias*, „Journal of Applied Psychology”, 76(2), 1991, s. 229–238; D.C. Raskin, J.C. Kircher, C.R. Honts, S.W. Horowitz, *A Study of The Validity of Polygraph Examinations in Criminal Investigations. Final Report to the National Institute of Justice*, Salt Lake City 1988.

Źródło: opracowanie własne autora.

Tab. 17. Wyniki badań terenowych na temat dokładności testów CQT ocenianych przez osoby przeprowadzające testy

Opracowanie*	Winni (% prawidłowych identyfikacji)	Niewinni (% prawidłowych identyfikacji)
Horvath (1977)	100	100
Honts i Raskin (1988)	92	100
Raskin <i>et al.</i> (1988)	95	96
Patrick i Iacono (1991)	100	90
Honts (1996)	94	100
Średnia	98	97

* Poza wymienionymi w poprzednim przypisie zob. F. Horvath, *The Effects of Selected Variables on Interpretation of Polygraph Records*, „Journal of Applied Psychology”, 62(2), 1977, s. 127–136.

Źródło: opracowanie własne autora.

Trudno jednak przesądzić, czy zapoznanie analizującego dane testowe ze szczegółami sprawy i przebiegiem badania poligraficznego wpływa korzystnie czy nie na dokładność opiniowania. Wykazano wcześniej (część I, rozdział 1), że informacje stawiające osobę badaną w lepszym lub gorszym świetle mogą wpływać na oceny numeryczne przypisywane przez badających, gdy zapisy reakcji nie są całkowicie jednoznaczne. W eksperymencie z lat 50., przeprowadzonym przez W. Holmesa, udostępnienie poligraferom informacji o sprawach, z których pochodziły oceniane poligramy, spowodowało niewielki wzrost dokładności diagnoz o 8%⁷⁵. Sprawdzali to eksperymentalnie także D. Wicklander i F. Hunter, którzy w co najmniej dwumiesięcznym odstępie czasu zorganizowali dwukrotne analizy 20 zestawów poligramów (technika Reida) przez 6 niezależnych od siebie poligraferów. Za pierwszym razem przekazali tylko lakonicznie problem, którego dany test dotyczył (np. kradzież pieniędzy ze skrytki w biurze). Przed drugą interpretacją poszerzyli zakres informacji o zarys historii zdarzenia, podstawowe dane biograficzne osoby badanej, jej werbalne i niewerbalne zachowania i wykaz relewantnych pytań testowych. Średnia dokładność diagnoz z pierwszej analizy danych wyniosła 88,33% [0,70–0,95], natomiast gdy poligraferzy posiadali dodatkowe informacje, okazali się oni nieco dokładniejsi – diagnozowali z dokładnością 92,5% [0,80–1,00]. Czterech z sześciu oceniających poprawiło swoje wyniki, jeden pozostał na tym samym poziomie, a jeden wypadł gorzej niż przy ślepej interpretacji. O połowę zmniejszyła się też liczba wyników nierozstrzygniętych⁷⁶. Ten eksperyment jest kolejnym dowodem na to, że poligraferzy w ramach swojej pracy mają przestrzeń na swobodną interpretację, informacje o kontekście badanej sprawy wpływają w pewnym stopniu na ich decyzje, natomiast subiektywizm nie za każdym razem ujemnie wpływa na dokładność diagnoz.

Wszystko jednak zależy zapewne od tego, czy informacje kształtujące wstępny pogląd biegłego na temat sprawy będą spójne czy sprzeczne ze stanem faktycznym. Jeśli będą pokrywały się z rzeczywistością – pomogą. Jeśli nie – mogą zaszkodzić. Tego z góry przeciez

⁷⁵ W.D. Holmes, *The Degree of Objectivity in Chart Interpretation*, [w:] V.A. Leonard, *Academy Lectures on Lie-detection*, vol. 2, Springfield 1957, s. 67–70.

⁷⁶ D.E. Wicklander, F.L. Hunter, *The Influence of Auxiliary Sources of Information in Polygraph Diagnosis*, „Journal of Police Science and Administration”, 3(4), 1975, s. 405–409.

nigdy nie wiadomo. Szczególnego znaczenia nabierają wówczas osobiste zdolności poligrafera, włącznie z intuicją.

4. Algorytmy komputerowe

4.1. Algorytmiczna całościowa analiza danych testowych

Jednym ze sposobów ograniczenia subiektywizmu, będącego immanentną cechą analizy danych testowych, dokonywanej przez człowieka, jest zastosowanie algorytmów przeliczających wartości, w których wyrażane są poszczególne reakcje mierzone przez poligraf. Korzystanie z algorytmów ułatwia komputeryzacja, przez co obliczenia mogą być dokonywane automatycznie, w sprawny sposób, z perfekcyjną rzetelnością – ten sam test, przy niezmiennych kryteriach, za każdym razem przyniesie tożsame wyniki.

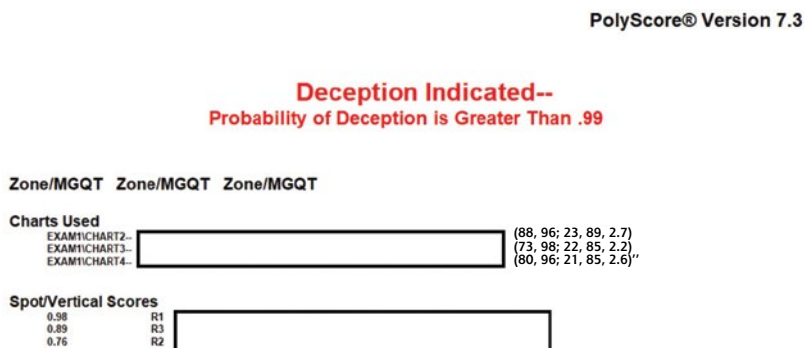
Pierwszy taki algorytm komputerowy, oparty na modelu regresji logistycznej, został opracowany na zlecenie Departamentu Obrony USA przez J. Harrisa i A. McQarrie – uczonych z Laboratorium Fizyki Stosowanej (Applied Physics Laboratory) Uniwersytetu Johna Hopkinsa w Baltimore. Prototypowi algorytmu nadano nazwę PASS – *Polygraph Automated Scoring System*. Dane surowe zarejestrowane przez poligraf program przetwarza na cechy świadczące o sygnałach emocjonalnych. Z każdą kolejną wersją programu przestrzeń na te sygnały rosła (np. od 4888 w wersji 3.0 do ok. 10 tys. w wersji 5.1)⁷⁷. Na podstawie tych danych algorytm wylicza prawdopodobieństwo wprowadzania w błąd. Jeżeli prawdopodobieństwo to osiąga co najmniej 90% – następuje wskazanie na nieszczerłość. Gdy wynosi 10% lub mniej – test nie wskazuje na wprowadzanie w błąd (zob. fragment przykładowego raportu – ryc. 17).

W 1993 r. wersja 2.0 programu została szerzej udostępniona do badań naukowych, natomiast w 1994 r. produkt przekazano praktykom pod nazwą handlową Polyscore 2.3. W kolejnych latach powstawały zaktualizowane wersje programu, które sprawdzano empirycznie. Dokładność PASS 2.0 potwierdzono na poziomie 78,95%; Polyscore 2.3 – 79,21%; Polyscore 2.9 – 83,5%; Polyscore 3.0 – 82,83% (wszystkie

⁷⁷ National Research Council – National Academies Press, *The Polygraph and Lie Detection*, op. cit., s. 309.

dane z wyłączeniem wyników nierozstrzygniętych)⁷⁸. Oprogramowanie współpracuje obecnie z systemami produkowanymi przez firmy Lafayette Instrument Co. oraz Axciton Systems Inc.

Ryc. 17. Fragment raportu z programu Polyscore (v. 7.3)



Źródło: materiały własne autora.

Do badań nad algorytmem początkowo wykorzystywano dane z eksperymentów laboratoryjnych, a następnie włączono również dane z badań prowadzonych w rzeczywistych sprawach – w tym takich, w których stan faktyczny był potwierdzony innymi dowodami, np. przyznaniem się podejrzanego lub innej osoby do winy. Akceptowano również dane z testów, które zostały jednakowo zinterpretowane zarówno przez badającego, jak i ślepo oceniających recenzentów. Oszacowana w ten sposób dokładność identyfikacji (z wyłączeniem wyników nierozstrzygniętych) wyniosła 99,4%. Ten imponujący wynik nie potwierdził się w badaniach laboratoryjnych dotyczących wyreżyszerowanego przestępstwa, które prowadził Departament Obrony USA. Diagnozy były dokładne w 79% (bez wyników nierozstrzygniętych na poziomie 10%)⁷⁹. Warto jednak zaznaczyć, że eksperci oceniający poligramy manualnie poradzili sobie podobnie. Wiadomo też, że eksperymenty laboratoryjne mają swoje ograniczenia (choćby

⁷⁸ N.J. Blackwell, Department of Defense Polygraph Institute, *An evaluation of the effectiveness of the polygraph automated scoring system (PASS) in detecting deception in a mock crime analog study* (Report No. DoDPI94-R0003), 1994, apps.dtic.mil/sti/pdfs/ADA305755.pdf [dostęp: 9.06.2020].

⁷⁹ *Ibidem*.

przez brak realnych konsekwencji związanych z krytycznymi zagadnieniami) i zwykle w rzeczywistych sprawach badania poligraficzne sprawdzają się znacznie lepiej.

Badania weryfikujące dokładność i użyteczność algorytmu były kontynuowane. W 1999 r. opublikowano dane dotyczące Polyscore 3.3, wykorzystanego do analizy testów typu ZCT i MGQT z 200 rzeczywistych spraw, i porównano je z ocenami manualnymi poligraferów (w numerycznej skali 7- i 3-pozycyjnej). Zastosowanie algorytmu w testach ZCT (jednoproblemowych) dało lepsze rezultaty niż bazowanie na ocenach manualnych. Dokładność Polyscore 3.3 wyniosła 96,8% (z wyłączeniem wyników nierozstrzygniętych – 6,1%), a poligraferów: 87% i 88,1% (odpowiednio w skali 7-pozycyjnej i 3-pozycyjnej, przy czym dla tej ostatniej odsetek wyników nierozstrzygniętych przekraczał 20%). Dokładność identyfikacji osób winnych była niemal jednakowa przy każdej z metod. Istotna różnica zarysowała się przy badaniach osób niewinnych. Polyscore trafnie wytypował 93,8% takich osób (bez wyników nierozstrzygniętych), a eksperci najwyżej 60,3% (przy skali 7-pozycyjnej)⁸⁰. W przypadku testów MGQT (wieloproblemowych) oceny manualne okazały się nieco dokładniejsze niż za pomocą algorytmu, ale generalnie liczby były zbliżone (88,5% dokładności w skali 7-pozycyjnej i 86,8% w przypadku algorytmu)⁸¹. Natomiast ponownie algorytm skuteczniej identyfikował osoby „niewinne” (62,5% w porównaniu do 31,9% przy skali 7-pozycyjnej, bez wyników nierozstrzygniętych).

Inny algorytm – CPS (z j. ang. *computerized polygraph system*), znany również jako *Probability Analysis* (analiza prawdopodobieństwa) – został opracowany przez *Scientific Assessment Technologies* na gruncie badań prowadzonych w latach 80. XX wieku przez J. Kirchera i D. Raskina z Uniwersytetu Utah. Przy projektowaniu algorytmu wykorzystano dane ze spraw kryminalnych, udostępnione przez amerykańską Secret Service⁸². CPS uwzględnia zasady przewidziane w manualnym systemie numerycznej analizy danych w wersji Utah.

⁸⁰ N.J. Blackwell, *PolyScore 3.3 and Psychophysiological Detection of Deception Examiner Rates of Accuracy When Scoring Examinations from Actual Criminal Investigations*, „Polygraph”, 28(2), 1999, s. 156.

⁸¹ *Ibidem*, s. 158.

⁸² Zob. J.C. Kircher, D.C. Raskin, *Computer Methods for the Psychophysiological Detection of Deception*, [w:] M. Kleiner (red.), *Handbook...*, *op. cit.*, s. 287–326.

Początkowo było to 12 diagnostycznych cech zapisów, a w późniejszych wersjach nastąpiło uproszczenie do trzech: amplitudy zmian w przewodnictwie elektrycznym skóry, podniesienia linii bazowej w kanale kardio oraz długości linii oddechu. Algorytm opiera się na wielowymiarowej liniowej analizie funkcji dyskryminacyjnej, po której następują szacunki prawdopodobieństwa prawdomówności lub wprowadzania w błąd. Według opublikowanego w 2005 r. raportu z badań prowadzonych na Uniwersytecie Utah – z wyłączeniem 11% wyników nierozstrzygniętych wskazania CPS były prawidłowe w przypadku 90,9% badanych prawdomównych i 83,3% nieszczerych⁸³.

Ciekawą perspektywą dla badaczy jest połączenie oceny manualnej (ludzkiej) z algorytmami wspomaganymi komputerowo. D. Krapohl i W. Goodson sprawdzili, czy zastosowanie algorytmu (PASS – prototypu *polyscore*) wpłynie, a jeśli tak to w jaki sposób – na decyzje poligraferów oceniających testy przesiewowe w formie R/I⁸⁴. Wybór takiego testu do porównania wydaje się o tyle ważny, że test R/I ocenia się metodą jakościową, a swego czasu był on szeroko wykorzystywany w badaniach kadrowych (przedzatrudnieniowych oraz w ramach kontroli personelu) w wielu instytucjach państwowych, w tym oczywiście w USA i również w Polsce. Jednocześnie dokładność tego testu raportowana w różnych badaniach nie była zbyt zachęcająca (w granicach 60%)⁸⁵.

D. Krapohl z W. Goodsonem wykorzystali próbę 100 badań kandydatów ubiegających się na początku lat 2000 o pracę w ochronie lotniska w amerykańskim stanie Georgia. Punktem odniesienia dla ustalenia stanu faktycznego i szczerości odpowiedzi badanych było jedno z pytań testowych – o palenie marihuany w ciągu ostatnich 30 dni (co zostało zweryfikowane poprzez badania moczu). W ten sposób wybrano 50 badań, w których osoby badane wprowadzały w błąd w tej kwestii. Drugą grupę, w równej liczbie, stanowiły testy osób odpowiadających szczerze na wszystkie pytania relewantne (w sumie były cztery pytania R). Jako

⁸³ J.C. Kircher, S.D. Kristjansson, M.K. Gardner, A. Webb, *Human and Computer Decision-making in the Psychophysiological Detection of Deception (Final Report. DoDPI02-R-0001)*, University of Utah, 2005.

⁸⁴ D.J. Krapohl, W. Goodson, *Decision Accuracy for the Relevant-Irrelevant Screening Test. Influence of an Algorithm on Human Decision-Making*, „European Polygraph”, 9, 4(34), 2015, s. 189–208.

⁸⁵ Por. D.J. Krapohl, S. Senter, B. Stern, *An Exploration of Methods...*, *op. cit.*; D.J. Krapohl, T. Rosales, *Decision Accuracy for the Relevant-Irrelevant Screening Test. A Partial Replication*, „Polygraph”, 41(1), 2014, s. 20–29.

oceniających metodą ślepej interpretacji zaangażowano 11 poligraferów Departamentu Bezpieczeństwa Publicznego Teksasu.

W omawianym eksperymencie zgodność ocen manualnych z komputerowymi wyniosła 66,5% (z czego 72,5% w przypadku badanych nieuczciwych i 60,7% – uczciwych). Zaznajomienie poligraferów z obliczeniami algorytmu zwiększyło nieco dokładność diagnoz (o 3%) i zgodność między oceniającymi (o 3,6%), ale różnice nie okazały się statystycznie istotne. W innym wariantcie poligraferzy podejmowali decyzje tylko w przypadku, gdy wynik analizy algorytmicznej był nierozstrzygnięty. Wówczas zwiększył się znacząco procent poprawnych wskazań badanych uczciwych (o 26,4%), a przez to ogólna dokładność (o 16%). Wyższa była też zgodność oceniających (o ponad 30%). Nie było natomiast różnicy w przypadku wskazań nieuczciwości. Takie efekty nie dziwią, jeśli weźmiemy pod uwagę, że rzeczywiście test R/I oceniany manualną metodą jakościową, z jej immanentnym subiektywizmem, daje bardzo słabe wyniki, niekorzystne dla osób prawdomównych (nawet poniżej poziomu statystycznej szansy⁸⁶). Algorytm pokazuje potencjał na częściową redukcję tych fałszywych identyfikacji i zmniejszenie zakresu subiektywnych interpretacji. Ze-stawienie danych z eksperymentu przedstawia tab. 18.

Tab. 18. Wyniki eksperymentu D. Krapohla i W. Goodsona dotyczącego trafności decyzyjnej dla testu przesiewowego R/I, łącznie z wykorzystaniem algorytmu *polyscore* (wartości wyrażono w procentach)

	Globalna ocena manualna	Algorytm <i>polyscore</i>	Globalna ocena manualna + zapoznanie z wynikami algorytmu	I etap: algorytm II etap: ocena manualna
Prawidłowe wskazania (NSR, SR ogółem)	63,6	70,0	66,6	79,5
Fałszywe wskazania	31,5	30,0	21,8	19,5
Odsetek nierozstrzygniętych (NO)	5,0	15,0	11,5	1,0
Zgodność między oceniającymi	60,8	–	64,6	93,0

Źródło: opracowanie własne na podst. D.J. Krapohl, W. Goodson, *Decision Accuracy...*, op. cit., s. 195–203.

⁸⁶ Zob. D.J. Krapohl, T. Rosales, *Decision Accuracy...*, op. cit.

Na komputerową analizę danych nie mają wpływu subiektywne czynniki i słabości natury ludzkiej, jak: uprzedzenia, nastroje, zmęczenie, braki w doświadczeniu czy wykszoleniu. Dzięki temu ten sposób analizy może poprawiać rzetelność testowania. Człowiek natomiast poradzi sobie lepiej w rozpoznawaniu wzorów reakcji zapisywanych na poligramach, artefaktów czy ocenie stabilności zarejestrowanych danych. Tak jak przy prognozowaniu pogody łączy się dane statystyczne z lokalnymi odchyleniami – podobnie w badaniach poligraficznych biegły z jednej strony stosuje wystandaryzowane metody analizy danych, a z drugiej rozważa specyfikę indywidualną – wszelkie czynniki, które mogą mieć wpływ na dane rejestrowane przez poligraf w konkretnym przypadku⁸⁷.

Naukowcy z Uniwersytetu Utah nie stwierdzili istotnej różnicy między dokładnością diagnoz biegłych z zakresu badań poligraficznych i tych otrzymywanych komputerowo⁸⁸. Jednak – jak słusznie zauważa N. Blackwell – nie wszyscy praktykujący poligraferzy faktycznie są *biegłymi*, prezentują zróżnicowany poziom wykszolenia, doświadczenia zawodowego i subiektywizmu⁸⁹. W tym kontekście algorytm komputerowy może nawet przewyższać zdolności przeciętnego poligrafera, ale nawet najlepszy algorytm nie sprawdzi się, jeśli pytania testowe nie będą odpowiednio zadane, a poligramy odpowiednio skompletowane, z klarownymi zapisami. W dalszym ciągu standardy praktyki i procedury wszystkich instytucji, w których przeprowadza się badania poligraficzne, dopuszczają posługiwanie się algorytmami przy analizie zapisów jedynie pomocniczo, a podstawową pozostaje metoda manualna.

Wynalezienie komputerowych programów do analizy danych testowych zainspirowało niektórych badaczy do wysunięcia kolejnej – po metodach numerycznych oraz klasyfikowania reakcji za pomocą rang – propozycji obiektywizacji manualnych metod oceniania zapisów na poligramach. Tym razem system miał być całkowicie obiektywny. D. Krapohl i B. McManus zaproponowali rozwiązanie polegające najpierw na pomiarze zmian reakcji z wykorzystaniem oprogramowania, którego twórcami byli naukowcy z Uniwersytetu Johna Hopkinsa, a następnie dane z tych pomiarów miały być przekształcane na wskaźniki R/C, czyli

⁸⁷ Zob. *ibidem*, s. 190–192.

⁸⁸ J.C. Kircher, D.C. Raskin, *Human versus Computerized Evaluations of Polygraph Data in a Laboratory Setting*, „Journal of Applied Psychology”, 73(2), 1988, s. 291–302.

⁸⁹ Por. s. 193.

wielkość reakcji na pytanie krytyczne była dzielona przez odpowiednią wartość dla pytania kontrolnego⁹⁰. W konsekwencji otrzymuje się liczbę obrazującą relatywną intensywność reakcji na pytanie krytyczne. Zgodnie z regułami DoDPI reakcje na dane pytanie krytyczne porównywano z reakcjami na to z sąsiadujących pytań kontrolnych, przy którym były one bardziej znaczące, a gdy nie można było wykorzystać danych z sąsiednich pytań kontrolnych (np. z powodu artefaktów), brano pod uwagę możliwie najbliższe pytanie kontrolne w sekwencji. Wskaźnik R/C większy od jednego oznacza, że istotniejsze zmiany reakcji wystąpiły przy pytaniu krytycznym. Z kolei wartość mniejsza od jednego świadczy o istotniejszych zmianach reakcji przy pytaniu kontrolnym.

Mierzone miały być tzw. cechy diagnostyczne Kirchera, czyli długość linii oddechu (RLL – z j. ang. *respiration line lenght*), amplituda EDA/EDR oraz amplituda objętości krwi (BV). Cechy te mają jasno określone okna reakcji. RLL mierzona jest przez 10 sekund z chwilą wprowadzenia bodźca (pytania). Reakcja w kanale EDA musi rozpocząć się 0,5 sekundy od wprowadzenia bodźca do 5 sekund po udzieleniu odpowiedzi, a pomiar trwa do 20 sekund po rozpoczęciu reakcji. Objętość krwi natomiast jest średnią z fali tętna, mierzoną od wprowadzenia jednego bodźca do kolejnego. Wszelkie artefakty są oczywiście pomijane w analizie. Te same pomiary są możliwe również w sposób manualny, ale komputer czyni ten proces dokładniejszym, a przede wszystkim szybszym, co zwiększa praktyczną użyteczność metody.

Punktację przeprowadza się w skali 7-pozycyjnej. Dla poszczególnych pozycji, odrębnie w każdym parametrze wyznaczono odpowiednie przedziały wskaźnika R/C (zob. tab. 19). Parametrom przypisano także określone wagi – po 25% dla RLL i BV oraz 50% dla EDA, przez co możliwe oceny w kanale EDA zostały podwojone (+/- 0, 2, 4, 6).

Obiektywna metoda manualnej oceny zapisów (OSS – z j. ang. *objective scoring system*) pozwoliła na trafne wytypowanie spośród badanych 93,2% osób szczerych i 94,6% nieszczerych w próbie 300 potwierdzonych spraw, a w kolejnej próbie 60 potwierdzonych spraw było to odpowiednio: 93,3% i 84,6% (z wyłączeniem wyników nierozstrzygniętych). Zastosowano przy tym progi decyzyjne +/- 6 dla oceny całkowitej, pomijając dodatkowe reguły systemu federalnego odnoszące się do ocen cząstkowych.

⁹⁰ D.J. Krapohl, B. McManus, *An Objective Method for Manually Scoring Polygraph Data*, „Polygraph”, 28(3), 1999, s. 209–222.

Gdy porównano oceny manualne trzech poligraferów, którzy analizowali wykresy ze 100 spraw tradycyjną metodą (numeryczną, w federalnej skali 7-pozycyjnej) i oceny dwóch innych poligraferów, którzy posłużyli się zaproponowaną *metodą obiektywną* (choć bez pomiarów wspomaganym komputerowo) – okazało się, że ci ostatni byli zarówno bardziej zgodni między sobą (0,86 wobec zgodności na poziomie 0,70–0,73 pośród oceniających metodą tradycyjną), jak również bliżsi ustalenia stanu faktycznego (0,82–0,83 wobec 0,63–0,71).

Tab. 19. Zasady punktacji w obiektywnej metodzie manualnej oceny danych poligraficznych (OSS), opracowane na podstawie testów z 300 spraw

7-Position Value	Ratios (R/C)		
	RLL	EDR	BV
3	> 1.25	< .44	< .54
2	1.11 – 1.25	.44 – .67	.54 – .71
1	1.04 – 1.10	.68 – .92	.72 – .88
0	.97 – 1.03	.93 – 1.20	.89 – 1.05
-1	.88 – .96	1.21 – 1.60	1.06 – 1.29
-2	.79 – .89	1.61 – 2.44	1.30 – 1.67
-3	< .79	> 2.44	> 1.67

RLL – Respiration Line Length

EDR – Electrodermal Response

BV – Blood Volume

Źródło: D.J. Krapohl, B. McManus, *An Objective Method...*, op. cit., s. 212.

Zastąpienie subiektywnych szacunków relatywnej intensywności reakcji obiektywnymi pomiarami jasno określonych cech diagnostycznych stwarza warunki do przedstawienia prawdopodobieństwa błędu dla poszczególnych rezultatów liczbowej analizy (zob. tab. 20), z zastrzeżeniem, że przyjęty na potrzeby obliczeń wskaźnik bazowy osób szczerych i nieszczerych w populacji jest równy 50/50.

Tab. 20. Szacunki prawdopodobieństwa występowania określonych rezultatów testów w populacji osób szczerych i nieszczerych przy zastosowaniu obiektywnego systemu oceniania (OSS)

Wynik	Prawdopodobieństwo, że prawdziwy badany osiągnie podany wynik lub niższy	Prawdopodobieństwo, że nieszczerzy badany osiągnie podany wynik lub wyższy
-40	0.01	
-38	0.01	
-36	0.01	
-34	0.01	
-32	0.01	
-30	0.01	
-28	0.01	
-26	0.01	
-24	0.01	
-22	0.01	
-20	0.02	
-18	0.02	
-16	0.02	
-14	0.03	
-12	0.04	
-10	0.05	
-8	0.06	0.20
-6	0.07	0.18
-4	0.09	0.15
-2	0.11	0.13
0	0.13	0.11
+2	0.15	0.09
+4	0.18	0.08
+6	0.21	0.06
+8	0.24	0.05
+10		0.04
+12		0.03
+14		0.03
+16		0.02
+18		0.02
+20		0.01
+22		0.01
+24		0.01
+26		0.01
+28		0.01
+30		0.01
+32		0.01
+34		0.01
+36		0.01
+38		0.01
+40		0.01

Źródło: D. Krapohl, B. McManus, *An Objective Method...*, op. cit., s. 218.

Autorzy obiektywnej metody manualnej oceny zapisów ostrożnie sugerowali w 1999 r., że ich propozycję należy traktować jako

uzupełnienie dotychczasowych metod, a nie alternatywę dla nich. Zaznaczali jednocześnie, że obiektywizacja jest konieczna ze względu na wymogi obowiązujące w środowisku naukowym i z pewnością mieli nadzieję na większe upowszechnienie tej metody. Z perspektywy czasu można jednak powiedzieć, że koncepcja nie została wdrożona w powszechnej praktyce analizy manualnej, ale na pewno była jedną z inspiracji dla późniejszych metod, które wniosły mocny walor naukowy i niemal wyparły inne metody (mowa tu o manualnym Empirycznym Systemie Oceniania, wraz z jego ostatnią aktualizacją w postaci systemu ESS-M). OSS przebił się natomiast w wersji skomputeryzowanej. Producenci sprzętu do badań poligraficznych zaczęli uzupełniać funkcje oprogramowania o ten system. Równolegle trwały prace nad jego doskonaleniem. W 2002 r. opublikowano aktualizację, którą nazwano OSS-2⁹¹. Zmiana objęła wskaźniki R/C (zob. tab. 21) oraz wyliczenia prawdopodobieństwa błędu dla poszczególnych rezultatów analizy numerycznej (zob. tab. 22). Poprzednie wskaźniki R/C były dostosowane wyłącznie do jednego formatu testu – ZCT, nauczanego przez DoDPI. Zaktualizowane miały być odpowiednie również dla innych formatów testów ZCT (tych, gdzie każde pytanie krytyczne porównywane jest zawsze z jednym sąsiadującym pytaniem kontrolnym – np. Utah ZCT). Empirycznie potwierdzona dokładność identyfikacji przy wykorzystaniu systemu OSS-2 była podobna jak przy pierwszej wersji (tym razem w granicach 92,6–94,5% po odrzuceniu wyników nierozstrzygniętych na poziomie kilkunastu procent)⁹².

Tab. 21. Zasady punktacji w systemie OSS-2 (oceny numeryczne w skali 7-pozycyjnej w zależności od wartości wskaźnika R/C)

RLL	0.00 – 0.77	0.78 – 0.86	0.87 – 0.93	0.94 – 1.02	1.03 – 1.10	1.11 – 1.21	1.22 – 9999
score →	-3	-2	-1	0	+3	+2	+1
EDR	9999 – 2.60	2.59 – 1.72	1.71 – 1.27	1.26 – 0.95	0.94 – 0.70	0.69 – 0.45	0.44 – 0.00
score →	-6	-4	-2	0	+2	+4	+6
BV	9999 – 1.81	1.80 – 1.39	1.38 – 1.14	1.13 – 0.93	0.92 – 0.74	0.73 – 0.56	0.55 – 0.00
score →	-3	-2	-1	0	+3	+2	+1

Źródło: D.J. Krapohl, *Short Report Update...*, op. cit., s. 299.

⁹¹ D.J. Krapohl, *Short Report Update for the Objective Scoring System*, „Polygraph”, 31(4), 2002, s. 298–302.

⁹² Zob. *ibidem*, s. 300.

Tab. 22. Szacunki prawdopodobieństwa występowania określonych rezultatów testów w populacji osób szczerych i nieszczerych przy zastosowaniu systemu OSS-2

Wynik	Prawdopodobieństwo, że prawdziwy badany osiągnie podany wynik lub niższy	Prawdopodobieństwo, że nieszczerzy badany osiągnie podany wynik lub wyższy
-40	<0.01	
-38	0.01	
-36	0.01	
-34	0.01	
-32	0.01	
-30	0.01	
-28	0.01	
-26	0.01	
-24	0.01	
-22	0.01	
-20	0.02	
-18	0.02	
-16	0.02	
-14	0.03	
-12	0.03	
-10	0.04	
-8	0.05	0.23
-6	0.06	0.20
-4	0.07	0.17
-2	0.09	0.15
0	0.11	0.13
+2	0.13	0.11
+4	0.16	0.09
+6	0.19	0.07
+8	0.22	0.06
+10		0.05
+12		0.04
+14		0.03
+16		0.02
+18		0.02
+20		0.01
+22		0.01
+24		0.01
+26		0.01
+28		0.01
+30		0.01
+32		0.01
+34		0.01
+36		0.01
+38		0.01
+40		<0.01

Źródło: D.J. Krapohl, *Short Report Update...*, op. cit., s. 301.

Stosunkowo najnowszym i najczęściej wykorzystywanym systemem do algorytmicznej analizy danych z testów poligraficznych jest *objective scoring system* w wersji trzeciej (OSS-3). Pozostaje kompatybilny z większością programów przeznaczonych do pracy z poligrafami wiodących producentów na świecie, a autorzy udostępnili go bezpłatnie do powszechnego użytku. Algorytm opiera się na tych samych podstawach co OSS przedstawiony przez D. Krapohla i B. McManusa⁹³ i późniejsza

⁹³ D.J. Krapohl, B. McManus, *An Objective Method...*, op. cit.

modyfikacja D. Krapohla⁹⁴. Nowa wersja zawiera natomiast istotne ulepszenia modelu matematycznej transformacji i agregacji danych⁹⁵.

Teoretycznie ewaluacja zapisów na poligramach na podstawie reguł OSS-3 mogłaby być wykonana manualnie, ale złożoność obliczeń sprawia, że *de facto* – jest to algorytm już tylko komputerowy. Dane są uśredniane i odpowiednio ważone, więc można go stosować przy wszystkich testach typu CQT, zawierających od dwóch do czterech pytań krytycznych, zadawanych w czasie od trzech do pięciu serii testu (wykresów). Reguły decyzyjne dotyczące rezultatów analizy numerycznej nie są tożsame z tymi z oceny manualnej, ale opierają się na statystycznych wartościach *p*. Nowością było też wprowadzenie możliwości wyboru m.in. oceny dwuetapowej (czyli na podstawie sumy całkowitej w pierwszym etapie i sum cząstkowych – w etapie drugim). Poszczególnym kanałom zapisu przypisano następujące wagi: 0,5 dla EDA; 0,33 dla kardio i 0,17 dla pneumo⁹⁶. Dokładność identyfikacji szczerości i nieszczerości, przy wykorzystaniu OSS-3, oszacowano w próbie walidacyjnej na poziomie 94,7% (przy 5% wyników nierozstrzygniętych, czułości testu wynoszącej 93,3% i swoistości 86,7%)⁹⁷.

Przed uruchomieniem w programie komputerowym obliczeń algorytmicznych należy dokonać wyboru właściwego typu testu (ryc. 18). Wśród nich są te najważniejsze: ZCT, MGQT, DLST, POT i CIT. Następnie należy zaznaczyć opcję oczekiwanych reguł decyzyjnych (ryc. 19). Kolejnym krokiem jest przejście przez wszystkie pytania testowe i manualne zaznaczenie tych kanałów zapisu, przy których stwierdza się artefakty lub inny powód do wyłączenia danych z analizy. Finalnie zostaje zaprezentowany raport z wynikami (zob. ryc. 20 i 21). Taki raport zawiera m.in. wskazanie testu (np. znaczące reakcje – z j. ang. *significant responses*), prawdopodobieństwo szczerości osoby badanej, typ testu, przyjęte reguły analityczne, ocenę różnic w rozmieszczeniu artefaktów, statystyczne wartości *p*, wagi poszczególnych komponentów, miary i współczynniki logarytmiczne.

⁹⁴ D.J. Krapohl, *Short Report...*, *op. cit.*

⁹⁵ R. Nelson, M. Handler, D. Krapohl, *Introduction to the Objective Scoring System – version 3*, www.oss3.info/index.html [dostęp: 27.02.2021].

⁹⁶ R. Nelson, M. Handler, D. Krapohl, *Development and Validation of the Objective Scoring System, version 3*, www.oss3.info/poster1.html [dostęp: 27.02.2021].

⁹⁷ *Ibidem.*

Ryc. 18. Wybór typu ocenianego testu w skomputeryzowanym systemie OSS-3

1) Select the exam type and scoring

Exam Type:

Scoring Method:

2) Select three to five charts to score

Charts:

- #1: Exam 1 Chart 1 (Zone)
- #2: Exam 1 Chart 2 (Zone)
- #3: Exam 1 Chart 3 (Zone)
- #4: Exam 1 Chart 4 (Zone)
- #5: Exam 1 Chart 5 (Zone)
- #6: Exam 1 Chart 6 (Zone)

Only show charts with two or more relevant questions.

Only color the questions that are selected for scoring.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	2	3	4	5	6	4	3	2	4	XX		
X	I	ScR	N1	C1	R1	R2	C2	R3	R4	C3	N2		
X	I	ScR	N2	C2	R1	R4	C1	R2	R3	C3	N1	XX	
X	I	ScR	N1	C3	R4	R2	C1	R3	R1	C2	N2	XX	

Źródło: materiały własne autora.

Ryc. 19. Wybór reguł decyzyjnych dla ocenianego testu w skomputeryzowanym systemie OSS-3

1) Select the exam type and scoring

Exam Type:

Scoring Method:

2) Select three to five charts to score

Charts:

- #1: Exam 1 Chart 1 (Rank Acquaintance)
- #2: Exam 1 Chart 2 (Zone)
- #3: Exam 1 Chart 3 (Zone)
- #4: Exam 1 Chart 4 (Zone)
- #5: Exam 1 Chart 5 (Zone)
- #6: Exam 1 Chart 6 (Zone)

Only show charts with two or more relevant questions.

Only color the questions that are selected for scoring.

	1	2	11	12	13
X	I		4	XX	
X	I		C3	N2	
X	I		C3	N1	XX
X	I		C2	N2	XX
I	SR2				
I	SR2				

Źródło: materiały własne autora.

Ryc. 20. Fragment raportu w programie OSS-3 (wynik badania, wskazanie m.in. typu testu, reguł decyzyjnych, prawdopodobieństwa szczerości badanego)

Lafayette Instrument Company
Objective Scoring System - Version 3
 By Raymond Nelson, Mark Handler and Donald Krapohl (2007)

Result: **Significant Reactions**

Description: **p-value: 0.024 - Probability this result was produced by a truthful person**

Exam Type: Event Specific/Single Issue (Zone)

Scoring Method: OSS-3 Two-stage (Senter 2003)

Test of Proportions: None - No significant differences in artifact distribution

PF Name:

Report Date:

Subject:

Examiner: System Administrator

Spot Scores		Decision Alpha (1 tailed)		Components		
ID	p-value	Result	Setting	Value	Component	Weight
R1	< 0.001		NSR	0.050	Pneumo	0.19
R2	0.303		SR	0.050	EDA	0.53
R3	0.020		Bonferroni corrected alpha	0.017	Cardio	0.28
			Test of Proportions (1 tailed)	0.050		

Relevant Questions
Exam 1 Chart 2

Źródło: materiały własne autora.

Ryc. 21. Fragment raportu w programie OSS-3 (pomiar, obliczenia logarytmiczne)

Measurements (Kircher and Raskin 1988)						
Exam 1 Chart 2						
	C1	R1	C2	R2	C3	R3
P1	783	608	862	776	497	774
P2	762	584	700	802	544	710
EDA	27	126	15	36	365	179
Cardio	44	75	30	39	48	34
SE	118	96	114	113	93	132
Exam 1 Chart 3						
	C1	R1	C2	R2	C3	R3
P1	792	829	755	676	742	764
P2	762	847	630	684	756	707
EDA	14	399	37	21	63	30
Cardio	28	22	24	84	44	66
SE	191	200	212	187	185	230
Exam 1 Chart 4						
	C1	R1	C2	R2	C3	R3
P1	1052	840	770	819	1018	915
P2	870	737	825	753	930	816
EDA	69	156	192	109	74	140
Cardio	34	68	61	56	44	64
SE	371	306	300	307	383	337

Standardized Lognormal Ratios			
Exam 1 Chart 2			
	R1	R2	R3
P	-1.14	2.06	1.11
EDA	0.19	2.70	-0.52
Cardio	-3.00	0.13	0.85
WMean	-0.96	1.86	0.18
Mean		0.36	
Exam 1 Chart 3			
	R1	R2	R3
P	1.93	-0.77	0.37
EDA	-3.00	1.23	0.51
Cardio	1.88	-3.00	-3.00
WMean	-0.69	-0.34	-0.50
Mean		-0.51	
Exam 1 Chart 4			
	R1	R2	R3
P	-1.24	-1.04	0.00
EDA	-0.63	0.09	-0.41
Cardio	-2.12	-1.09	-1.80
WMean	-1.16	-0.46	-0.72
Mean		-0.78	

Źródło: materiały własne autora.

Na Uniwersytecie Śląskim w Katowicach prowadzono badania porównawcze między oceną manualną w skali 7-pozycyjnej według reguł systemu federalnego, opisanego przez J. Swinforda⁹⁸, a wskazaniem algorytmu OSS-2. Były to klasyczne badania laboratoryjne z wykorzystaniem zainscenizowanej kradzieży. W grupie badanych znalazło się 43 uczestników (23 mężczyzn i 20 kobiet). Spośród tych osób 19 przyjęło role winnych, a 23 – niewinnych. Zaangażowano 8 poligraferów, którzy mieli przeprowadzić testy (w formie Utah ZCT DLT), natomiast dwóch eksperymentatorów dokonywało pomiarów zmian w reakcjach fizjologicznych i oceniało zapisy.

Przy ocenie manualnej poprawnie zidentyfikowano 93% badanych (z wyłączeniem wyników nierozstrzygniętych na wysokim poziomie 35%). Natomiast algorytmy w ramach programu OSS-2 dały wskazania z dokładnością 78% (z wyłączeniem wyników nierozstrzygniętych, których odsetek wyniósł 21). Jeśli odrzucimy wyniki bez jednoznacznego wskazania – za pomocą oceny manualnej trafniej zidentyfikowano osoby nieszczerze (nawet w 100%, choć aż przy 42% wyników nierozstrzygniętych – wobec zaledwie 38% poprawnych identyfikacji pozytywnych wyłącznie manualnie). Natomiast algorytm OSS-2 lepiej sprawdził się przy badaniach osób prawdomównych (95% dokładnych

⁹⁸ J. Swinford, *Manually Scoring Polygraph Charts...*, op. cit.

wskazań, przy jedynie 3% wyników nierozstrzygniętych – wobec 80% trafnych identyfikacji za pomocą oceny manualnej)⁹⁹.

Autorzy zwracali uwagę, że wyższość oceny manualnej mogła w tym przypadku wynikać m.in. z różnej jakości zapisów (algorytm potrzebuje znacznie klarowniejszych poligramów niż człowiek), a także oparcia algorytmu w znacznej mierze na jednym parametrze – EDA, podczas gdy przy ocenie manualnej na równi analizowano wszystkie mierzone parametry. Z kolei niższa dokładność i zawyżona liczba wyników nierozstrzygniętych ogółem – w porównaniu do wyników raportowanych w badaniach amerykańskich – były najpewniej konsekwencją nie tylko ograniczeń związanych z eksperymentem typu laboratoryjnego (w tym niską motywacją badanych do skutecznego kłamstwa), ale również – jak się faktycznie okazało – zróżnicowanym poziomem umiejętności poligraferów, którzy przeprowadzali testy.

Poza ww. na przestrzeni lat powstały też inne algorytmy do analizy danych z testów poligraficznych – m.in. *chart analysis*, *identifi*, *white star* – oferowane przez Axciton Systems czy *Quest* – udostępniany przez Lafayette Instrument. Porównanie kilku wybranych algorytmów (AXCON 1.2, Chart Analysis 5.1-090-17-097, CPS 2.2, Identifi 1.43, Polyscore 4.0) nie wykazało istotnych różnic w dokładności identyfikacji i odsetkach wyników nierozstrzygniętych. Po wyłączeniu tych ostatnich dokładność mieściła się w przedziale od 0.88 do 0.91. Ponadto wszystkie programy, z wyjątkiem CPS, częściej błędnie klasyfikowały osoby szczerze jako nieszczerze niż odwrotnie¹⁰⁰. Dokonano także analizy porównawczej systemów OSS i *probability analysis* (PA). Skuteczność identyfikacji przy obu metodach również okazała się podobna (ok. 90%), z lekkim wskazaniem na korzyść OSS. Systemy dawały zgodne ze sobą poprawne wskazania w ok. 72% przypadków. Potwierdzono, że nadają się nie tylko do analizy danych z testów jednoprotokolowych typu ZCT, ale także wieloprotokolowych – typu MGQT¹⁰¹.

⁹⁹ M. Leśniak, M. Zubańska, *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)*, „European Polygraph”, 4, 2(12), 2010, s. 81–86.

¹⁰⁰ A.B. Dollins, D.J. Krapohl, D.W. Dutton, *Computer Algorithm Comparison*, „Polygraph”, 29(3), 2000, s. 237–247.

¹⁰¹ A.K. Webb, M.D. Handler, D.J. Krapohl, J.C. Kircher, *A Comparison of the Objective Scoring System and Probability Analysis*, „Polygraph”, 37(4), 2008, s. 250–255.

Od kilku lat trwają prace nad zautomatyzowaną wersją systemu ESS, która ma być włączona do oprogramowania współpracującego z poligrafem (zarysowano perspektywę ok. roku od wydania niniejszej monografii) i naśladować manualny sposób ewaluacji danych testowych dokonywanej przez poligraferów¹⁰². Zastosowano 15-sekundowe okno oceny reakcji (z j. ang. *evaluation window*) dla wszystkich kanałów zapisu. Dla EDA i kardio okno zaistnienia reakcji (z j. ang. *response-onset-window*) wyznaczono od wprowadzenia bodźca do 5 sekund następujących po odpowiedzi osoby badanej, albo 5 sekund od zakończenia prezentacji bodźca testowego, jeśli w teście nie następowały werbalne odpowiedzi.

W kanale pneumograficznym wyłączono z analizy dane w przedziale od 1,5 sekundy przed i 1,5 sekundy po odpowiedzi. Ma to na celu uniknięcie uwzględniania często pojawiających się zniekształceń cyklu oddechowego, związanych z momentem udzielania werbalnej odpowiedzi. Pomiar danych w przedmiotowym kanale odbywa się z wykorzystaniem odchylenia linii oddychania (z j. ang. *respiration line excursion*). Chodzi tu o wychwycenie zmniejszenia częstotliwości lub tłumienia oddychania.

W kanale EDA brane są pod uwagę dane od momentu wprowadzenia bodźca (z uwzględnieniem latencji wynoszącej 0,5 sekundy) do końca okna oceny reakcji. Szczyt reakcji algorytm identyfikuje jako zmianę kierunku krzywej z wznoszącej na opadającą, w przedziale od 2,5 sekundy po wprowadzeniu bodźca do końca okna oceny reakcji. Dodatkowy wierzchołek krzywej włącza się do analizy również po zakończeniu okna oceny reakcji, jeśli krzywa pozostaje wznosząca od 13,5 sekundy do pierwszego szczytu po granicy okna oceny reakcji. Z kolei, jeśli krzywa wznosi się jeszcze przed wprowadzeniem bodźca, algorytm szuka znaczącego, wyraźnego odchylenia po jego wprowadzeniu.

W kanale kardio reakcja mierzona jest jako maksymalna odległość pomiędzy jej rozpoczęciem w ramach okna zaistnienia reakcji do szczytu w czasie okna oceny reakcji. Uwzględnia się także taki szczyt reakcji, który zarysował się po oknie oceny reakcji, jeżeli podniesienie linii bazowej kardio wciąż trwa od 14,5 sekundy do pierwszego szczytu reakcji po oknie oceny.

¹⁰² Zob. R. Nelson, *Accuracy Effects for ESS...*, *op. cit.*; E. Lebrija, *Evidence-based Best Scoring Practices using ESS* [prezentacja podczas Seminarium Polskiego Towarzystwa Badań Poligraficznych], Warszawa, 5.11.2022.

Zautomatyzowany algorytm ESS wykorzystuje takie same progi decyzyjne jak dla manualnego wielomianowego systemu ESS-M. W dwóch próbach badawczych obejmujących łącznie 160 potwierdzonych przypadków ani razu nie nastąpiła zmiana diagnozy, zależna od przyjętej metody oceniania testów w formacie Federal ZCT (porównywano reguły obowiązujące w tradycyjnym 3-pozycyjnym systemie federalnym z systemem ESS w postaci zautomatyzowanej). Gdy wykorzystywano reguły systemu ESS oraz wielomianowe progi decyzyjne, korzystniej natomiast prezentowały się wskaźniki czułości i swoistości testów, a zarazem mniejszy był odsetek wyników nierozstrzygniętych – co przemawia za nowszą metodą analizy danych rejestrowanych przez poligraf.

4.2. Narzędzia wspomagające manualną interpretację niektórych danych

Oprócz algorytmów całościowo analizujących dane rejestrowane przez poligraf utworzono jeszcze narzędzia wspomagające manualną ocenę reakcji w pojedynczych parametrach, względnie trudniejszych w interpretacji, a mianowicie w kanałach zapisu zmian w amplitudzie pulsu oraz cyklu oddechowym (a ściślej – długości linii oddechu).

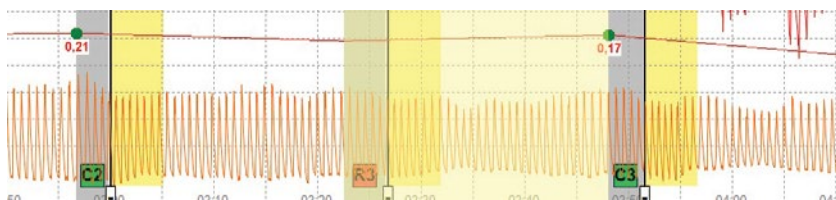
Cechą diagnostyczną poszukiwaną w kanale PLE/PPG jest redukcja amplitudy pulsu w reakcji na bodziec testowy. Zmiana ta jest często trudno uchwytna gołym okiem poligrafera, a w dalszym ciągu przydatna w diagnostyce. Fotopletysmograf wciąż pozostaje fakultatywnym czujnikiem poligrafu i przez poziom skomplikowania analizy danych rejestrowanych przez ten czujnik bywa po prostu pomijany, a jeśli już zostaje uwzględniony – występują spore rozbieżności między ocenianymi, co obniża rzetelność testu i podtrzymuje fakultatywny status komponentu PLE. Tracone są w ten sposób dane wartościowe, bo wysoko skorelowane z tzw. kryterium winy. Większość badań empirycznych potwierdza, że miary reakcji naczynioruchowych pozwalają na istotne statystycznie rozróżnienie badanych „winnych” i „niewinnych”. Część dowodzi, że korelacja z kryterium winy w przypadku amplitudy pulsu jest nawet wyższa niż zmian w ciśnieniu krwi i cyklu oddechowym¹⁰³. Dlatego naukowcy związani z Lafayette Instrument Co. skon-

¹⁰³ Zob. m.in. C.R. Honts, R. Reavy, *The Comparison Question Polygraph Test. A Contrast of Methods and Scoring*, „Physiology & Behavior”, 143, 2015, s. 15–26.

struowali komputerowe narzędzie do obliczeń stopnia supresji amplitudy pulsu przed i po wprowadzeniu bodźca testowego, a następnie porównania tych wskaźników. To narzędzie nazwano PLE Amplitude Tool, a obecnie dostępne jest, poza LX Software, również w ramach programów służących do obsługi poligrafów innych producentów.

Precyzyjnie rzecz ujmując – amplituda pulsu mierzona jest w przedziale 3 sekund przed wprowadzeniem bodźca oraz w 5-sekundowym przedziale rozpoczynającym się 5 sekund po wprowadzeniu bodźca (zadaniu pytania)¹⁰⁴. Pomiar amplitudy pulsu nie ma związku z fizycznymi wartościami ciśnienia krwi, a jest to efekt obliczenia jednostek cyfrowych odzwierciedlających różnicę między punktami skurczowymi i rozkurczowymi w wyżej określonych fragmentach zapisu. Wynik dla danego pytania krytycznego porównywany jest do każdego z pytań testowych. Na poligramie wskaźniki połączone są linią, która tworzy krzywą i wskazuje, które punkty są wyższe względem pozostałych. Im wyższy punkt – tym relatywnie największa redukcja amplitudy pulsu. Po zaznaczeniu wybranego pytania krytycznego przy pytaniach porównawczych zostaną naniesione kropki w określonych kolorach (zob. ryc. 22).

Ryc. 22. Przykładowy fragment poligramu z oznaczeniami wykonanymi z wykorzystaniem narzędzia PLE Amplitude Tool



Źródło: materiały własne autora.

Kolor czerwony oznacza istotniejszą reakcję przy pytaniu krytycznym, a zatem rekomendację dla ujemnej oceny numerycznej (-1). Zielony – odwrotnie (+1). Biały natomiast świadczy o tym, że nie wystąpiły na tyle znaczące różnice, aby brać je pod uwagę w punktacji (ocena 0). Wynik nadający się do wykorzystania jako podstawa do przyznania

¹⁰⁴ R. Nelson, *Tech Talk. Frequently Asked Questions about the LXSoftware PLE Scoring Tool*, „Lafayette Instrument”, 19.08.2016, www.researchgate.net/publication/306292694_Tech_Talk_Frequently_asked_questions_about_the_LXSoftware_PLE_scoring_tool [dostęp: 10.02.2021].

dotatniej lub ujemnej oceny musi przekroczyć 10% najmniejszego ze wszystkich możliwych losowych wyników (nie chodzi tu więc o miarę fizyczną, lecz o proporcję względem wszystkich możliwych rezultatów). W tym przypadku progi decyzyjne wynoszą $\pm 0,0993$. Pojawiające się przy tym wskaźniki liczbowe oznaczają stosunek amplitudy pulsu w analizowanych przedziałach (przed i po wprowadzeniu bodźca).

Ewentualnie uwzględnienie wskazań algorytmu pozostaje zawsze do decyzji poligrafera. Narzędzie nie jest w stanie odróżnić danych niestabilnych oraz artefaktów od prawidłowych zapisów. Biegły powinien więc zignorować wskazania, jeśli zakłócenia wystąpiły w obrębie przedziałów czasowych, w których dokonywane są automatyczne obliczenia lub wcześniejsze zdarzenia podczas testu miały wpływ na kształt krzywych w tych obszarach. Generalnie wykorzystanie algorytmu do analizy PLE przynosi jednak więcej pożytku. Zwiększa się wskaźnik niezawodności ocen i dokładność identyfikacji na podstawie danych z kompletu czujników. Raportowano ponadto, że zastosowanie narzędzia PLE Amplitude Tool pozwoliło na zwiększenie częstotliwości przyznawania punktacji różnej od zera – z 30% do 74,5%¹⁰⁵.

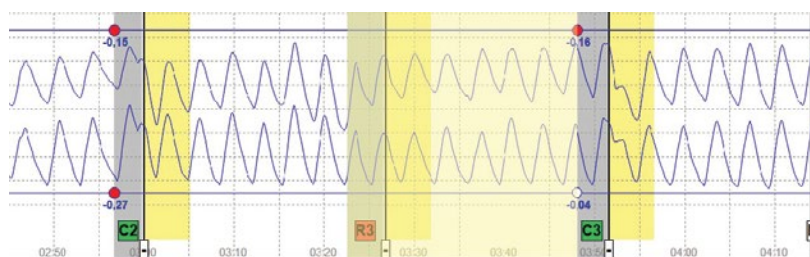
Na podobnej zasadzie jak powyższe narzędzie funkcjonuje rozwiązanie dla analizy komponentu oddechowego – *respiration line excursion tool*. Zmiany w cyklu oddechowym przy ocenie manualnej stosunkowo rzadko przekładają się na inne oceny numeryczne niż zero. Główną potwierdzoną naukowo przesłanką diagnozowania jest krótsza linia oddechu i jej graficzne przejawy w postaci takich zapisów, które pokazują różne formy supresji oddechowej. Jeśli jednak dana cecha diagnostyczna występuje zarówno przy pytaniu krytycznym, jak i dobranym do niego pytaniu porównawczym – nie pozostaje nic innego jak zestawić ze sobą długości linii oddechu. Bez dodatkowych narzędzi poligrafer musi to robić w sferze wyobraźni, rozciągając zapis danego odcinka krzywej do linii prostej. Algorytm komputerowy zrobi to jednak – rzecz jasna – bardziej precyzyjnie.

Co więcej – narzędzie RLE Tool mierzy nie tyle długość linii oddechu, co wyłącznie pionowe odchylenia krzywej związane z aktywnością fizjologiczną, podczas gdy linia oddechu uwzględnia także zmiany poziome, które nie wynikają z fizjologii, lecz sposobu zaprojektowania

¹⁰⁵ C.R. Honts, M. Handler, P. Shaw, M. Gougler, *The Vasomotor Response in the Comparison Question Test*, „Polygraph”, 44(1), 2015, s. 62–71.

poligrafu¹⁰⁶. Wynik RLE jest logarytmicznym stosunkiem reakcji na analizowane pytanie względem reszty pytań testowych. Ujemną ocenę numeryczną rekomenduje się, gdy rezultat RLE mieści się w przedziale pomiędzy $-0,0488$ i $-0,4055$. Te współczynniki odpowiadają liniowym stosunkom od 1,05 do 1,5. Dodatnia ocena numeryczna jest wskazana dla przedziału wyników od $+0,2231$ do $+0,4055$. To z kolei odpowiada stosunkom liniowym od 1,25 do 1,5. Natomiast ocena zerowa przypadnie dla wyników wykraczających poza przedstawione współczynniki. Graficzna prezentacja wyników wygląda tak jak w przypadku PLE Amplitude Tool (zob. ryc. 23). Nie inaczej niż w przypadku algorytmu dla PLE – posiłkując się algorytmem dla pneumo – ostateczny werdykt, jakie pytania ze sobą porównać, które zapisy uwzględnić, a które odrzucić – należy do biegłego.

Ryc. 23. Przykładowy fragment poligramu z oznaczeniami wykonanymi z wykorzystaniem narzędzia Respiration Line Excursion Tool



Źródło: materiały własne autora.

5. Badania empiryczne dotyczące zgodności ocen i diagnoz biegłych z zakresu badań poligraficznych

5.1. Zgodność ocen poligraferów - przegląd badań w piśmiennictwie

Poziom rzetelności testu wydaje się głównym wskaźnikiem sygnalizującym potencjalny zakres subiektywizmu biegłych, którzy się nim posłużyli. Początki badań w tym zakresie, udokumentowane w literaturze, sięgają lat 40. XX wieku. Prowadzono je przede wszystkim poprzez udostępnianie pewnej liczbie poligraferów jednakowych

¹⁰⁶ R. Nelson, *Using the Lafayette Respiration Line Excursion Tool*, „Polygraph Magazine”, 49(4), 2016, s. 59–62.

zestawów poligramów do ewaluacji na zasadzie ślepej interpretacji. Pierwsze porównania odbywały się przy zaangażowaniu bardzo małej liczby poligraferów, więc ich znaczenie mogło co najwyżej pilotażowo sygnalizować potencjalne tendencje. Na przykład F. Rouke przekazał dwóm biegłym do oceny zapisy reakcji elektrodermalnych uzyskanych w toku eksperymentu prowadzonego w warunkach laboratoryjnych, ze znanym eksperymentatorowi stanem faktycznym. Poligraferzy dokonali identyfikacji na temat „winy” sprawcy wyreżyserowanego czynu z dokładnością 91% i 88%. Rozpiętość diagnoz wyniosła tu tylko 3%, ale jednak wystąpiła¹⁰⁷.

Na większej próbie biegłych (choć wciąż dalekiej od pożądanej, bo sześciu) podobne badanie przeprowadził W. Holmes w latach 50. XX wieku. Posłużył się poligramami z 25 rzeczywistych spraw kryminalnych prowadzonych przez policję w Miami. Rozpiętość trafnych diagnoz wyniosła tu 12% [0,69–0,81]¹⁰⁸. W eksperymencie J. Kubisa z początku lat 60. było to 19% [0,73–0,92], a deklarowanym parametrem, który budził najmniej wątpliwości był GSR¹⁰⁹.

Bardziej reprezentatywne były z pewnością wyniki eksperymentu realizowanego na zlecenie Departamentu Obrony USA w drugiej połowie lat 60. Tym razem zaangażowano 30 poligraferów, którzy oceniali zapisy z testów wykonanych trzema różnymi technikami (Backstera, Reida i POT). Otrzymali po 30 kompletów poligramów dla każdej z nich. W równych proporcjach przydzielono testy, które dotyczyły badanych szczerych i nieprawdomównych. Zgodność ocen mierzono współczynnikiem Kappa Cohena. Najwyższy odnotowano dla rezultatów testów wykonanych techniką Backstera (0,51), a najniższy dla testów szczytowego napięcia (0,15)¹¹⁰.

Na początku lat 70. wyniki analogicznych badań opublikowano w Japonii. A. Suzuki *et al.* raportowali, że zgodność między ocenianymi poligramy, wyrażona współczynnikiem Spearmana–Browna, wyniosła 0,798 (w próbie uczestniczyło 26 poligraferów, którzy

¹⁰⁷ F.L. Rouke, *Evaluation of Indices of Deception in the Psychogalvanic Technique* [niepublikowana rozprawa doktorska], Fordham University, 1941 – za: G.H. Barland, *The Reliability of Polygraph...*, *op. cit.*, s. 192.

¹⁰⁸ W.D. Holmes, *The Degree of Objectivity...*, *op. cit.*, s. 67–70.

¹⁰⁹ J.F. Kubis, *Studies in Lie-detection Computer...*, *op. cit.*

¹¹⁰ P.J. Bersh, R.A. Brisentine, *The Reliability of Blind Interpretation of Polygraph Record for Lie-detection Purposes (Report Prepared for DoD)*, 1968; za: J. Widacki, *Wprowadzenie do problematyki...*, *op. cit.*, s. 135–136.

analizowali zapisy z 30 spraw)¹¹¹. W podobnym czasie inni badacze – F. Horvath i J. Reid – wyselekcjonowali 40 zestawów poligramów zarejestrowanych przy wykonywaniu testów techniką Reida¹¹² (po 20 zweryfikowanych jako pochodzących od osób winnych i niewinnych), a następnie przekazali je do ewaluacji poligraferom o zróżnicowanym doświadczeniu zawodowym. W ramach tej selekcji odrzucili te zapisy, które były tak jednoznaczne, że nawet laik dostrzegłby różnice w reakcjach na pytania krytyczne i kontrolne, a także te nienadające się do interpretacji nawet przez wykwalifikowanego poligrafera. Oceniającym przekazano przy poszczególnych testach jedynie ogólnikowe informacje o tematyce badań. Na wykonanie zadania otrzymali tylko jeden dzień roboczy. W gronie siedmiu poligraferów, którzy legitymowali się co najmniej rocznym stażem praktyki, dokładność diagnoz wyniosła 91,4%, a rozpiętość trafnych ocen – 12,5% [0,85–97,5]. Trzech pozostałych, którzy stawiali dopiero swoje pierwsze kroki na ścieżce kariery, oceniało z dokładnością na poziomie 79,1% [0,70–0,90]. Ogółem odsetek poprawnych diagnoz oszacowano na 87,5%, a rozbieżność między trafnymi diagnozami poligraferów na 27,5%¹¹³. Lepsze rezultaty osiągały osoby bardziej doświadczone. Ponadto trafniej identyfikowano osoby niewinne (9,5% wyników fałszywych pozytywnych wobec 15% wyników fałszywych negatywnych). F. Horvath z J. Reidem dowiedli przede wszystkim tego, że eksperci z zakresu badań poligraficznych mogą osiągać dokładne i rzetelne wyniki jedynie na podstawie analizy zapisów na poligramach, bez znajomości szczegółowych okoliczności sprawy i bez osobistej interakcji z osobą badaną. Uważali jednak, że w lepszej sytuacji są eksperci osobiście przeprowadzający testy i w pełni zaznajomieni ze sprawą. Dodatkowo wskazówki behawioralne uznali za coś, co sprzyja trafniejszym diagnozom w połączeniu z zarejestrowanymi danymi fizjologicznymi. Trzeba przy tym zauważyć, że taki pogląd sformułowali dość arbitralnie.

¹¹¹ A. Suzuki, S. Watanabe, K. Ohnishi, K. Matsuno, M. Arasuna, *Polygraph Examiners' Judgements in Chart Interpretation – Reliability of Judgement*, „Kagaku Keisatsu Kenkyunsho” (Police Science Report), 26(1), 1973, s. 34 i nast.

¹¹² Zob. opis techniki Reida w: J. Reid, F.E. Inbau, *Truth and Deception...*, *op. cit.*

¹¹³ F. Horvath, J. Reid, *The Reliability of Polygraph Examiner Diagnosis of Truth and Deception*, „Journal of Criminal Law, Criminology and Police Science”, 62(1), 1971, s. 276–281.

Czasem zachowanie osoby badanej pomaga w prawidłowej ocenie wiarygodności wypowiedzi, a czasem może jednak przeszkadzać (*vide* błąd Otella). Prawidłowo postępujący poligrafer świadomie sugeruje się zachowaniem osoby badanej, obok zapisów na poligramach, jedynie przy prowadzeniu wywiadu przedtestowego i rozmowy po testach, a także formułowaniu pytań kontrolnych. Analiza sposobu reagowania na pytania krytyczne powinna odbywać się wyłącznie na zapisach zarejestrowanych przez poligrafa.

F. Horvath opublikował kilka lat później wyniki kolejnego eksperymentu, w którym brało udział 10 poligraferów zatrudnionych w policji¹¹⁴. Sklasyfikowano ich w dwóch równych liczebnie podgrupach: ze stażem praktyki powyżej trzech lat i do trzech lat. Dokonywali „ślepej” interpretacji zapisów na poligramach z badań 112 podejrzanych w sprawach kryminalnych z lat 1969–1972, dotyczących przestępstw przeciwko osobie oraz przeciwko mieniu. Połowa z poligramów była z oryginalnym wskazaniem na winę, a druga połowa na niewinność osoby badanej. W obu grupach po równo rozdzielono sprawy z potwierdzonym stanem faktycznym (kiedy w sprawie uzyskano przyznanie się faktycznego sprawcy do winy) i bez takiego potwierdzenia. Oceniający dokonali ogółem 63,1% trafnych identyfikacji (przy czym w odróżnieniu od większości eksperymentów wyniki nierozstrzygnięte potraktowano jako nietrafne). Zgodność między oceniającymi oszacowano współczynnikiem Hoyta na 0,89 w sprawach potwierdzonych i 0,85 w niepotwierdzonych. Jeżeli chodzi o inne ważne wnioski z tego eksperymentu – nie wykazano istotnych statystycznie różnic w dokładności diagnoz w zależności od doświadczenia poligraferów. Lepiej identyfikowano osoby winne (uzyskano mniej błędów fałszywych negatywnych niż fałszywych pozytywnych – co w tym wypadku odbiegało od wcześniejszych eksperymentów, gdzie było odwrotnie). Oceniający wykazywali się też większym poziomem pewności, kiedy oceniali zapisy z badań osób faktycznie nieszczerych. Łatwiejsze do interpretacji były poligramy ze spraw dotyczących przestępstw przeciwko osobie (prawdopodobnie ze względu na silniejszy ładunek emocjonalny) niż przeciwko mieniu. Najmniej problemów z interpretacją poligraferzy mieli z kanałem oddechowym, a najwięcej z kanałem GSR (co współcześnie praktykujących powinno raczej zaskakiwać).

¹¹⁴ F. Horvath, *The Effects of Selected Variables...*, *op. cit.*

Tab. 23. Rozpiętość trafnych opinii biegłych z zakresu badań poligraficznych w wybranych badaniach naukowych (chronologicznie)

Autorzy (rok publikacji)	Rozpiętość ocen trafnych (%)	Rozbieżność ocen między biegłymi (%)
Rouke (1941)	88–91	3
Holmes (1957)	69–81	12
Kubis (1962)	73–92	19
Horvath, Reid (1971)	70–97,5	27,5
Hunter, Ash (1973)	75–90	15
Slowik, Buckley (1975)	70–100	30
Wicklender, Hunter (1975)	70–95	25
Widacki (1977)	67,5–85	17,5
Średnia		18,6

Źródło: opracowanie własne autora.

Eksperyment nad rzetelnością testów z wykorzystaniem poligrafu przeprowadzili również F. Hunter i P. Ash, ale nie dotyczył on zgodności między oceniającymi, lecz aspektu powtarzalności wyników u tego samego oceniającego. Grupie siedmiu poligraferów badacze przedłożyli do ewaluacji 20 zestawów poligramów z rzeczywistych spraw, w których stan faktyczny był niezależnie potwierdzony. Zapisy (z testów przeprowadzonych techniką Reida) dotyczyły po połowie osób prawdomównych i nieszczerých. Poligraferzy analizowali poligramy w co najmniej trzymiesięcznym odstępie czasu. Nie wiedzieli, że po raz drugi otrzymali te same zapisy, ułożone w przypadkowej kolejności. Wyłączne informacje, jakie posiadali – to ogólny temat sprawy, np. *kradzież pewnej sumy pieniędzy w jednej z krajowych sieci restauracji*¹¹⁵. Zostali poinstruowani, aby nie kontaktowali się z pozostałymi oceniającymi. Zadanie wykonali średnio w ciągu 76 minut¹¹⁶. Zgodność pomiędzy rezultatami pierwszych analiz a kolejnymi wystąpiła w 85%

¹¹⁵ F.L. Hunter, P. Ash, *The Accuracy and Consistency of Polygraph Examiners' Diagnoses*, „Journal of Police Science and Administration”, 1(3), 1973, s. 370–371.

¹¹⁶ Tak krótki czas może zastanawiać i wskazywać, że analizy nie były zbyt wnikliwe – a to z kolei mogło przełożyć się na poprawność diagnoz. Z osobistej praktyki zawodowej autor monografii wie, że tyle czasu może zajmować porządna analiza jednego zestawu poligramów, a nie dwudziestu, a w przypadku bardziej skomplikowanych zapisów – nawet wielokrotnie dłużej. Tylko przy wyjątkowo prostych i jednoznacznych danych – ich ewaluacja trwa krócej.

przypadków. U poligrafera, który był najmniej spójny w ocenach, było to 75%, a u najbardziej konsekwentnego – 90%. Ten pierwszy miał najkrótszy staż praktyki, a drugi – najdłuższy, co może sygnalizować znaczenie doświadczenia zawodowego dla rzetelności analizowanych testów. Ogólna dokładność wszystkich powtórnych diagnoz wyniosła 86% (na przeciętnym poziomie).

Większość badań nad zgodnością diagnoz biegłych z zakresu badań poligraficznych, które prowadzono do lat 70. XX wieku, odnosiła się do testów wykonanych technikami: pytań kontrolnych Reida, szczytowego napięcia (POT) oraz pytań związanych i niezwiązanych (R/I). Badania oceniane były metodą jakościową. Natomiast metoda numeryczna była dopiero w fazie pierwszych prób, a nieodłącznie towarzyszyła jedynie technice Backstera (w skali 7-pozycyjnej). Do nielicznych, którzy próbowali w tym czasie zastosować częściowo zobiektywizowaną metodę numeryczną, należeli wspomniany wcześniej J. Kubis, a także G. Barland z D. Raskinem.

Dwaj ostatni naukowcy (z Uniwersytetu Utah) przeprowadzili eksperyment, w którym badaniom poligraficznym techniką Backstera poddano 72 studentów na okoliczność zainscenizowanego zdarzenia (chodziło o zabranie 10 dolarów, które można było zatrzymać, jeśli uda się przekonać poligrafera, że się tego nie zrobiło). Studentów podzielono na dwie tożsame liczebnie grupy (winnych i niewinnych), a następnie ustalono losową kolejność badań. Analiza uzyskanych zapisów reakcji fizjologicznych osób badanych metodą numeryczną w skali 7-pozycyjnej (przyjęto progi decyzyjne: +4 dla NDI i -4 dla DI) doprowadziła do 81% poprawnych identyfikacji, przy 35% rezultatach bez rozstrzygnięcia. Gdy punkt odcięcia ustalono na wartości zerowej, dokładność obniżyła się do 71%, ale za to liczba wyników nierozstrzygniętych została zredukowana do 4%. Zaobserwowano przy okazji, że zapisy w kanale GSR były miarodajne dla wszystkich trzech zarejestrowanych wykresów. Kanał oddechowy efektywnie różnicował zmiany reakcji na pytania krytyczne i kontrolne w czasie dwóch pierwszych wykresów, a kanał kardio jedynie podczas pierwszej serii testu. W drugim etapie eksperymentu wykorzystano 26 zestawów poligramów z pierwszego etapu i udostępniono je do ślepej analizy sześciu praktykującym poligraferom, którzy osiągnęli dokładność na poziomie 79% (przy 38% wyników nierozstrzygniętych). Były to więc rezultaty zbliżone do tych z pierwszego etapu. W ten sposób wykazano rzetelność sposobu oceniania zapisów metodą numeryczną. Generalnie bardziej

klarowne wyniki pochodziły z testów dotyczących osób winnych. Autorzy tłumaczyli to dwojako – z jednej strony większą motywacją tej części probantów, a ponadto prawdopodobną niedoskonałością pytań porównawczych, które nie tak dobrze stymulowały niewinnych badanych niż pytania krytyczne tych uczestników, którzy zabrali pieniądze¹¹⁷. Wykorzystując dane z tego eksperymentu, G. Barland dokonał jeszcze innego ważnego porównania, a mianowicie oszacował zgodność oceniających w poszczególnych parametrach mierzonych przez poligraf. Zatem w kanale pneumo było to 0,65; w kanale EDA – 0,90, a kardio – 0,76. Łącznie dla wszystkich parametrów – 0,86¹¹⁸.

Korelację ocen poligraferów przydzielanych indywidualnym parametrom sprawdzał również B. Jayne. W jego badaniu trzech poligraferów oceniało 100 zestawów poligramów (50 pochodziło od badanych szczerych, a 50 od nieszczerych). Testy były wykonane techniką Reida, a oceniający mieli za zadanie przeprowadzić dwojaką analizę – ilościowo zmierzyć wielkość reakcji na pytania krytyczne oraz kontrolne (czas trwania istotnej zmiany w kanałach: pneumo i kardio oraz amplitudę reakcji w kanale GSR), a także zastosować metodę analizy numerycznej w skali 3-pozycyjnej i na jej podstawie postawić diagnozę. Korelacja ocen ilościowych między parami poligraferów, wyrażona współczynnikiem Pearsona, wyniosła średnio: dla kanału pneumo – 0,73; dla kanału kardio – 0,68; dla kanału GSR – 0,79¹¹⁹. Ogółem w trzech parametrach: 0,75. Poprzez pomiary ilościowe trafnie zidentyfikowano 81% prawdomównych badanych i 85% nieszczerych. W obu przypadkach uzyskano 6% wyników nierozstrzygniętych. Natomiast jakościowo-ilościową metodą numeryczną diagnozowano z dokładnością 82% (jednakowo osoby szczerze i nieszczerze; przy rezultatach nierozstrzygniętych wynoszących odpowiednio: 9% i 12%) oraz zgodnością na poziomie 85%.

E. Edel i J. Jacoby zaprezentowali jeszcze inne podejście do zagadnienia rzetelności ocen biegłych z zakresu badań poligraficznych. Byli zainteresowani tym, jak zgodni będą poligraferzy w podejmowaniu decyzji – czy w danej strefie zapisów reakcji na wykresie wystąpiła istotna reakcja fizjologiczna czy nie, a jeśli tak – to jakiego typu

¹¹⁷ G.H. Barland, D.C. Raskin, *An Experimental Study of Field Techniques in „Lie Detection”*, [prezentacja dla Society for Psychophysiological Research, St. Louis, 24.10.1971], „Psychophysiology”, 9, 1972, s. 275 i nast.

¹¹⁸ G.H. Barland, *The Reliability of Polygraph...*, *op. cit.*, s. 199.

¹¹⁹ B.C. Jayne, *Contributions of Physiological Recordings...*, *op. cit.*, s. 108.

(w jakim parametrze). W tym celu wykorzystano zapisy z 40 badań (z łączną liczbą 2530 pytań). Poligraferzy osiągnęli wysoki stopień spójności ($p < 0,0001$). Wykazali się ogólną zgodnością na poziomie 95% (96% w kanale kardio; 93% – GSR; 96% – pneumo)¹²⁰.

Blizsze współczesności badania naukowe w zakresie zgodności ocen poligraferów były prowadzone w związku z potrzebą walidacji różnych technik i – w odróżnieniu od tych początkowych – uwzględniały w zasadzie już tylko metody numeryczne. Poza Empirycznym Systemem Oceniania nie były to badania, których głównym przedmiotem były sposoby interpretacji danych testowych, lecz standaryzowane testy. Natomiast w związku z istnieniem kilku potwierdzonych naukowo systemów ewaluacji zapisów poligrafu, w różnych badaniach wykorzystywano inny system. W ten sposób dysponujemy danymi dla tych systemów w powiązaniu z określonym typem testu, a podstawowe znaczenie mają tu testy typu ZCT – diagnostyczne, o charakterze jednoprotokolowym lub wieloaspektowym, a zatem najczęściej wykorzystywane na potrzeby dowodowe w postępowaniach karnych.

Zgodność ocen numerycznych przypisywanych dla każdego mierzonego parametru w testach ZCT, analizowanych z wykorzystaniem systemu ESS (3-pozycyjnej metody numerycznej), zmierzona współczynnikiem Fleiss Kappa¹²¹ przez M. Handlera *et al.*, wyniosła 0,59 (95% CI = .52, .65)¹²². W tym eksperymencie brało udział 19 niedoświadczonych poligraferów z meksykańskiej policji federalnej, którzy oceniali 100 zestawów poligramów udostępnionych przez Departament Obrony USA. W innych badaniach szacowano to w przedziałach: 0,56¹²³ – 0,61¹²⁴ u niedoświadczonych poligraferów i 0,57¹²⁵ – 0,61¹²⁶ u tych z dłuższym stażem praktyki. Z kolei zgodność

¹²⁰ E.C. Edel, J. Jacoby, *Examiner Reliability in Polygraph Chart Analysis. Identification of Physiological Responses*, „Journal of Applied Psychology”, 60(5), 1975, s. 632–634.

¹²¹ Współczynnik Fleiss Kappa < 0 oznacza słabą zgodność; 0,01–0,20 – niewielką; 0,21–0,40 – przeciętną; 0,41–0,60 – umiarkowaną; 0,61–0,80 – znaczącą; 0,81–1,00 – niemal perfekcyjną.

¹²² M. Handler *et al.*, *Empirical Scoring System...*, *op. cit.*, s. 205.

¹²³ B. Blalock, B. Cushman, R. Nelson, *A Replication and Validation...*, *op. cit.*, s. 281–286.

¹²⁴ Zob. R. Nelson, D.J. Krapohl, M. Handler, *Brute Force Comparison...*, *op. cit. Ibidem.*

¹²⁶ R. Nelson, B. Blalock, M. Oelrich, B. Cushman, *Reliability of the Empirical Scoring System with Expert Examiners*, „Polygraph”, 40(3), 2011, s. 134.

decyzji (diagnoz) u Handlera *et al.* ustalono na średnim poziomie 0,84 (95% CI = .73, .95)¹²⁷. W kolejnym badaniu (z udziałem 25 oceniających), przy wykorzystaniu tych samych poligramów, średnia zgodność wyniosła 0,95¹²⁸.

Przy ocenianiu testów ZCT, na podstawie reguł systemu Rządu Federalnego USA, zgodność ocen poligraferów badała m.in. N. Blackwell, która zaangażowała trzech profesjonalnych poligraferów do „ślepej” interpretacji 100 zestawów poligramów. W skali 7-pozycyjnej zgodność ocen (Fleiss Kappa) badających oszacowano na 0,57. W skali 3-pozycyjnej tego samego systemu było to 0,36. Średnia zbieżność decyzji w parach ewaluatorów wyniosła 0,80 przy skali 7-pozycyjnej i 0,66 przy skali 3-pozycyjnej. Z wynikami manualnej analizy poligraferów zestawiono także wyniki ewaluacji za pomocą algorytmów w programie komputerowym Polyscore 3.3. Dla szerszej skali uzyskano zbieżność na poziomie 0,75. Dla węższej – 0,66¹²⁹. Patrząc na cząstkowe oceny reakcji w różnych kanałach zapisu – według danych z jednego z eksperymentów realizowanych na Uniwersytecie Utah – średnia zgodność między 32 oceniającymi za pomocą 7-pozycyjnej skali federalnej wyniosła: 0,89 dla kanału EDA, 0,80 dla kardio i 0,50 w kanale oddechowym¹³⁰.

Z kolei w testach typu MGQT zgodność decyzji poligraferów w skali 7- oraz 3-pozycyjnej wyniosła odpowiednio: 0,57 i 0,49¹³¹. Interesujące, że – biorąc pod uwagę ostateczne diagnozy, a nie poszczególne oceny numeryczne – w eksperymentach Blackwell to w skali szerszej występowała wyższa zgodność biegłych niż w węższej, z teoretycznie mniejszą przestrzenią na rozbieżności.

Przy wykorzystaniu systemu Utah korelacja ocen numerycznych wahała się w granicach 0,92–0,97, w innych badaniach – wyrażona współczynnikiem Fleiss Kappa została oszacowana na 0,73. Natomiast średnią zbieżność decyzji między poligraferami raportowano na poziomie 96–100% (po odrzuceniu wyników nierozstrzygniętych)¹³².

¹²⁷ M. Handler *et al.*, *Empirical Scoring System...*, *op. cit.*

¹²⁸ R. Nelson *et al.*, *Reliability...*, *op. cit.*, s. 135.

¹²⁹ N.J. Blackwell, *PolyScore 3.3 and...*, *op. cit.*, s. 149–175.

¹³⁰ J.C. Kircher *et al.*, *Human and Computer Decision-making...*, *op. cit.*, s. 17.

¹³¹ N.J. Blackwell, *PolyScore 3.3 and...*, *op. cit.*, s. 160.

¹³² Zob. J.C. Kircher, D.C. Raskin, *Human versus Computerized Evaluations...*, *op. cit.*; C.R. Honts, D.C. Raskin, J.C. Kircher, *Effects of Physical Countermeasures and Their Electromyographic Detection During Polygraph Tests for Deception*, „Psychophysiology”, 1, 1987, s. 241–247; B.G. Bell *et al.*, *The Utah Numerical...*, *op. cit.*, s. 2.

W badaniach S. Sentera *et al.* zgodność oceniających, którzy posługiwali się regułami systemu Utah, była – choć nieznacząco, ale jednak – wyższa niż przy zastosowaniu reguł wywodzących się z Departamentu Obrony USA (odpowiednio, uśredniając: 0,82 i 0,71)¹³³.

Jeżeli zaś chodzi o tzw. jakościową metodę interpretacji zapisów na poligramach – obecnie dopuszczalna jest właściwie tylko w dwóch rodzajach testów: POT oraz R/I. Dlatego nie znajdziemy współczesnych badań dotyczących rzetelności testów typu ZCT tą metodą. Jak wspomniano wyżej – w badaniach prowadzonych na rzecz Departamentu Obrony USA w latach 60. zgodność ocen jakościowych w testach typu POT, wyrażona współczynnikiem Kappa Cohena, wyniosła zaledwie 0,15¹³⁴. Przy 30 wykresach zbieżność decyzji w porównywanych parach wahała się między 0,30 a 0,93¹³⁵.

Natomiast zgodność decyzji poligraferów posługujących się inną techniką z oceną globalną – R/I – weryfikowali empirycznie D. Krapohl oraz jego współpracownicy w dwóch kolejnych eksperymentach, których wyniki opublikowano w 2014 r. i 2015 r. W pierwszym eksperymencie wzięło udział czterech certyfikowanych poligraferów z minimum 14-letnim doświadczeniem w stosowaniu tej techniki. Przekazano im zapisy ze 100 zweryfikowanych spraw i – aby kontrolować zmęczenie – pozwolono analizować materiały z nie więcej niż dziesięciu badań na dobę. Poligramy pochodziły z wspomnianego już projektu badawczego z lat 90., w którym przesiewowym badaniom poligraficznym poddawano kandydatów do pracy na stanowiskach w firmie ochroniarskiej świadczącej usługi na rzecz lotniska w Atlancie, w stanie Georgia. Standardowy test zawierał wówczas cztery zagadnienia krytyczne. Jako reprezentatywne dla grupy badanych nieszczerych wybrano poligramy z takich testów, w których osoby badane kłamały tylko na temat jednego zagadnienia – zażywania marihuany. Pozostałe (w równej liczbie) dotyczyły osób odpowiadających szczerze. Drugi eksperyment opierał się na tych samych zapisach, natomiast analizowała je inna, nieco liczniejsza, 11-osobowa grupa poligraferów

¹³³ Zob. S.M. Senter, A.B. Dollins, D.J. Krapohl, *A Comparison of Polygraph Data Evaluation Conventions Used at the University of Utah and the Department of Defense Polygraph Institute*, „Polygraph”, 33(4), 2004, s. 219.

¹³⁴ J. Bresh, R.A. Brisentine, *The Reliability of Blind Interpretation...*, *op. cit.*

¹³⁵ U.S. Department of Defense, *Present Status of DoD Research on the Polygraph: Report of the DoD Joint Services Group On a Coordinated R&D Program of Lie Detection Research*, 28.08.1968, s. 9.

z Departamentu Bezpieczeństwa Publicznego w Teksasie, o przeciętnym 6-letnim stażu praktyki. Miesięcznie wykonywali ok. 12 badań przesiewowych. Rezultaty w zakresie zgodności oceniających porównują od techniki R/I wiele do zyczenia. Średnia zbieżność opinii w pierwszym eksperymencie wyniosła 0,597¹³⁶, zaś w drugim – 0,604¹³⁷.

Na mankamenty techniki R/I, gdzie obowiązuje jakościowa ewaluacja zapisów, zwracał uwagę Ch. Sevilla, mówiąc:

Jeśli użyłeś techniki R/I – zakładając, że prawnik po drugiej stronie odrobił swoją pracę domową – zapyta cię, dlaczego nie skorzystałeś z techniki pytań kontrolnych. [...] Jeśli ta osoba naprawdę jest niewinna, reaguje na pytania relewantne, a ty nie używasz pytań kontrolnych – skąd wiesz, że jest niewinna? [...] Potem umieści na tablicy uproszczony przykład techniki pytań kontrolnych, gdzie człowiek wytwarza taką samą ilość reakcji na pytanie relewantne, a pytania kontrolne pokazują gigantyczne reakcje. Następnie sąd zada ci kilka pytań, dlaczego wykorzystałeś technikę R/I oraz kilka pytań w odniesieniu do jej trafności¹³⁸.

Tab. 24. Zgodność między oceniającymi poligramy z wykorzystaniem różnych współczesnych metod interpretacji danych testowych w wybranych eksperymentach i analizach naukowych

Autorzy opracowania (eksperymentu)	Rodzaj techniki badawczej (testu)	Metoda analizy danych testowych	Zgodność ocen (Fleiss Kappa)	Zbieżność decyzji (diagnoz) w parach oceniających
Nelson <i>et al.</i> (2008); Blalock <i>et al.</i> (2009); Handler <i>et al.</i> (2010); Nelson <i>et al.</i> (2011)	ZCT	numeryczna – ESS	0,56–0,61	84–95%
Blackwell (1999)		numeryczna – Federal – 7-poz.	0,57	80%
		numeryczna – Federal – 3-poz.	0,36	66%
Honts <i>et al.</i> (1987); Kircher, Raskin (1988)			numeryczna – Utah	0,73
U.S. Department of Defense (1968)	POT	jakościowa	b.d.	30–93%
Krapohl, Rosales (2014); Krapohl, Goodson (2015)	R/I	jakościowa	b.d.	60%

Źródło: opracowanie własne autora.

¹³⁶ D. Krapohl, T. Rosales, *Decision...*, *op. cit.*, s. 24.

¹³⁷ D. Krapohl, W. Goodson, *Decision Accuracy...*, *op. cit.*, s. 201.

¹³⁸ C. Sevilla, *The Polygraph Examiner...*, *op. cit.*, s. 127.

5.2. Własne badania dotyczące dokładności i zgodności diagnoz poligraferów posługujących się różnymi metodami analitycznymi

Choć – jak z powyższego wynika – ogólna zgodność między oceniającymi w badaniach poligraficznych była w przeszłości weryfikowana empirycznie niejednokrotnie – pozostawała przestrzeń naukowa do porównania efekty w odniesieniu do jednakowych danych testowych przyniesie zastosowanie różnych metod analizy. W latach 2014–2018 uczestniczyłem w projekcie naukowo-badawczym pt. *Instrumentalne i nieinstrumentalne metody detekcji nieuczciwości – problemy kryminalistyczne, etyczne i prawne*. Projekt realizowano w Krakowskiej Akademii im. Andrzeja Frycza Modrzewskiego w ramach grantu Narodowego Centrum Nauki¹³⁹. W związku z przyjętą tematyką niniejszej pracy zajmowałem się m.in. zagadnieniem subiektywizmu w badaniach poligraficznych. Do tego czasu w Polsce właściwie tylko jedno badanie eksperymentalne (i to z udziałem zaledwie dwóch biegłych) poświęcono zgodności ocen testów poligraficznych. Pod koniec lat 70. dwóch poligraferów o zbliżonych kwalifikacjach podjęło się niezależnie od siebie analizy danych z 80 badań laboratoryjnych z rozwiązaniem znanym organizatorowi eksperymentu. Pierwszy z poligraferów wydał wówczas 85% dokładnych diagnoz, drugi – 67,5%¹⁴⁰.

W omawianym grantie NCN do badań eksperymentalnych zaangażowano 15 profesjonalnych poligraferów, z których trzech osobiście przeprowadzało badania probantów, a pozostali otrzymali zadanie ślepej oceny poligramów różnymi metodami. Grupę probantów tworzyło 39 ochotników z grona studentów KAAF. Wśród nich było 13 mężczyzn i 26 kobiet w wieku od 20 do 43 lat. Na potrzeby eksperymentu zainscenizowano zdarzenie w postaci oddania trzech strzałów do sylwetki młodej kobiety na kolorowym plakacie w obiekcie strzelnicy uczelnianej. Dokonało tego 15 probantów i utworzyli oni grupę osób winnych, którzy następnie mieli nie przyznawać się do sprawstwa tego czynu podczas badań poligraficznych. W ramach dodatkowej motywacji otrzymali banknot 50 zł, który mogli zachować w sytuacji negatywnej identyfikacji przez badającego. „Niewinni” (24 osoby) w ogóle nie byli na strzelnicy i nie wiedzieli, co się tam wydarzyło.

¹³⁹ Decyzja NCN numer DEC-2013/11/B/H55/03856.

¹⁴⁰ J. Widacki, *Wartość diagnostyczna badania poligraficznego i jej znaczenie kryminalistyczne*, Kraków 1977.

Mieli odpowiadać zgodnie z prawdą w czasie badań. Również otrzymali gratyfikację finansową, ale mieli ją zwrócić w przypadku fałszywej identyfikacji pozytywnej (chodziło o zapewnienie, żeby trzymali się swojej roli w eksperymencie, a poza tym w realnych warunkach niewinnie podejrzany również ponosi konsekwencje ewentualnej nie-trafnej diagnozy biegłego)¹⁴¹.

Tab. 25. Dane dotyczące wartości diagnostycznej wyników testów ocenianych różnymi metodami w ramach eksperymentu (z wyłączeniem przypadków niekonkluzywnych)

Metoda analizy danych testowych	N (sprawca / niewinny świadek)	Czułość [CI]	Swoistość [CI]	PPV [CI]	NPV [CI]	Procent trafnych diagnoz winnych (n)	Procent trafnych diagnoz niewinnych (n)	Dokładność [Overall Fraction Correct]
oceny badających (ESS)	34 (12/22)	0.50 [0.26, 0.67]	0.86 [0.73, 0.96]	0.67 [0.34, 0.90]	0.76 [0.64, 0.85]	50% (6)	86,4% (19)	0.74 [0.56, 0.86]
metoda globalna – ślepa ocena	34 (12/22)	0.67 [0.40, 0.86]	0.77 [0.63, 0.88]	0.62 [0.37, 0.80]	0.81 [0.70, 0.92]	67% (8)	77% (17)	0.74 [0.55, 0.87]
ESS – ślepa ocena	27 (9/18)	0.78 [0.47, 0.92]	0.89 [0.73, 0.97]	0.78 [0.47, 0.94]	0.89 [0.73, 0.97]	78% (7)	89% (16)	0.85 [0.65, 0.96]
Utah – ślepa ocena	33 (11/22)	0.69 [0.46, 0.82]	0.90 [0.75, 0.98]	0.82 [0.54, 0.97]	0.82 [0.68, 0.89]	69% (9)	90% (18)	0.82 [0.63, 0.92]
wybrane algorytmy komputerowe w ramach OSS-3	35 (12/23)	0.50 [0.26, 0.64]	0.91 [0.79, 0.98]	0.75 [0.39, 0.95]	0.78 [0.67, 0.83]	50% (6)	91% (21)	0.77 [0.61, 0.86]

Źródło: opracowanie własne autora.

Poza testem demonstracyjnym (w postaci testu szczytowego napięcia) badający wykorzystywali jednakowy test o charakterze diagnostycznym, jednoproblemowym, w standaryzowanym formacie z pytaniami porównawczymi – *Utah zone comparison test*. Interpretacji danych testowych dokonywali samodzielnie, natomiast 12 innych

¹⁴¹ *Idem* (red.), *Kierunki rozwoju instrumentalnej i nieinstrumentalnej detekcji kłamstwa...*, op. cit., s. 65.

poligraferów – bez znajomości sprawy, na zasadzie ślepej oceny – analizowało poligramy w trzech podgrupach. Każda z nich wykorzystywała inną metodę analizy – systemy: ESS, Utah oraz metodę globalną. Obok manualnych ocen biegłych uwzględniono również rezultaty wynikające z zastosowania programu analitycznego OSS-3 (algorytmy skonstruowane na podstawie tzw. reguł Sentera i analizy prawdopodobieństwa Raskina). Rezultaty pracy poligraferów zostały poddane analizom statystycznym.

Uzyskane w toku eksperymentu wyniki wskazują, że pod względem ogólnej dokładności najlepsze rezultaty dała ślepa interpretacja danych testowych zgodna z regułami systemu ESS (0,85). Z drugiej strony – nie w ujęciu procentowym, lecz patrząc na liczbę trafnie zidentyfikowanych przypadków – metoda ta nie przewyższała, a wręcz ustępowała pozostałym w konsekwencji stosunkowo największej liczby wyników nierozstrzygniętych (zob. tab. 26).

Tab. 26. Udział wyników nierozstrzygniętych w eksperymencie

Metoda analizy danych testowych	Nierozstrzygnięte (spośród ogólnej liczby przypadków n = 39)	Udział procentowy
oceny badających (ESS)	5	13%
metoda globalna – ślepa ocena	5	13%
ESS – ślepa ocena	12	31%
Utah – ślepa ocena	6	15%
algorytmy komputerowe OSS-3	4	10%

Źródło: opracowanie własne autora.

Krytycy sposobu prezentacji danych dotyczących swoistości i czułości metod badawczych z wyłączeniem wyników nierozstrzygniętych podnoszą, że taki zabieg może sztucznie zawyżać te parametry. Jednakże dla celów praktycznych ma to uzasadnienie, ponieważ wyniki nierozstrzygnięte niczego nie wnoszą do sprawy – nie są interpretowane ani na korzyść, ani na niekorzyść osoby badanej. Ta grupa wyników rzutuje na użyteczność metody, a nie na jej dokładność. Dlatego, w przypadku badań poligraficznych, zwykle raportuje się je osobno. Patrząc na powyższe rezultaty, uwagę zwraca niemal trzykrotnie wyższy odsetek wyników nierozstrzygniętych w przypadku analizujących poligramy, z wykorzystaniem systemu ESS, metodą ślepej interpretacji

(31%) – w porównaniu do analizujących tym samym systemem poligraferów, którzy osobiście przeprowadzili badania (13%), a także wobec ustalonej metaanalizy średniej liczby takich wyników przy przeprowadzaniu testów typu ZCT (9,8%)¹⁴². Świadczy to o tym, że albo „ślepo oceniający” byli przesadnie ostrożni podczas swoich ewaluacji, albo – to oni poprawniej analizowali poligramy uzyskane w toku eksperymentu, a badający byli bardziej zmotywowani do podejmowania rozstrzygających decyzji, w niektórych przypadkach tendencyjnie (świadomie lub nie) naciągając wyniki. W świetle obliczeń przedstawionych w tab. 25 można bardziej skłonić się ku tej drugiej hipotezie, ponieważ eksperci dokonujący ślepej interpretacji wydali procentowo więcej dokładnych opinii. Ich zadaniem była wyłącznie analiza poligramów. Nie znali tematyki badań. Z kolei badający próbowali przyporządkować konkretne osoby do roli, którą odegrały w eksperymencie – winnego lub niewinnego.

W przypadku badań poligraficznych najważniejsza wydaje się minimalizacja nietrafnych identyfikacji – nawet jeśli miałyby to oznaczać, że nieco więcej badań okazywałoby się nieprzydatnych z powodu braku jednoznacznych wskazań. Testy ślepo interpretowane systemem ESS charakteryzowały się też najwyższą czułością (0,78) i wartością predykcyjną negatywną (czyli prawdopodobieństwem, że osoba badana nie jest sprawcą, gdy tak wskazuje wynik testu; 0,89). Z kolei najlepszą swoistość osiągnięto przy zastosowaniu algorytmów komputerowych (0,91), a wartość predykcyjną dodatnią (prawdopodobieństwo, że osoba badana jest winna, gdy tak wskazuje wynik testu; 0,82) – przy ślepej interpretacji systemem Utah.

Zgodnie z przyjętą hipotezą eksperci dokonujący ślepej interpretacji metodami numerycznymi (ESS i Utah) byli dokładniejsi niż ci, którzy sami przeprowadzali testy (odpowiednio: 0,85 i 0,82 wobec 0,74). W każdym przypadku okazało się to nieco poniżej średniej dokładności testu Utah raportowanej przez APA na podstawie metaanalizy wyników rozmaitych opublikowanych badań (92,1–93%)¹⁴³ – co należy łączyć raczej ze specyfiką tej konkretnej sytuacji eksperymentalnej, ponieważ kwalifikacje zaangażowanych poligraferów nie odbiegały od światowych standardów. Powody, dla których badający wykazali się mniejszą skutecznością w trafnym diagnozowaniu niż ślepo

¹⁴² *Ibidem*.

¹⁴³ Zob. M. Gougler *et al.*, *Meta-Analytic Survey...*, *op. cit.*

oceniający, mogły być rozmaite. Ci pierwsi działali pod presją czasu i oczekiwania kategorycznej decyzji. Niewykluczone, że sugerowali się czynnikami subiektywnymi, wynikającymi z bezpośredniej interakcji z osobami poddanymi testom. Analizujący poligramy na ślepo byli od tego wszystkiego oddzieleni.

Porównując poszczególne metody analizy danych z badań poligraficznych na podstawie kryterium dokładności, nie zaobserwowano istotnych różnic między nimi – a przynajmniej na tyle, by jednoznacznie wykazać, że którakolwiek góruje nad pozostałymi. Potwierdzono, że ewaluacja zapisów na poligramach różnymi uznanymi metodami numerycznymi prowadzi do diagnoz o podobnym poziomie dokładności, ale nie dowiedziono przy tym hipotezy, że metody numerycznej interpretacji co do zasady pozwalają na osiągnięcie większej liczby dokładnych rezultatów testów niż metoda globalna (jakościowa). Interesujące w tym kontekście byłoby sprawdzenie w przyszłości, czy sytuacja nie zmieniałaby się w przypadku porównania diagnoz poligraferów o zróżnicowanym doświadczeniu zawodowym (włącznie z takimi, którzy legitymują się niewielkim stażem praktyki).

Ogółem trafniej identyfikowano osoby niewinne (tak jak np. we wspomnianym wcześniej eksperymencie F. Horvatha i J. Reida z lat 70.). Statystycznie zbadano zależność między dokładnością testu przy różnych metodach interpretacji danych a faktyczną szczerością osób badanych. Dla jednej z metod – algorytmów komputerowych – zaobserwowano tzw. *truth bias*. W tym przypadku różnica między trafną identyfikacją winnych i niewinnych okazała się istotna statystycznie: $\chi^2 \text{ Walda} = 4.90$, $\beta = -1.66$, $p = .027$. Z kolei dla ocen poligramów dokonywanych przez poligraferów, którzy przeprowadzali badania, zaobserwowano efekt na granicy istotności statystycznej: $\chi^2 \text{ Walda} = 3.12$, $\beta = -1.05$, $p = .077$ – co może sugerować, że w ten sposób również trafniej identyfikowano niewinnych niż winnych (zob. tab. 27).

Jeżeli zaś chodzi o analizy statystyczne najbardziej interesujące z punktu widzenia celów niniejszej pracy – posłużono się współczynnikiem alfa Krippendorpha, który miał określić poziom zgodności między oceniającymi poligramy (zob. tab. 28).

Tab. 27. Współczynniki beta, odchylenia standardowe, przedziały ufności oraz istotności dla zależności pomiędzy dokładnością testów przy różnych metodach analizy poligramów a szczerością osób badanych

Metoda analizy danych testowych	β	SD	CI		Chi-kwadrat Walda	istotność
			dolna granica	górną granica		
ESS – ślepa ocena	-0,13	0,82	-1,74	1,47	0,03	0,870
algorytmy komputerowe OSS-3	-1,66	0,75	-3,13	-0,19	4,90	0,027
metoda globalna – ślepa ocena	-0,06	0,59	-1,22	1,10	0,01	0,918
oceny badających – ESS	-1,05	0,59	-2,21	0,11	3,12	0,077

Źródło: opracowanie własne autora.

Tab. 28. Współczynniki zgodności (rzetelności) między poligraferami, którzy na potrzeby eksperymentu posługiwali się różnymi metodami interpretacji danych testowych

Metoda ewaluacji poligramów	alfa Krippendorpha [CI]
ESS	.57 [.47, .66]
Utah	.43 [.33, .54]
globalna / jakościowa	.43 [.32, .53]
algorytmy OSS-3	.82 [.58, 1.00]

Źródło: J. Widacki, M. Gołaszewski, *Subiektywizm w badaniach poligraficznych*, „Studia Prawnicze. Rozprawy i Materiały”, 2(21), 2017, s. 210.

Najwyższy współczynnik zgodności odnotowano dla algorytmów komputerowych w ramach oprogramowania OSS-3. Nie jest to zaskakujące – zważywszy na to, że jedyne różnice powinny wynikać z reguł algorytmu, a czynnik subiektywny tu nie istnieje. Spośród metod manualnych największa zgodność występowała przy systemie analizy numerycznej ESS (skala 3-pozycyjna).

Większa rozbieżność diagnoz zaistniała przy skali 7-pozycyjnej systemu Utah, jak również przy metodzie globalnej analizy zapisów na poligramach. W przypadku tych dwóch ostatnich metod współczynnik alfa Krippendorpha okazał się jednakowy. Tym samym nie potwierdziła się w pełni hipoteza o tym, że metody numeryczne z zasady pozostawiają poligraferom mniejszy margines subiektywizmu niż

metoda jakościowa. Hipoteza jest trafna w odniesieniu do Empirycznego Systemu Oceniania. Natomiast kiedy pojawiają się szersze możliwości punktowania ocen zmian reakcji fizjologicznych przy pytaniach krytycznych – rzetelność testu spada. Tym samym pozytywnie zweryfikowano hipotezę o bardziej obiektywnych metodach numerycznych ze skalą 3-pozycyjną wobec metod ze skalą 7-pozycyjną.

Porównano ponadto zgodność ocen pomiędzy poszczególnymi systemami analizy danych testowych (zob. tab. 29). Najbardziej zbliżone do siebie rezultaty dawały testy oceniane przez badających systemem ESS i te, które były analizowane na ślepo z wykorzystaniem systemu Utah. Najśłabsza korelacja była również między tymi samymi systemami, ale wtedy gdy ESS posłużyli się ślepo oceniający (niewątpliwie przyczyniła się do tego sygnalizowana wcześniej zwiększona liczba wyników nierozstrzygniętych).

Tab. 29. Współczynniki rzetelności *Krippendorph's alpha* dla całościowych ocen poligraferów (z włączeniem wyników nierozstrzygniętych) – zestawienie różnych metod analizy danych

Metoda analizy danych testowych	Oceny badających (ESS) [CI]	Algorytmy komputerowe OSS-3 [CI]	ESS – ślepa ocena [CI]	Utah – ślepa ocena [CI]
metoda globalna – ślepa ocena	0.59 [0.36, 0.82]	0.53 [0.24, 0.76]	0.35 [0.10, 0.59]	0.56 [0.35, 0.78]
oceny badających (ESS)	-	0.54 [0.28, 0.80]	0.41 [0.15, 0.66]	0.63 [0.40, 0.86]
algorytmy komputerowe w ramach OSS-3	-	-	0.47 [0.21, 0.69]	0.53 [0.28, 0.76]
ESS – ślepa ocena	-	-	-	0.30 [0.06, 0.58]

Źródło: opracowanie własne autora.

W raporcie przygotowanym w 2001 r. dla Instytutu Badań Poligraficznych Departamentu Obrony USA porównano dokładność i zgodność diagnoz w badaniach poligraficznych z badaniami psychologicznymi (DSM III i IV) i medycznymi (badania rentgenem, ultrasonograficzne, tomograficzne i rezonansem magnetycznym). Wyrażona współczynnikiem Kappa zgodność ocen poligraferów wynosiła 0,79. W przypadku psychologów było to 0,77, a dla diagnoz medycznych 0,56¹⁴⁴.

¹⁴⁴ P.E. Crewson, *A Comparative Analysis of Polygraph with Other Screening and Diagnostic Tools*, „Polygraph”, 32(2), 2003, s. 57–85.

Podsumowując powyższe eksperymenty, można powiedzieć, że rzetelność testów wykorzystywanych w badaniach poligraficznych jest stosunkowo wysoka i na pewno nie gorsza niż w innych powszechnie uznawanych badaniach – w tym psychologicznych czy wielu nawet zaawansowanych technologicznie badaniach medycznych, ale jednocześnie widać, że przestrzeń dla rozbieżności niewątpliwie istnieje. Dlatego tak istotne jest podejmowanie wszelkich możliwych kroków, aby ta przestrzeń była jak najmniejsza.

III. Sposoby ograniczenia marginesu subiektywizmu w badaniach poligraficznych

1. Walidacja i standaryzacja testów, ujednoczenie procedur badań poligraficznych

1.1. Trafność i rzetelność testów

Poprawność diagnoz w ocenie wiarygodności przy użyciu poligrafu zależy m.in. od tego, czy metoda badań i konstrukcja przeprowadzanych testów dostarczają takich danych fizjologicznych, co do których możliwe jest trafne przypisanie ich przyczyn. Jeśli metoda nie ma naukowego potwierdzenia, nie można istotnie polegać na uzyskiwanych za jej pomocą rezultatach. Gdy reguły postępowania przy danej metodzie nie są dostatecznie precyzyjne i jednolite, mogą być subiektywnie interpretowane i stosowane w sposób prowadzący do niepotrzebnych rozbieżności. Dlatego, poza postępowaniem w technologii badań poligraficznych, coraz większą wagę (szczególnie od czasu opublikowania raportu amerykańskiej National Research Council w 2003 r.¹) przykładana się do walidacji metod wykorzystywanych w badaniach poligraficznych. Walidacja metody pomiarowej – według normy PN-EN ISO/IEC 17025:2005² – jest potwierdzeniem przez zbadanie i przedstawienie obiektywnego dowodu, że zostały spełnione wymagania dotyczące zamierzonego zastosowania. W tym kontekście szerszego omówienia wymagają, zasygnalizowane już w rozdziale 1 tej części monografii, takie pojęcia jak trafność (w tym fasadowa,

¹ National Research Council – National Academies Press, *The polygraph and lie detection*, op. cit.

² Polski Komitet Normalizacyjny, *Ogólne wymagania dotyczące kompetencji laboratoriów badawczych i wzorcujących*, Warszawa 2005.

konstrukcji, kryterialna, wewnętrzna i zewnętrzna) oraz rzetelność (wewnętrzna i zewnętrzna)³.

Trafność fasadowa opisuje, w jaki sposób dany test, wskaźnik czy narzędzie pomiarowe są subiektywnie postrzegane – przez badających (w jakim stopniu test mierzy to, co zakłada się, że powinien mierzyć), a w szczególności przez osoby badane (co im wydaje się, że jest przez test mierzone; czy jest to adekwatne do celu tego testu). Przekonanie przedstawicieli środowiska naukowego oraz praktykujących biegłych (np. panelu czy zrzeszenia zawodowego) w kwestii efektu zastosowania określonego testu i poprawności metodologii badań poligraficznych rzutuje na stopień ich akceptacji, upowszechnienie i znaczenie dowodowe. Prawdopodobnie u większości ludzi wiedza o badaniach poligraficznych sprowadza się do potocznego wyobrażenia poligrafu jako wprost – wykrywacza kłamstw. Część (włącznie z przedstawicielami zawodów prawniczych) utożsamia cele tych badań z ustalaniem związku emocjonalnego z jakimś zdarzeniem albo pomiarami poziomu stresu przy pytaniach testowych. Tymczasem wzór kłamstwa nie istnieje, a poligraf nie jest miernikiem emocji czy stresu. Poligraf rejestruje mimowolne zmiany reakcji fizjologicznych osoby badanej, które pozwalają na ocenę przypisywanego przez nią znaczenia określonym bodźcom testowym względem pozostałych. Dopiero pośrednio, na podstawie statystycznych danych referencyjnych, zaobserwowany sposób reagowania można zakwalifikować do zbioru przypadków typowych, czyli tych najczęściej występujących w populacji osób odpowiadających szczerze albo nieszczerze.

Jeżeli osoby badane nie będą przekonane co do skuteczności metody, ich motywacja najpewniej obniży się, co może rzutować na końcowy wynik. Niektórzy stawiają znak równości między poligrafem a szklaną kulą, podnoszą rzekomą łatwość w *oszukiwaniu urzędnika*. Stąd m.in. wkomponowanie w procedurę badania poligraficznego testu demonstracyjnego (stymulacyjnego), którego jednym z podstawowych celów jest upewnienie osoby badanej w tym, że *poligraf działa*.

Trafność teoretyczna (zamiennie – *konstrukcyjna*) odnosi się do tego, czy test mierzy określoną cechę tak, jak został zaprojektowany; czy rzeczywiście jest miarą pożądanego atrybutu, czy właściwości. Badanie poligraficzne jest instrumentalną metodą detekcji nieszczerości. Nieszczerłość powinna więc być odpowiednio zdefiniowana

³ Por. J.A. Matte, *Forensic Psychophysiology...*, *op. cit.*, s. 104–107.

i odróżniona od innych pojęć, np. od winy, a test musi być osadzony w ramach jakiejś teorii. Trafność konstrukcyjną weryfikuje się poprzez sprawdzanie hipotez, m.in. o zachowaniu się wytypowanych grup (np. szczerych i nieszczerych), uzyskujących różne wyniki testu. Sprawdza się, jakie zależności występują między zmiennymi i jaka jest korelacja testu z innymi, które mierzą tak podobne, jak i odmienne właściwości.

W badaniach poligraficznych trafność konstrukcyjna jest rozpatrywana, jeżeli chodzi o strukturę (rodzaje bodźców – pytań i ich sekwencję) i sposób przeprowadzania testu oraz metodę analizy danych testowych. W każdym z tych aspektów można dostrzec pewne różnice. Na przykład wprowadzone przez Backstera tzw. pytania symptomatyczne obecnie przedstawia się jako bez znaczenia dla dokładności wyników. Sposoby przeprowadzania testów też nie są jednolite – np. w kwestii omawiania pytań między poszczególnymi seriami testu (wykresami), rotacji pytań w kolejnych seriach, momentu przeprowadzenia testu stymulacyjnego (na początku lub po pierwszym wykresie testu właściwego). Z kolei trzeci ze wskazanych aspektów trafności teoretycznej – analiza danych – został szerzej przedstawiony w niniejszej pracy. Najważniejsze wydaje się przy tym obranie takich metod, które będą oparte na czytelnych, możliwie jednolitych i popartych dowodami naukowymi cechach diagnostycznych zapisów na poligramach oraz regułach oceniania i podejmowania decyzji.

Trafność kryterialna określa, w jakim stopniu wynik testu jest powiązany ze zmienną pozatestową (której test nie mierzy bezpośrednio), na temat której chcemy wnioskować, opierając się na tym wyniku. Kryterium szczerości/nieszczerości (czasem, nie do końca adekwatnie, nazywane również kryterium winy) – jest zatem czymś, co ustala się na podstawie pomiarów zmian reakcji fizjologicznych osoby badanej. O trafności kryterialnej świadczy współczynnik korelacji między wynikiem testu a faktyczną wiarygodnością osoby badanej. Właściwie możemy tu postawić znak równości między trafnością kryterialną a dokładnością wskazania testu (trafnością diagnostyczną, dotyczącą cechy aktualnie występującej u osoby badanej, czyli odpowiadającej na pytania krytyczne szczerze lub nieszczерze, rozpoznającej mimo zaprzeczania lub nierozpoznającej ulokowanego w teście bodźca krytycznego – szczegółu związanego z danym zdarzeniem). Omawiany rodzaj trafności w przypadku testów z wykorzystaniem poligrafu jest bardzo dobrze zbadany, nawet lepiej niż dla wielu innych powszechnie uznanych metod badań kryminalistycznych.

Trafność wewnętrzna pozwala na ustalenie, na ile uzyskane wyniki testu są efektem wpływu zmiennej niezależnej na zmienną zależną, a nie jakichś innych czynników pozostających poza kontrolą badającego. Im test jest bardziej trafny wewnętrznie, tym istnieje mniej alternatywnych możliwości wyjaśnienia przyjętej hipotezy badawczej czy wyniku testu. W eksperymentach laboratoryjnych trafność wewnętrzną można weryfikować i wzmacniać poprzez zaangażowanie grup kontrolnych. W ten sposób obserwuje się, czy dana zmienna (np. typ pytania, format testu, liczba powtórzeń, sposób omawiania pytań, metoda analizy danych) rzeczywiście istotnie wpływa na wynik testu i ewentualnie w jaki sposób, czy też nie wywiera żadnego efektu. Na pewno jest to zadanie trudniejsze przy badaniach terenowych, gdzie nie ma doskonałej metody ustalenia stanu faktycznego (przyznanie się do winy, wyrok sądu czy ocena innych dowodów zgromadzonych w sprawie pozwalają na przyjęcie takiego założenia w kwestii prawdopodobności osoby badanej, które graniczy z pewnością, ale absolutnej pewności, czyli pełnej kontroli nad warunkami badawczymi nie dają).

Z kolei *trafność zewnętrzna* opisuje możliwość generalizacji wyników badań w danej grupie na szerszą populację; na ile te wyniki będą miały zastosowanie wobec innych osób (np. z odmiennego kręgu kulturowego) albo w innych warunkach (np. wyniki z badań dotyczących zainscenizowanych zdarzeń w rzeczywistych sprawach karnych). Wśród sposobów na potwierdzenie tego rodzaju trafności znajdują się przede wszystkim badania replikacyjne. Takowe były wielokrotnie prowadzone dla przedmiotowej specjalności, a za minimum uznaje się przynajmniej dwie niezależne publikacje z wynikami badań, poddane niezależnej recenzji. APA przyjęło owe kryterium dla uznania danej techniki badawczej za potwierdzoną naukowo i dopuszczoną do innych badań niż tylko eksperymentalne (dowodowe, w celach śledczych i przesiewowych)⁴. Trafność zewnętrzna jest zazwyczaj wyższa, gdy warunki eksperymentalne są możliwie bliskie realnej sytuacji życiowej, a próba badawcza liczna i reprezentatywna. Przyznać trzeba, że większość badań empirycznych prowadzono zazwyczaj z różnych przyczyn na małych próbach. Na przykład w Polsce – poważną przeszkodą była i jest niewielka populacja poligraferów, hermetyczne środowisko funkcjonujące w większości w ramach służb policyjnych i specjalnych (czasem konkurencyjnie nastawionych). Badania poligraficzne (poza pewnymi

⁴ Zob. M. Gougler *et al.*, *Meta-Analytic Survey...*, *op. cit.*

wyjątkami – jak Kraków czy Katowice) nie cieszyły też specjalnym zainteresowaniem ośrodków akademickich⁵.

Kolejną istotną właściwością oceny testu jest jego rzetelność, powtarzalność pomiarów – wskazująca, że test konsekwentnie mierzy tożsame właściwości. Osoba ponownie poddawana badaniu poligraficznemu w tej samej sprawie powinna być zdiagnozowana w poprawny, jednaki sposób, a dany zbiór danych testowych analizowany przez różnych badających powinien prowadzić do zbieżnych konkluzji. Z jednej strony mamy więc rzetelność, rozpatrywaną na przestrzeni czasu – czyli jaka będzie korelacja ocen między kolejnymi testami przeprowadzonymi w innych terminach. Inną miarą rzetelności jest spójność wewnętrzna, czyli stopień zgodności wyników dla różnych fragmentów testu. Zatem, jeżeli wykonuje się np. test jednoprotokółowy z dwoma lub trzema pytaniami krytycznymi – to sposób reagowania osoby badanej nie powinien być rozbieżny, ponieważ wszystkie pytania krytyczne dotyczą tej samej sprawy i pozostają względem siebie zależne. Stąd właśnie to suma całkowita ocen numerycznych przesądza wówczas o końcowym wyniku. Wreszcie korelacja wyników różnych badających dysponujących tymi samymi danymi wskazuje na ich stopień zgodności między sobą. Czynniki subiektywne obniżają ten wskaźnik, a zakres subiektywizmu powinna zmniejszać standaryzacja, czyli ujednolicanie procedur badawczych, ustalenie odpowiednich norm w oparciu o możliwie największą liczbę danych z badań reprezentatywnych jednostek. Poza publikacjami naukowymi, dokumentującymi dane empiryczne, istotnym źródłem poszukiwania norm są wytyczne zrzeszeń poligraferów, w szczególności APA.

1.2. Walidacja metod badań poligraficznych

Trafność i rzetelność metod wykorzystywanych w badaniach poligraficznych potwierdzają badania walidacyjne. Na ogół tych badań

⁵ Jeżeli chodzi o inne niż Kraków czy Katowice ośrodki zajmujące się badaniami poligraficznymi, to oczywiście znane są prace prof. M. Kulickiego z Torunia, czy prof. R. Jaworskiego z Wrocławia. Choć publikacje tych autorów w gruncie rzeczy albo zawierały informacje podręcznikowe o badaniach poligraficznych, albo stanowiły opis praktyki badań wykonywanych na użytek organów ścigania – to były jak najbardziej pożyteczne. Bardziej należy wyrazić obawę o to, że ten dorobek nie jest kontynuowany. Wspominając o dwóch jedynie ośrodkach akademickich, przemawia przez autora niniejszej publikacji troska o przyszłość zaplecza naukowego, bez którego praktyka badań poligraficznych w Polsce wiele by straciła.

składają się wszystkie poprawne pod względem naukowym historycznie przeprowadzone badania empiryczne – omówione szeroko w literaturze przedmiotu, w tym także częściowo w niniejszej pracy i innych publikacjach autora⁶. Były to zarówno badania laboratoryjne, terenowe, jak i mieszane. Prowadzono je na populacjach z różnych kręgów kulturowych. Nie sposób przy tym pominąć jednego z przełomowych badań, zrealizowanych na Uniwersytecie Jagiellońskim w latach 70. XX wieku, które prowadziło do wniosku, że badania poligraficzne charakteryzują się nie mniejszą wartością diagnostyczną i użytecznością niż inne, powszechnie uznane, metody identyfikacji (zeznanie świadka naoczego, badania daktyloskopijne, badania pisma ręcznego). Wyniki z tego eksperymentu były raportowane w 1983 r. w amerykańskim Kongresie⁷.

Zważywszy na wielość różnych badań empirycznych, przy stosunkowo małych próbach badawczych, zasadne było stosowanie metaanalizy danych. Doniosłe znaczenie miały wyniki opublikowane w raporcie National Research Council w 2003 r. Oszacowano w nim m.in. dokładność testów jednowątkowych w granicach 0,81–0,91 (na podstawie średnich 26 wartości z 52 analizowanych zbiorów danych)⁸. Sam raport wskazywał jednak również, że naukowe podstawy badań poligraficznych są dalekie od tego, czego oczekiwaloby się od testu, który ma istotną wagę w podejmowaniu decyzji dotyczących bezpieczeństwa narodowego. Stanowił więc asumpt do tego, by środowisko poligraferów podjęło wysiłki zmierzające do udowodnienia tezy przeciwnej. APA powołało specjalny komitet, który miał zająć się

⁶ Por. m.in. M. Gołaszewski, *Przesłuchanie biegłego z zakresu badań poligraficznych (wariograficznych) w postępowaniu jurysdykcyjnym – tryb, najczęściej zadawane pytania i prawidłowe odpowiedzi*, [w:] K. Banasik, A. Kargol, A. Kubiak-Cyruł, M. Lubelski, E. Plebanek, A. Strzelec (red.), *W poszukiwaniu prawdy. Rozważania o prawie, historii i sprawiedliwości*, Kraków 2018, s. 269–298; *idem*, *Validated Techniques and Scoring Models...*, *op. cit.*; *idem*, *Współczesne standardy...*, *op. cit.*, s. 21–25.

⁷ Zob. J. Widacki, F. Horvath, *An Experimental Investigation of the Relative Validity and Utility of the Polygraph Technique and Three Other Common Methods of Criminal Identification*, „Journal of Forensic Sciences”, 23(3), 1978; Congress of the United States: Office of Technology Assessment, *Scientific Validity of Polygraph Testing. A Research Review and Evaluation – a Technical Memorandum*, Washington 1983, s. 69.

⁸ National Research Council – National Academies Press, *The Polygraph and Lie Detection*, Washington 2003, s. 148.

przeprowadzeniem własnej solidnej metaanalizy, służącej weryfikacji dotychczas praktykowanych technik badawczych. Ta metaanaliza objęła ogromną ilość danych – 12 665 rezultatów rozmaitych testów, wykonanych w trakcie ponad czterech tysięcy badań poligraficznych. W ten sposób sporządzono wykaz uznanych technik, rekomendowanych do różnych typów badań i ustalono:

- średnią dokładność wszystkich typów testów poligraficznych na poziomie 87,1% przy średniej ilości wyników nierozstrzygniętych (INC): 12,7%;
- średnią dokładność testów jednoprotymowych: 92,1%, przy wynikach INC: 8,8%,
- średnią dokładność testów wieloprotymowych: 85%, przy wynikach INC: 12,5%.

Tab. 30. Lista technik dopuszczonych do badań poligraficznych zgodnie ze standardem APA na podstawie rezultatów metaanalizy opublikowanej w 2012 r.⁹

Techniki dowodowe / metoda analizy danych	Techniki badań wieloprotymowych, konfrontacyjnych ¹⁾ / metoda analizy danych	Techniki wykrywcze (do celów dochodzeniowo-śledczych, eliminacyjnych) / metoda analizy danych
US Federal You-Phase / ESS • średnia dokładność: 90,4% • nierozstrzygnięte (INC): 19,2% • czułość: 84,5% • swoistość: 75,7%	USAF MGQT²⁾, LEPET, Utah MGQT³⁾ / ESS • średnia dokładność: 87,5% • nierozstrzygnięte (INC): 17% • czułość: 72,9% • swoistość: 70%	USAF MGQT, LEPET, Utah MGQT / skala 7-poz. • średnia dokładność: 81,7% • nierozstrzygnięte (INC): 19,7% • czułość: 78,3% • swoistość: 53,8%
ZCT (Federal, Utah) / ESS • dokładność: 92,1% • INC: 9,8% • czułość: 81,7% • swoistość: 84,6%	Federal You-Phase / skala 7-poz. • dokładność: 88,3% • INC: 16,8% • czułość: 84,5% • swoistość: 75,7%	CIT (GKT) / system Lykkena • dokładność: 82,3% • INC: 0,1% • czułość: 81,5% • swoistość: 83,2%

⁹ W pierwszej kolumnie umieszczono techniki spełniające kryteria: $\geq 90\%$ dokładności (trafności) i $\leq 20\%$ wyników nierozstrzygniętych. W drugiej kolumnie znalazły się techniki o przynajmniej 86% dokładności i dające najwyżej 20% wyników nierozstrzygniętych. Natomiast w trzeciej – techniki o co najmniej 80% dokładności i maksymalnie 20% wyników nierozstrzygniętych. Mając na uwadze opisane wymagania statystyczne dotyczące właściwości testu – te z pierwszej kolumny mogą być stosowane również w tych badaniach wyszczególnionych w drugiej kolumnie, zaś do badań określonych w trzeciej kolumnie można wykorzystywać testy wymienione we wszystkich kolumnach.

Techniki dowodowe / metoda analizy danych	Techniki badań wielopodmiotowych, konfrontacyjnych ¹⁾ / metoda analizy danych	Techniki wykrywcze (do celów dochodzeniowo-sledczych, eliminacyjnych) / metoda analizy danych
Utah ZCT (różne wersje ogółem) / Utah <ul style="list-style-type: none"> • dokładność: 93% • INC: 10,7% • czułość: 85,3% • swoistość: 80,9% 	Federal ZCT / skala 7-poz. <ul style="list-style-type: none"> • dokładność: 86% • INC: 17,1% • czułość: 85,8% • swoistość: 58,1% 	DLST (TES) / skala 7-poz. <ul style="list-style-type: none"> • dokładność: 84,4% • INC: 8,8% • czułość: 74,8% • swoistość: 79,2%
Utah ZCT DLC / Utah <ul style="list-style-type: none"> • dokładność: 90,2% • INC: 7,3% • czułość: 81,5% • swoistość: 85,7% 	Federal ZCT / skala 7-poz. dowodowa <ul style="list-style-type: none"> • dokładność: 88% • INC: 8,5% • czułość: 80,4% • swoistość: 80,9% 	DLST (TES) / ESS <ul style="list-style-type: none"> • dokładność: 85,8% • INC: 9% • czułość: 80,9% • swoistość: 75,1%
Utah ZCT PLC / Utah <ul style="list-style-type: none"> • dokładność: 93,1% • INC: 7,7% • czułość: 86,7% • swoistość: 83,3% 	Backster You-Phase / Backster <ul style="list-style-type: none"> • dokładność: 86,2% • INC: 19,6% • czułość: 83,6% • swoistość: 55,6% 	
Utah ZCT RCMP (v.1) / Utah <ul style="list-style-type: none"> • dokładność: 93,9% • INC: 18,5% • czułość: 83,3% • swoistość: 70% 		
*IZCT / HSS <ul style="list-style-type: none"> • dokładność: 99,4% • INC: 3,3% • czułość: 97,7% • swoistość: 94,6% 		
*MQTZCT / Matte <ul style="list-style-type: none"> • dokładność: 99,4% • INC: 2,9% • czułość: 96,7% • swoistość: 96,3% 		
* zastrzeżenia ze względu na tzw. statystyczne wartości oddalone i wątpliwości metodologiczne		

¹⁾ Z j. ang. – *paired testing* – „testy dobrane w pary”, tzw. protokół *Marin* – gdzie dwóch badających (wyjątkowo – jeden) bada co najmniej dwie osoby przedstawiające sprzeczne wersje zdarzenia w taki sposób, że jedna z osób z całą pewnością kłamie. ²⁾ Dane uśrednione dla obu wersji AFMGQT (v.1 i v.2). ³⁾ Testy LEPET i Utah MGQT nie były poddane odrębnej metaanalizie danych, natomiast uznano je za dopuszczalne, o ile będą zastosowane te same metody analizy danych testowych jak przy AFMGQT – ze względu na niemal jednakową strukturę i reguły postępowania.

Źródło: opracowanie własne na podst.: APA, *Meta-Analytic Survey of Criterion Accuracy of Validated Techniques*, „Polygraph”, 40(4), 2011, s. 93–305.

2. Doprecyzowanie kryteriów diagnostycznych

2.1. Naukowe potwierdzenie cech diagnostycznych zapisów na poligramach – rys historyczny

Dla ograniczenia marginesu subiektywizmu, rzutującego na trafność konstrukcyjną, a przede wszystkim na rzetelność testów poligraficznych, podstawowe znaczenie mają – poza omówionymi w części II monografii (rozdział 2) regułami przyznawania ocen numerycznych i podejmowania decyzji – odpowiedni dobór (opierając się na dowodach naukowych) i precyzyjny opis kryteriów diagnostycznych przy analizie zapisów na poligramach.

Wraz z rozwojem instrumentalnych metod detekcji nieuczciwości od początku XX wieku liczba cech uznawanych za diagnostyczne systematycznie zwiększała się, by następnie ulec zmniejszeniu – w imię uproszczenia zasad analizy danych fizjologicznych rejestrowanych przez poligraf, a także w wyniku odrzucenia tych cech, które nie znalazły dostatecznego potwierdzenia we współczesnych badaniach naukowych. Do połowy XX wieku wśród cech diagnostycznych u większości badaczy najczęściej powtarzały się – w przypadku kanału pneumograficznego: tłumienie oddychania i wzrost amplitudy oddechu; w kanale kardio: wzrost ciśnienia krwi, z bezpośrednio po nim następującym spadkiem. Natomiast w kanale GSR (do którego nie przywiązywano jeszcze szczególnej wagi) – wzrost amplitudy.

C. Lombroso – jeden z pionierskich twórców *aparatów do wykrywania kłamstw* (korzystał z wynalezionej przez siebie hydrofigmografu) zaobserwował jako symptomatyczny – spadek ciśnienia krwi i tętna po pytaniu krytycznym (1911 r.)¹⁰. W 1914 r. V. Benussi wskazywał, że z nieuczciwością skorelowany jest większy wskaźnik w postaci stosunku czasu wdychania powietrza do wydychania (I/E – z j. ang. *inhalation/exhalation*)¹¹.

W latach 20. J. Larson brał pod uwagę tłumienie oddychania oraz wzrost, a następnie spadek ciśnienia krwi, a dla obu krzywych zapisu:

¹⁰ C. Lombroso, *Crime. Its Causes and Remedies*, Boston 1911; przedruk fragmentu w: „Polygraph”, 20(1), 1991, s. 52–53.

¹¹ V. Benussi, *Die Atmung Asymptome der Lüge (The respiratory symptoms of lying)*, „Archiv für die gesamte Psychologie”, 11, 1914, s. 244–273; przedruk w: „Polygraph” 4(1), 1975, s. 52–76.

wzrost (podniesienie linii bazowej), zwiększenie częstotliwości, spowolnienie, nieregularne wahania¹². W latach 30. odnajdujemy u L. Keelera ponownie: wzrost i spadek ciśnienia oraz tłumienie oddychania¹³. U J. Wintera – zmiany w tempie oddychania (regularny/nieregularny, płytki/głęboki) i podobnie w kanale kardio: regularny i nieregularny zapis, średnie i silne odchylenia¹⁴. Z kolei zdaniem W. Marstona do wskazania na wprowadzenie w błąd uprawniały zmiany w oddychaniu towarzyszące zmianom w ciśnieniu krwi¹⁵.

Kompleksowy wykaz cech diagnostycznych, dla trzech kanałów zapisu (kardio, pneumio i GSR), przedstawił na początku lat 40. P. Trovillo. W kanale kardio wyszczególnił: wzrost ciśnienia krwi, po którym następuje albo powrót do linii bazowej, albo cykliczny wzrost na przestrzeni całego wykresu; lekki wzrost i gwałtowny spadek ciśnienia; zwężenie amplitudy pulsu, zmiany (zwłaszcza gwałtowne) w rytmie serca. W przypadku oddychania: tłumienie (w chwili nieszczerzej odpowiedzi lub tuż przed, po czym następuje głębszy oddech), blok, podniesienie linii bazowej, regularny sposób oddychania do momentu krytycznego bodźca i dalej odchylenia lub odwrotnie. W kanale GSR: stosunkowo duża lub rozległa reakcja, wyraźne różnice w przebiegu reakcji, stopniowy wzrost amplitudy, reakcje do momentu krytycznego bodźca i dalej zanik lub wyraźna zmiana w reakcjach jedynie w obrębie bodźca krytycznego (dwie ostatnie cechy typowe dla testów szczytowego napięcia). P. Trovillo charakteryzował ponadto reakcje niejednoznaczne, np. celowe głębokie wdechy, zniekształcenia zapisów spowodowane strachem czy ruchami ciała¹⁶.

F. Inbau wskazywał w swoich publikacjach najpierw takie cechy diagnostyczne jak: wzrost i spadek ciśnienia oraz tłumienie oddychania, po którym następuje westchnienie związane z ulgą¹⁷, a pod koniec lat 40.: wzrost ciśnienia krwi, ostry spadek, spowolnienie tętna,

¹² J.A. Larson, *The Cardio-pneumo-psychogram and Its Use in the Study of the Emotions, with Practical Applications*, „Journal of Experimental Psychology”, 5(5), 1923, s. 323–328; *idem*, *The Cardio-pneumo-psychogram in Deception*, „Journal of Experimental Psychology”, 6(6), 1923, s. 420–454.

¹³ L. Keeler, *A Method...*, *op. cit.*

¹⁴ J.E. Winter, *A Comparison...*, *op. cit.*

¹⁵ W.M. Marston, *The Lie Detector Test*, New York 1938; przedruk: APA, 1989.

¹⁶ P.V. Trovillo, *Deception Test Criteria. How One Can Determine Truth and Falsehood from Polygraphic Records*, „Journal of Criminal Law and Criminology”, 33(4), 1942, s. 338–358.

¹⁷ F.E. Inbau, *Lie Detection and Criminal Interrogation*, Baltimore 1942.

a także tłumienie oddechu i ciężki oddech przez ok. 20–25 sekund po odpowiedzi¹⁸. W podobnym okresie H. Baesen *et al.* wyszczególniali w kanale kardio: zmiany w tętnie, spadki ciśnienia, czas wzrostu i spadku ciśnienia, umiejscowienie szczyrby dykrotycznej. Z kolei w kanale pneumo: zmiany linii bazowej, blokowanie, tłumienie¹⁹. W 1950 r. R. Pierce, szef spółki L. Keelera, opisując spostrzeżenia z jednego z przeprowadzonych badań poligraficznych, wymienił takie przesłanki jak: uspokojenie i powrót do regularnego oddychania oraz spadek ciśnienia krwi po pytaniu krytycznym (reakcje toniczne), a ponadto wzrost amplitudy w kanale EDA – *pisak ostro wychylił się w górę* (reakcja fazowa)²⁰.

Od lat 60., czyli od momentu upowszechnienia dorobku C. Backstera, zestawy cech diagnostycznych stanowią integralną część kompletnych systemów analizy danych testowych. Zatem mówimy tu najpierw o systemie Backstera, jego modyfikacjach w postaci systemu stosowanego przez amerykańskie wojsko i później wszystkie instytucje federalne, systemie Utah oraz Empirycznym Systemie Oceniania. Reguły dotyczące numerycznej ewaluacji danych zostały przedstawione w poprzedniej części monografii. W tym miejscu natomiast omówione zostaną szczegółowo kryteria diagnostyczne. W ten sposób możemy przyjrzeć się, jaki jest stopień rozbudowania zbiorów cech, jakie różnice występują między systemami i na ile precyzyjnie określone są owe przesłanki diagnozowania. Przy tym – rzecz jasna – im bardziej skomplikowane i niedookreślone – tym z założenia większy powinien być margines subiektywizmu oceniających.

2.2. Przesłanki diagnozowania w systemie Backstera

W systemie Backstera wyróżniamy następujące fragmenty zapisu świadczące o zaistnieniu reakcji:

w kanale pneumo:

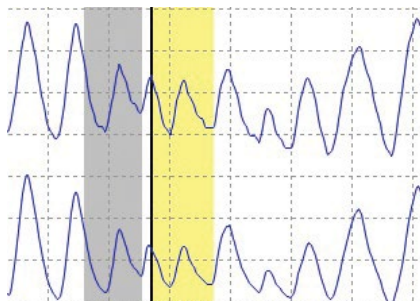
- tłumienie cyklu oddechowego, włącznie z blokiem, bezdechem (przykłady – ryc. 24 i 25),

¹⁸ *Idem, Lie Detection and Criminal Interrogation*, wyd. 2, Baltimore 1948.

¹⁹ H.V. Baesen, C. Chung, C. Yang, *A Lie Detector Experiment*, „Journal of Criminal Law, Criminology, and Police Science”, 39, 1948–1949, s. 532–537.

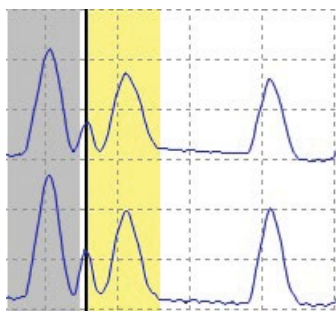
²⁰ R.W. Pierce, *The Peak of Tension Test*, „ISDD Bulletin”, 3(3), 1950, s. 1, 4–5.

Ryc. 24. Tłumienie oddechu



Źródło: materiały własne autora.

Ryc. 25. Blokowanie oddechu (bezdech)



Źródło: materiały własne autora.

– podniesienie linii bazowej,

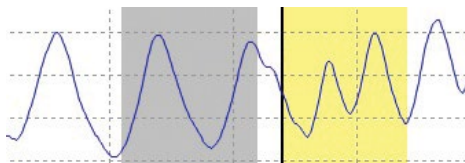
Ryc. 26. Czasowe podniesienie linii bazowej krzywych oddychania (w przykładzie również połączone ze zmniejszeniem amplitudy)



Źródło: materiały własne autora.

– wznoszące się tłumione cykle oddechowe,

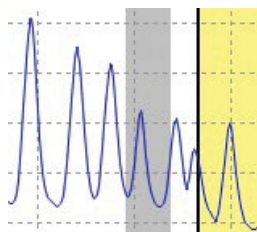
Ryc. 27. Wznoszące się tłumione cykle oddechowe



Źródło: materiały własne autora.

– opadające tłumione cykle oddechowe,

Ryc. 28. Opadające tłumione cykle oddechowe

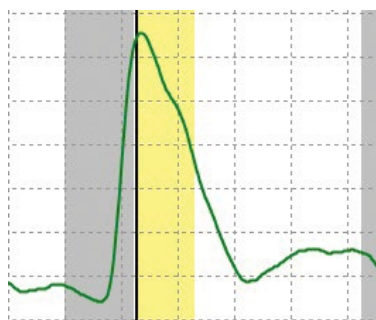


Źródło: materiały własne autora.

w kanale GSR/EDA:

– amplituda (odcinek od linii bazowej do szczytu reakcji),

Ryc. 29. Wzrost amplitudy krzywej GSR

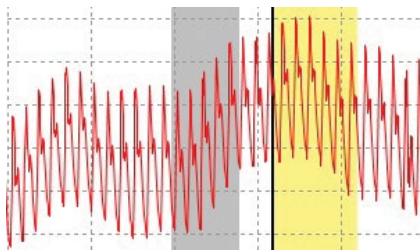


Źródło: materiały własne autora.

w kanale kardio:

– wzrost ciśnienia krwi, rozpatrywany od strony skurczowej,

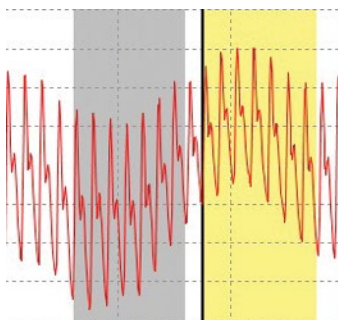
Ryc. 30. Wzrost ciśnienia krwi



Źródło: materiały własne autora.

– zmiany amplitudy pulsu (redukcja),

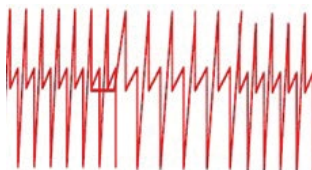
Ryc. 31. Redukcja amplitudy pulsu (w przykładzie również w połączeniu ze wzrostem ciśnienia)



Źródło: materiały własne autora.

– zmiany w tętnie,

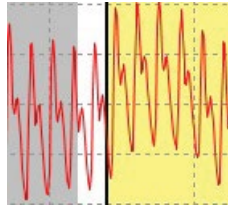
Ryc. 32. Spowolnienie tętna



Źródło: M. Gołaszewski (red.), *Współczesne standardy...*, op. cit., s. 33.

– zmiana pozycji szczyrby dykrotycznej (zwykle związana ze wzrostem ciśnienia),

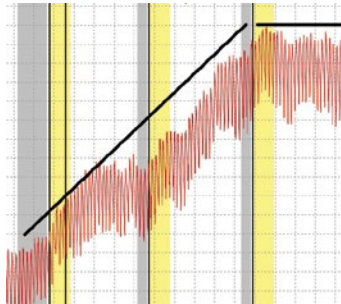
Ryc. 33. Zmiana pozycji szczyby dykrotycznej



Źródło: materiały własne autora.

- zmiana ogólnego trendu ciśnienia krwi na przestrzeni szerszego fragmentu wykresu,

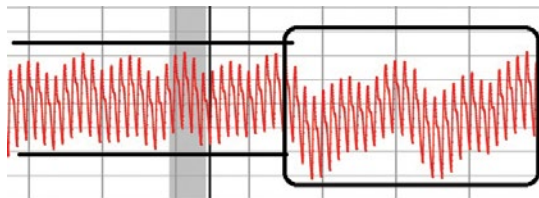
Ryc. 34. Przykład zmiany trendu krzywej kardio (systematyczny wzrost, potem stabilizacja)



Źródło: materiały własne autora.

- zmiana z ustabilizowanego cyklicznego trendu ciśnienia krwi (kryterium drugorzędne).

Ryc. 35. Przykład zmiany z ustabilizowanego trendu ciśnienia krwi (najpierw stabilna linia bazowa, a dalej zaczynają się większe fluktuacje)



Źródło: materiały własne autora.

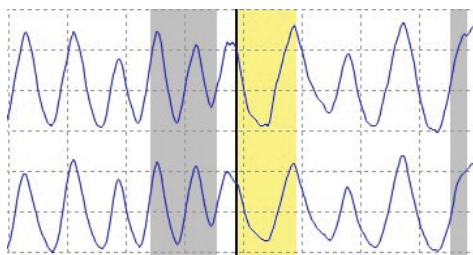
Poza powyższymi – dla kanałów pneumo i kardio obowiązywała jeszcze jedna z zasad pomocniczych, polegająca na stwierdzeniu reakcji poprzez dedukcję – kiedy fragment zapisu wskazujący na ulgę zaczyna się 5 sekund po udzieleniu odpowiedzi – co ma świadczyć o tym, że przed tym zaistniała najpewniej jakaś forma reakcji²¹.

2.3. Cechy diagnostyczne wyodrębnione przez szkołę USAMPS, system Rządu Federalnego USA

W systemie USAMPS mieliśmy z kolei takie kryteria:
w kanale pneumo:

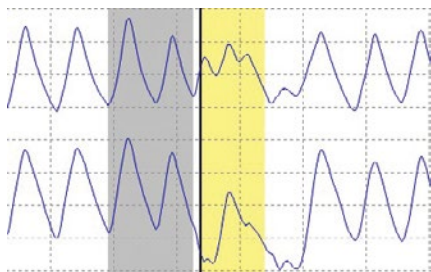
- zmiana rytmu lub regularności (przykłady – ryc. 36 i 37),

Ryc. 36. Zmiana rytmu oddychania



Źródło: materiały własne autora.

Ryc. 37. Zmiana (utrata) regularności w oddychaniu

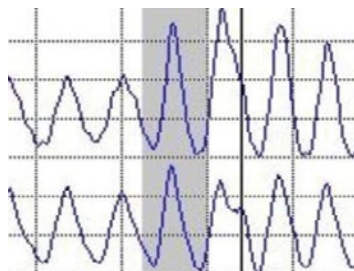


Źródło: materiały własne autora.

²¹ R.S. Weaver, *The Numerical Evaluation...*, op. cit., s. 100.

– zmiany w amplitudzie lub objętości,

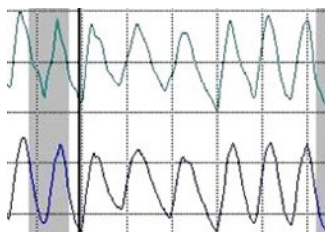
Ryc. 38. Zwiększenie amplitudy cyklu oddechowego



Źródło: materiały własne autora.

– zmiany w stosunkach wdechu do wydechu,

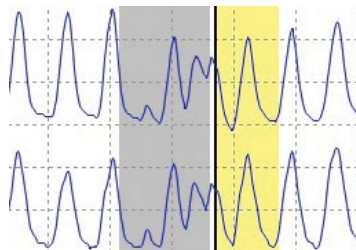
Ryc. 39. Zmiana w stosunkach wdechu do wydechu (I/E)



Źródło: materiały własne autora.

– ząbkowany zapis,

Ryc. 40. Ząbkowany zapis w kanale pneumo



Źródło: materiały własne autora.

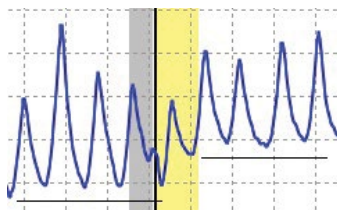
– zmiany linii bazowej,

Ryc. 41. Czasowe obniżenie linii bazowej krzywej oddychania



Źródło: Department of Defense Polygraph Institute, *Test Data Analysis. DoDPI Numerical Evaluation Scoring System*, 2004, s. 18.

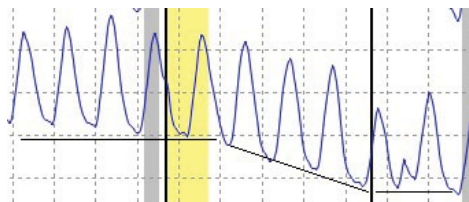
Ryc. 42. Zmiana (względnie trwałą) linii bazowej krzywej oddychania



Źródło: materiały własne autora.

– utrata linii bazowej,

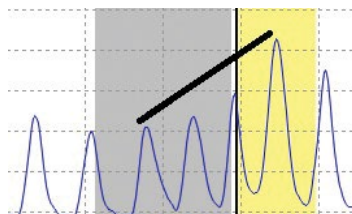
Ryc. 43. Utrata linii bazowej krzywej oddychania



Źródło: materiały własne autora.

– hiperwentylacja,

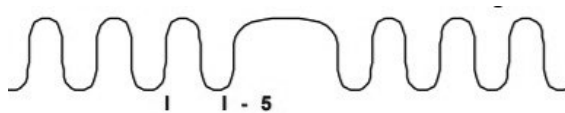
Ryc. 44. Hiperwentylacja (progressywne zwiększenie amplitudy cyklu oddechowego)



Źródło: materiały własne autora.

– wstrzymywanie lub blokowanie oddechu,

Ryc. 45. Wstrzymywanie oddechu (po wdechu)



Źródło: J. Swinford, *Manually Scoring Polygraph Charts...*, op. cit., s. 15.

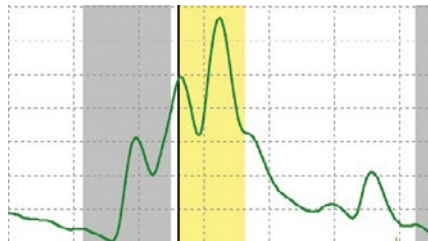
– tłumienie,

w kanale GSR/EDA:

– pionowy wzrost krzywej – amplituda,

– reakcja *podwójnego siodła*, złożona,

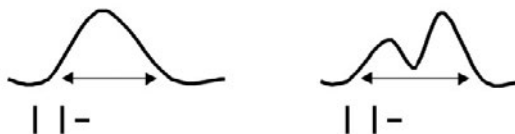
Ryc. 46. Reakcja złożona w kanale GSR (w przykładzie nawet *potrójne siodło*)



Źródło: materiały własne autora.

– długi czas trwania lub siła reakcji,

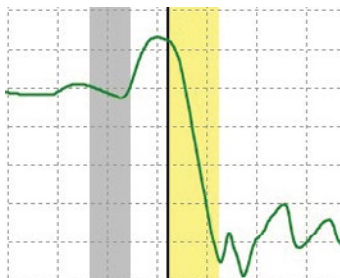
Ryc. 47. Porównanie czasu reakcji w kanale GSR



Źródło: Department of Defense Polygraph Institute, *Test Data Analysis...*, op. cit., s. 28.

– skaczący, gwałtownie opadający zapis krzywej,

Ryc. 48. Gwałtowny spadek krzywej GSR, a dalej fluktuacje

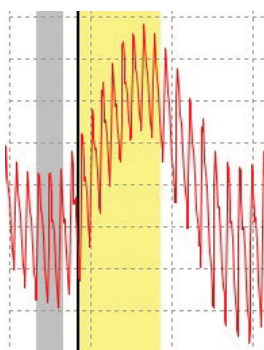


Źródło: materiały własne autora.

w kanale kardio:

– wzrost i następujący po nim spadek ciśnienia,

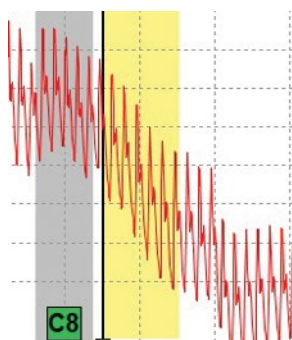
Ryc. 49. Wzrost i spadek ciśnienia



Źródło: materiały własne autora.

– tylko spadek ciśnienia,

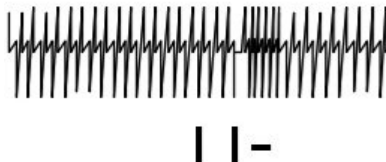
Ryc. 50. Spadek ciśnienia krwi



Źródło: materiały własne autora.

– wzrost tętna,

Ryc. 51. Wzrost tętna



Źródło: Department of Defense Polygraph Institute, *Test Data Analysis...*, op. cit., s. 36.

– wzrost amplitudy,

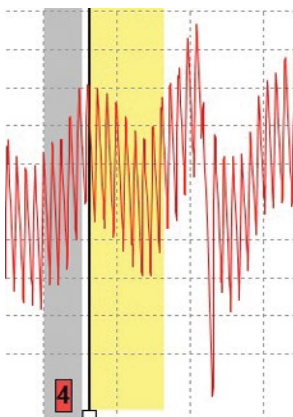
Ryc. 52. Wzrost amplitudy krzywej kardio



Źródło: Department of Defense Polygraph Institute, *Test Data Analysis...*, op. cit., s. 36.

– ekstrasystolie.

Ryc. 53. Ekstrasystolia



Źródło: materiały własne autora.

Ponadto w systemie USAMPS w kanale kardio jako diagnostyczne traktowano:

- spadek amplitudy,
- tylko wzrost ciśnienia,
- spadek tętna,
- zmiana pozycji lub zanik szczyrby dykrotycznej.

Dodatkowymi cechami, poza wykazem specyficznych reakcji na bodziec testowy, były:

- rozmieszczenie reakcji,
- siła reakcji,
- trend krzywej,
- stopień zmiany krzywej,
- okres latencyjny reakcji,
- czas trwania reakcji.

Łącznie powyższych cech diagnostycznych było zatem w systemie USAMPS aż 25. Liczba ta minimalnie spadła w kolejnych wersjach systemu federalnego – do 23 – i obowiązywała przynajmniej do 2004 r.²² W kanale pneumo nie ma stwierdzenia o *ząbkowanym* zapisie. Pozostałe cechy są właściwie tożsame, mogą być najwyżej nieco inaczej rozpisane – np. hiperwentylacja jako progresywny wzrost amplitudy, po którym następuje progresywne obniżenie lub powrót do homeostazy (ryc. 54 i 55). W kanale EDA nie ma mowy jedynie o gwałtownie opadającym zapisie. Z kolei w kanale kardio generalnie podaje się również te same cechy, nie wspomina się już o znaczeniu spadku ciśnienia i umiejscowienia szczyrby dykrotycznej. Inne cechy są bardziej szczegółowo wyodrębnione – np. podniesienie linii bazowej podzielono na gwałtowy wzrost jako wyraz aktywności fazowej (ryc. 56) oraz powolny wzrost wynikający z aktywności tonicznej (ryc. 57). Cechy pomocnicze sprowadzone są do czasu trwania reakcji, szybkości pobudzenia (wzrostu krzywej) oraz zaistnienia reakcji związanej ze współczulnym układem nerwowym w porównaniu do mniej wówczas istotnej aktywności układu przywspółczulnego (ryc. 58).

²² Department of Defense Polygraph Institute, *TEST DATA ANALYSIS. Do-DPI Numerical Evaluation Scoring System*, 2004.

Ryc. 54. Progresywne zwiększenie i zmniejszenie amplitudy krzywej oddychania



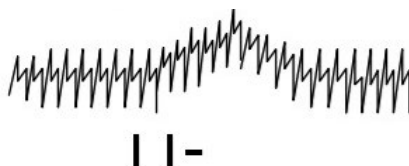
Źródło: Department of Defense Polygraph Institute, *Test Data Analysis...*, *op. cit.*, s. 17.

Ryc. 55. Progresywne zwiększenie amplitudy krzywej oddychania i powrót do homeostazy



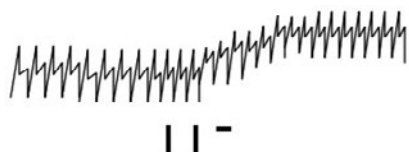
Źródło: Department of Defense Polygraph Institute, *Test Data Analysis...*, *op. cit.*, s. 17.

Ryc. 56. Reakcja fazowa w kanale kardio



Źródło: Department of Defense Polygraph Institute, *Test Data Analysis...*, *op. cit.*, s. 35.

Ryc. 57. Reakcja toniczna w kanale kardio



Źródło: Department of Defense Polygraph Institute, *Test Data Analysis...*, *op. cit.*, s. 35.

Ryc. 58. Reakcje spowodowane aktywnością współczulnego i przywspółczulnego układu nerwowego



Źródło: Department of Defense Polygraph Institute, *Test Data Analysis...*, *op. cit.*, s. 39.

W podręczniku dla amerykańskich instytucji federalnych z 2006 r. następuje istotna zmiana – wymienia się w nim już tylko 12 przesłanek diagnozowania, które pozostają aktualne do chwili obecnej (patrz tab. 31)²³.

Tab. 31. Aktualne kryteria diagnostyczne w systemie analizy danych testowych według Rządu Federalnego USA (tzw. system federalny)

Kanał	System federalny
Pneumo (ruchy klatki piersiowej odzwierciedlające przebieg cykli oddechowych)	<p>Czas rozpoczęcia reakcji: od wprowadzenia bodźca do jednego pełnego cyklu po odpowiedzi. Zakres reakcji: ≥ 3 cykle.</p> <ul style="list-style-type: none"> • tłumienie (zmniejszenie amplitudy); • bezdech (blok); • zmiana w stosunkach wdechu do wydechu; • stopniowe zmniejszenie amplitudy oddechu; • spowolnienie częstotliwości; • czasowa zmiana linii bazowej (cecha drugorzędna – w odróżnieniu od powyższych, bez związku z RLL – długością linii oddechu).
EDA (aktywność elektrodermalna – zmiany w przewodności lub oporności skóry)	<p>Czas rozpoczęcia reakcji: od wprowadzenia bodźca do odpowiedzi. Zakres reakcji: od rozpoczęcia reakcji do powrotu do linii bazowej poprzedzającej moment wprowadzenia bodźca.</p> <ul style="list-style-type: none"> • amplituda (cecha główna mierzona od linii bazowej do szczytu reakcji); • złożoność (jeśli krzywa nie powróci do linii bazowej, lecz nastąpi kolejne pobudzenie fizjologiczne); • czas trwania (czas między rozpoczęciem reakcji a powrotem do linii bazowej); • dwie ostatnie cechy bierze się pod uwagę jedynie przy podobnej amplitudzie porównywanych reakcji.
Kardio (względne ciśnienie krwi i tętno)	<p>Czas rozpoczęcia reakcji: od wprowadzenia bodźca do zakończenia odpowiedzi. Zakres reakcji: od rozpoczęcia reakcji do powrotu do linii bazowej (w części rozkurczowej).</p> <ul style="list-style-type: none"> • podniesienie linii bazowej (główna cecha); • spowolnienie tętna (jeśli nie zaistniała cecha główna); • czas trwania reakcji (pomocniczo – przy równych poziomach podniesienia linii bazowej).
PPG (reakcje naczynioruchowe; zmiany amplitudy pulsu)	<ul style="list-style-type: none"> • brak wytycznych

Źródło: opracowanie własne autora.

²³ Por. Department of Defense Polygraph Institute, *Federal Psychophysiological Detection of Deception Examiner Handbook* [2006]; przedruk w: „Polygraph”, 40(1), 2011.

2.4. Wytyczne w systemie Utah

Takiej ewolucji jak w federalnym systemie analizy danych nie podlegały natomiast kryteria opracowane przez naukowców związanych z Uniwersytetem Utah. Te drugie, znane już w latach 70., pozostają niezmiennie do dziś (zob. tab. 32)²⁴.

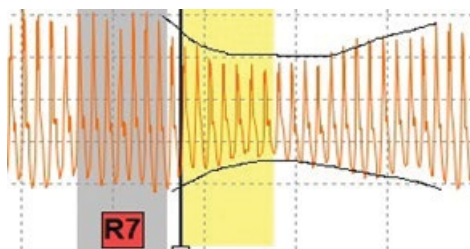
Tab. 32. Aktualne kryteria diagnostyczne w systemie opracowanym przez Uniwersytet Utah

Kanał	System UTAH
Pneumo (ruchy klatki piersiowej odzwierciedlające przebieg cykli oddechowych)	Czas rozpoczęcia reakcji: od wprowadzenia bodźca do 5 sekund po odpowiedzi (z wyjątkiem sytuacji, kiedy charakterystyczne dla badanego jest rozpoczynanie się reakcji nieco później niż u przeciętnego człowieka, tj. 5–8 sekund po odpowiedzi). Zakres reakcji: ≥ 3 cykle, nawet do 20 sekund, jeśli reakcja rozpoczęła się we właściwym czasie. <ul style="list-style-type: none"> • zmniejszenie amplitudy, • podniesienie linii bazowej, • bezdech, • spowolnienie częstotliwości.
EDA (aktywność elektrodermalna – zmiany w przewodności lub oporności skóry)	Czas rozpoczęcia reakcji: 0,5 sekundy od wprowadzenia bodźca do 5 sekund po odpowiedzi (z wyjątkiem sytuacji, kiedy charakterystyczne dla badanego jest rozpoczynanie się reakcji 5–8 sekund po odpowiedzi). Zakres reakcji: od rozpoczęcia reakcji do powrotu do linii bazowej. <ul style="list-style-type: none"> • amplituda (cecha główna mierzona od linii bazowej do szczytu reakcji), • pomocniczo: czas trwania i złożoność.
Kardio (względne ciśnienie krwi i tętno)	Czas rozpoczęcia reakcji: od wprowadzenia bodźca do 5 sekund po odpowiedzi (z wyjątkiem sytuacji, kiedy charakterystyczne dla badanego jest rozpoczynanie się reakcji 5–8 sekund po odpowiedzi). Zakres reakcji: od rozpoczęcia reakcji do powrotu do linii bazowej. <ul style="list-style-type: none"> • podniesienie linii bazowej (wzrost krzywej – wyraźniejszy po stronie rozkurczowej), • czas trwania.
PPG (reakcje naczynioruchowe; zmiany amplitudy pulsu)	Czas rozpoczęcia reakcji: 2 sekundy od wprowadzenia bodźca do 5 sekund po odpowiedzi (z wyjątkiem sytuacji, kiedy charakterystyczne dla badanego jest rozpoczynanie się reakcji 5–8 sekund po odpowiedzi). Zakres reakcji: do 20 sekund. <ul style="list-style-type: none"> • redukcja amplitudy (zwężenie krzywej) i czas trwania tej redukcji (ryc. 59).

Źródło: opracowanie własne autora.

²⁴ Zob. M. Handler, R. Nelson, *Utah Approach...*, *op. cit.*

Ryc. 59. Redukcja amplitudy krzywej reakcji naczynioruchowych



Źródło: materiały własne autora.

2.5. Kryteria diagnostyczne w ESS i ESS-M

W najmłodszym (wprowadzonym na początku obecnego wieku), a zarazem najpopularniejszym współcześnie, systemie analizy danych z testów poligraficznych – ESS – obowiązuje najmniejsza liczba cech diagnostycznych (zob. tab. 33)²⁵. Autorom systemu zależało na jego niezawodności i efektywnym wykorzystywaniu bez względu na poziom doświadczenia zawodowego. Oprócz względnie prostych reguł analizy numerycznej zapisów ustalono zwięzły katalog cech diagnostycznych, niewymagających skomplikowanej interpretacji. Żadnych spekulacji, autorskich koncepcji jak u C. Backstera – wszystko wynikające wyłącznie z badań empirycznych (to kryterium pełnej naukowości spełniał również system Utah, lecz ESS miał być przystępniejszy, nie tracąc na dokładności diagnoz). Odradzano korzystanie z jakichkolwiek narzędzi pomiarowych przy ocenie wielkości odchyłań krzywych na wykresach. Co więcej – uczyniono to wręcz regułą systemu.

Gdy zestawimy powyższe systemy analizy ze sobą, możemy wskazać na pewne różnice w doborze cech diagnostycznych. Zatem, w przypadku kanału pneumo, zwraca uwagę rozbudowana liczba cech w systemie amerykańskich instytucji rządowych. Do 2006 r. za diagnostyczną przesłankę uznaje się w tym systemie hiperwentylację (nazywaną inaczej progresywnym zwiększeniem amplitudy oddechu) – czego nie widzimy w żadnym innym. Tylko u Backstera obowiązuje zasada stwierdzenia reakcji poprzez dedukcję (na podstawie zauważonej ulgi). Natomiast wspólne w każdym z systemów pozostają: tłumienie oddechu, spowolnienie częstotliwości oraz podniesienie linii bazowej.

²⁵ Por. R. Nelson, M. Handler, *Empirical Scoring System: NPC Quick Reference*, Lafayette Instrument 2010.

W kanale EDA tożsamy dla wszystkich systemów jest wzrost amplitudy. Systemy federalny i Utah dodatkowo korzystają jeszcze z czasu trwania i złożoności reakcji. W starszej wersji systemu federalnego (USAMPS) mamy ponadto *skaczący*, gwałtownie opadający wzrost krzywej.

Tab. 33. Aktualne kryteria diagnostyczne w systemie ESS / ESS-M

Kanał	System ESS/ ESS-M
Pneumo (ruchy klatki piersiowej odzwierciedlające przebieg cykli oddechowych)	Czas rozpoczęcia reakcji: brak sztywnych ram; generalnie od wprowadzenia bodźca do 5 sekund po odpowiedzi. Zakres reakcji: ≥ 3 cykle, do 15–20 sekund. • zmniejszenie amplitudy, • spowolnienie częstotliwości oddychania, • podniesienie linii bazowej.
EDA (aktywność elektrodermalna – zmiany w przewodności lub oporności skóry)	Czas rozpoczęcia reakcji: 0,5 sekundy od wprowadzenia bodźca do 5 sekund po odpowiedzi. Zakres reakcji: do 15–20 sekund. • amplituda (mierzona od linii bazowej do szczytu reakcji).
Kardio (względne ciśnienie krwi i tętno)	Czas rozpoczęcia reakcji: od wprowadzenia bodźca do 5 sekund po odpowiedzi. Zakres reakcji: do 15–20 sekund. • amplituda (podniesienie linii bazowej).
PPG (reakcje naczynioruchowe; zmiany amplitudy pulsu)	Czas rozpoczęcia reakcji: 2 sekundy od wprowadzenia bodźca do 5 sekund po odpowiedzi. Zakres reakcji: do 15–20 sekund. • redukcja amplitudy (zwięźnienie krzywej).

Źródło: opracowanie własne autora.

W kanale kardio jednolite stanowisko przyjmuje się wszędzie w kwestii podniesienia linii bazowej (wzrostu ciśnienia krwi). Systemy federalny i Utah nie pomijają jeszcze czasu trwania zmiany w reakcjach. U Backstera lista cech jest szersza o zmiany tętna, zmiany trendu reagowania i pomocniczą zasadę stwierdzania reakcji poprzez dedukcję.

Przy reakcjach naczynioruchowych jako diagnostyczna jest wskazywana redukcja amplitudy. Mówią o tym systemy Utah i ESS. Natomiast pozostałe nie odnoszą się do tego kanału zapisu.

Subiektywizm badającego zaczyna się właściwie już na etapie wyboru jednej z uznanych metod interpretacji danych testowych. Pewne różnice między systemami mogą wprowadzać perturbacje dla badających, dlatego ważne jest, aby po dokonaniu wyboru konkretnej metody postępować już ściśle według jej reguł.

Patrząc na liczbę cech – w systemie federalnym mamy obecnie 12, w systemie Utah – 10, natomiast ESS – już tylko 6 (przy czym niektóre kryteria w kanale oddechowym obejmują *de facto* też kilka innych, odrębnie wyszczególnionych w poprzednich systemach, jak np. bezdech, który jest skrajną postacią tłumienia oddechu). Tak czy inaczej – widzimy, że systemy ewoluowały w kierunku ograniczenia cech diagnostycznych do tych niekwestionowanych, z solidnymi podstawami naukowymi i względnie klarownych, z ograniczeniem liczby niuansów w zapisach reakcji pozostawionych biegłym bez sztywnych wytycznych. Margines subiektywizmu też przez to niechybnie zmniejszył się, ale nie został wyeliminowany. Jednocześnie nie należy przez to deprecjonować wartości dowodu z opinii biegłego z zakresu badań poligraficznych. Subiektywizm towarzyszy w nie mniejszym stopniu innym badaniom kryminalistycznym – np. badaniom dokumentów. Jak zauważa E. Gruza – w takiej ekspertyzie,

opierając się na wynikach prowadzonych badań, dokonujemy jednocześnie ich oceny, opartej zarówno na doświadczeniu, jak i spostrzegawczości biegłego, które mogą być zabarwione subiektywizmem. Stosowane najczęściej metody graficzno-pomiarowe mimo, że traktowane są jako obiektywne, nacechowane są pewną niewielką dozą subiektywizmu, co oczywiście nie oznacza, że zezwalają na dowolność tłumaczenia wyników, ale jednak dość często odwołują się do interpretacji zapisu²⁶.

3. Zapewnienie jakości badań poligraficznych, znaczenie pracy zespołowej

3.1. Kontrola sądowa opinii biegłego

Gdy podmiot potrzebujący opinii powierza biegłemu wykonanie określonych badań, np. poligraficznych – czyni to z założeniem, że będą one wykonane w sposób kompetentny, rzetelny i obiektywny. Jak wykazano wyżej – ze względu na niedookreśloność niektórych cech diagnostycznych zapisów na poligramach oraz zasad oceniania w ramach systemów analizy danych testowych, a także ze względu na różnice między tymi systemami – poligrafer opiniuje w warunkach, w których istnieje przestrzeń na podejmowanie subiektywnych decyzji²⁷. Ta przestrzeń

²⁶ E. Gruza, *Granice opiniowania biegłego*, [w:] Z. Kegel (red.), *Problemy dowodu z dokumentu*, Wrocław 1999, s. 81.

²⁷ Wśród przykładów nieostrych reguł analitycznych można wymienić np. zasadę „zbyt piękne, żeby było prawdziwe” przy podejmowaniu decyzji, czy

dodatkowo zwiększy się, jeśli poligrafer nie będzie postępował ściśle według obiektywnych reguł przewidzianych dla poszczególnych metod interpretacji danych, np. poprzez wybiórcze korzystanie z tych reguł; pomijanie takich, które nie pasują do wcześniejszych oczekiwań, odwoływanie się w pierwszej kolejności do własnych doświadczeń z praktyki, autorskie modyfikacje technik badawczych.

Przykładowo – w jednej z publikacji biegły J. Bieńkuński opisuje własne arbitralne modyfikacje standaryzowanych technik badawczych, w uzasadnieniu odwołuje się wyłącznie do własnych doświadczeń i zapewnia *a priori*, że nie miało to wpływu na wartość diagnostyczną tych testów:

„Wprowadzona przez biegłego korekta polegała na tym, że zamiast dwóch pytań kontrolnych zastosowano tylko jedno, aby mieć większą możliwość wykorzystania pytań krytycznych dotyczących wyjaśnianej sprawy. Nie oznaczało to w żadnym wypadku, że taki test traci wartość diagnostyczną. [...] Korekta tego oryginalnego testu (dokonana przez biegłego w niniejszej sprawie) polegała jedynie na tym, że zamiast jednego pytania kontrolnego wprowadzono tu dwa pytania kontrolne. Ta modyfikacja również w żaden sposób nie wpływa na wartość diagnostyczną testu. Za powyższym przemawiają wyjaśniane sprawy, gdzie „oczyszczano” niesłusznie podejrzewane osoby o popełnienie ciężkich przestępstw”¹⁾.

Do wyrażonego stanowiska należy się odnieść krytycznie. Choć prawdą jest, że autorsko zmodyfikowany test nie traci wartości diagnostycznej (każdy jakąś ma – większą lub mniejszą), ale bez empirycznego poparcia staje się ona nieznaną. Oderwanie od standardów w badaniach poligraficznych prowadzi do zastępowania metody naukowej intuicyjnymi, subiektywnymi ocenami, wzmagając stopień niepewności co do wyników testów.

¹⁾ P. Stryżowski, J. Bieńkuński, *Wykorzystanie i znaczenie ekspertyz wariograficznych. Studium przypadku*, „Prokuratura i Prawo”, 7–8, 2014, s. 181.

mamy do czynienia z zapisem odzwierciedlającym naturalną reakcję organizmu na bodziec testowy, czy jednak z tzw. artefaktem; punktowaniu podlega „każda **wyraźna** różnica” między porównywanymi reakcjami. Do niejednolitego postępowania przyczyniają się też niekiedy nieprzejrzyste procedury instytucjonalne – w tym brak wyraźnych formalnych wytycznych w postępowaniach kwalifikacyjnych do służb, które określałyby jak postępować w zależności od różnych wariantów wyników testów oraz treści informacji udzielanych przez osoby badane – kiedy badanie kontynuować, a kiedy należy je ostatecznie przerwać.

W tym kontekście zarysowuje się wyraźna potrzeba mechanizmów, gwarantujących odpowiednią jakość badań poligraficznych oraz intersubiektywną kontrolowalność ich wyników – jednakowe zrozumienie i możliwość weryfikacji przez innych biegłych przedmiotowej specjalności. W przypadku biegłych indywidualnych w zasadzie możemy mówić jedynie o kontroli sądowej – w konkretnych postępowaniach, w których biegły opiniuje. Ewentualnie o kontroli *naukowej*, o ile treść opinii zostanie upubliczniona, albo udostępniona pracownikowi akademickiemu czy doktorantowi na potrzeby analityczne i poddana krytycznej recenzji²⁸. Jeżeli poligrafer należy do zrzeszenia zawodowego czy towarzystwa naukowego, może dobrowolnie podlegać procedurom kontrolnym w tych organizacjach, jeśli takowe zostały ustanowione. Część biegłych pozostaje jednak poza takimi strukturami i nikomu nie podlega.

W ramach kontroli sądowej – „jeżeli opinia jest niepełna lub niejasna albo gdy zachodzi sprzeczność w samej opinii lub między różnymi opiniami w tej samej sprawie, można wezwać ponownie tych samych biegłych lub powołać innych” (art. 201 k.p.k.). Sąd i strony procesowe mogą zweryfikować ewentualne wątpliwości w czasie przesłuchania biegłego (zob. art. 197 § 3 k.p.k. odsyłający do przepisów odnoszących się do przesłuchania świadka). Ponadto: „Jeżeli ujawnią się powody osłabiające zaufanie do wiedzy lub bezstronności biegłego albo inne ważne powody, powołuje się innego biegłego” (art. 196 § 3 k.p.k.).

Strona procesowa może również złożyć wniosek o odczytanie na rozprawie opinii *prywatnej* (na podstawie 393 § 3 k.p.k.) i w ten sposób posiłkować się w ewentualnym kwestionowaniu ustaleń zawartych w opinii biegłego powołanego przez organ procesowy. Według S. Waltośa prywatnej opinii biegłego można nie dopuścić w zasadzie tylko, opierając się na przesłankach wyrażonych w art. 170 § 2 k.p.k. – czyli wtedy, gdy „okoliczność, która ma być udowodniona, nie ma znaczenia dla rozstrzygnięcia sprawy albo jest już udowodniona zgodnie

²⁸ Przykładem kontroli naukowej – artykuł L. Stępki w monografii *Kryminalistyka a nowoczesne technologie* pod red. V. Kwiatkowskiej-Wójcikiewicz, D. Wilka i J. Wójcikiewicza, gdzie autor poddał analizie zarejestrowany audiowizualnie przebieg badania poligraficznego, przeprowadzonego przez biegłego z Laboratorium Kryminalistycznego KWP w Bydgoszczy, a także zeznania biegłego z późniejszego przesłuchania na rozprawie w styczniu 2018 r. (zob. L. Stępka, *Rejestracja i ocena poprawności ekspertyzy wariograficznej (poligraficznej)*, [w:] V. Kwiatkowska-Wójcikiewicz, D. Wilk, J. Wójcikiewicz (red.), *Kryminalistyka a nowoczesne technologie*, Kraków 2019, s. 99–117).

z twierdzeniem wnioskodawcy”. *Per analogiam* zastosowanie powinny mieć ponadto przesłanki wyłączenia biegłego (z art. 196 k.p.k.), a po dopuszczeniu takiego dowodu – kryteria jego oceny z art. 201 k.p.k.²⁹ Jak zauważa J. Widacki:

Gdy prywatna opinia jest załącznikiem do pisma procesowego strony, ustosunkowującego się do opinii biegłego powołanego przez sąd, może ewentualnie lepiej i bardziej fachowo niż czyniłaby to sama strona wskazać, że opinia biegłego zawiera sprzeczności, jest niepełna lub niejasna. Te zastrzeżenia i uwagi sąd ma obowiązek rozważyć i odnieść się do nich, tak samo, jak miałby obowiązek odnieść się do zastrzeżeń i uwag samej strony³⁰.

Gdy z kolei sąd zdecydowałby się wysłuchać biegłego prywatnego – od tego momentu przyjmie on rolę procesową biegłego *ad hoc* (w trybie art. 193 § 1 k.p.k.), a zastosowanie do niego będą miały w całej rozciągłości przepisy rozdziału 22 k.p.k.³¹ W innej roli autor opinii prywatnej może wystąpić jedynie jako świadek na okoliczność powstania dokumentu.

Ewentualne sprzeczności pomiędzy ustaleniami biegłych podlegają swobodnej ocenie organu prowadzącego postępowanie. Nie istnieje jakiś odrębny wykaz kryteriów, które stanowiłyby podstawę takiego rozstrzygnięcia. Biegłych się przesłuchuje, niekiedy włącznie z zastosowaniem konfrontacji. Zdarza się, że powoływany jest kolejny biegły i zazwyczaj poszukuje się wówczas osoby o możliwie najwyższych kwalifikacjach, autorytetu w danej dziedzinie. Zwykle rozbieżności między opiniami biegłych z zakresu badań poligraficznych powstają albo z powodu błędu statystycznego (niedokładnego wyniku testu), albo w konsekwencji błędów w metodyce przeprowadzonego badania. Bywa również, że sprzeczności między wnioskami z dwóch opinii mają jedynie pozorny charakter – zależnie od tego, jakie pytania były zadawane podczas badań poligraficznych (niekoniecznie tożsame), jakiej treści oświadczenia składała osoba badana, albo w jakim czasie prowadzono testy z wykorzystaniem poligrafu (może się zdarzyć, że w późniejszym czasie stan wiedzy osoby poddawanej badaniu na temat krytycznego zdarzenia jest już odmienny niż przy pierwszym badaniu). Każda opinia biegłego wymaga więc wnikliwej analizy.

²⁹ S. Waltoś, *Opinie biegłych...*, *op. cit.*, s. 36–37.

³⁰ J. Widacki, *Obrońca wobec opinii biegłego...*, *op. cit.*, s. 17–18.

³¹ Zob. S. Waltoś, *Opinie biegłych...*, *op. cit.*, s. 37.

Gdy chodzi o pracowników instytucji państwowych oraz prywatnych instytutów badawczych – formalny i (przynajmniej teoretycznie) merytoryczny nadzór nad poligraferem sprawuje właściwy przełożony. Wydaje się, że powyższe formy kontroli nie są wystarczające, aby zminimalizować liczbę opinii niewiarygodnych, błędnych, wykonanych z naruszeniem reguł sztuki. Problem dotyczy oczywiście także innych rodzajów badań, nie tylko poligraficznych, ale – zważywszy na przedmiot wywodu – to tym ostatnim autor niniejszej monografii poświęca uwagę.

Dla niektórych typów badań, w sposób zróżnicowany dla poszczególnych krajów, wymaga się, aby ekspertyzy sądowe (i nie tylko) były wykonywane wyłącznie przez akredytowane laboratoria (np. wykonujące badania DNA w Holandii), albo certyfikowanych biegłych (np. rewidentów w Polsce – zob. ustawa z dnia 11 maja 2017 r. o biegłych rewidentach, firmach audytorskich oraz nadzorze publicznym, Dz.U. z 2017 r., poz. 1089). Generalnie w polskim wymiarze sprawiedliwości zagadnienie certyfikacji biegłych i atestacji laboratoriów nie jest doceniane, a traci na tym jakość ekspertyz, uznawanych za dowody.

3.2. Instytucjonalne procedury kontroli jakości badań poligraficznych

Odpowiedzią na potrzeby badań i opiniowania na odpowiednim poziomie jest prowadzenie polityki tzw. zapewnienia jakości, które może obejmować: ustalenie procedur i kontrolę ich przestrzegania, działania zapobiegające przyczynom błędów, eliminację niezgodności z procedurami poprzez działania korygujące, a jeśli finalnie procedury okazałyby się zbyt lub za mało restrykcyjne, celowa mogłaby się okazać ich zmiana. J. Hebenstreit słusznie zauważa, że w naukach sądowych kluczową rolę odgrywają zdolności i doświadczenie biegłego. Przez to można by sądzić

że tego typu badania są subiektywne i nie spełniają wymagań systemu zapewnienia jakości. Ale i takie badania w wielu wypadkach można dostosować do tych wymagań, na przykład przez ich walidację, obowiązkową weryfikację wyników przez niezależnego badacza (być może nawet w innym laboratorium) i uczestnictwo w testach międzylaboratoryjnych, które pozwalają na porównanie wyników dotyczących tego samego materiału, uzyskiwanych w różnych laboratoriach³².

³² J. Hebenstreit, *Zapewnienie jakości w laboratoriach sądowych*, [w:] J. Wójcikiewicz (red.), *Ekspertyza sądowa. Zagadnienia wybrane*, Warszawa 2007, s. 610.

Potrzeba kontroli jakości nie powinna budzić wątpliwości – nie tylko ze względu na potencjalne skażenie subiektywizmem przy prowadzeniu badań poligraficznych (i w zasadzie wszelkich innych), ale z powodu istnienia innych źródeł błędnych opinii biegłych. Na przykład w różnych laboratoriach badawczych, poza subiektywizmem, spotyka się takie problemy jak: stosowanie technik i procedur, które są nie do zaakceptowania; kontaminację śladów; niedoświadczenie personelu; niejednoznaczne rezultaty prowadzące do sprzecznych interpretacji; niewłaściwe skalibrowanie sprzętu; wyolbrzymienie pewności badającego czy zlekceważenie wpływu czynników zewnętrznych na wyniki testów³³. Poza formalnymi procedurami dla różnych specjalności ustanawiane są rozmaite wytyczne zrzeczeń biegłych i tzw. katalogi dobrych praktyk. Takowe w 2015 r. ogłosiła np. Europejska Sieć Instytutów Badań Kryminalistycznych (ENFSI, European Network of Forensic Science Institutes) dla dziesięciu różnych typów badań³⁴. W kolejnych latach zbiór ten został poszerzony, ale wśród nich – jak dotąd – zabrakło miejsca dla badań poligraficznych. Dla tej specjalności punktem odniesienia są przede wszystkim standardy przyjęte i ogłoszone przez wiodące zrzeszenia poligraferów – o zasięgu międzynarodowym (APA) i krajowym (jak w Polsce – Polskie Towarzystwo Badań Poligraficznych), a także standardy opublikowane przez organizację normalizacyjną – ASTM International³⁵.

Mając na uwadze rozmaite aspekty kontroli jakości badań poligraficznych – można wyodrębnić jej kilka form. Na przykład ze względu na czas organizowania zadań kontrolnych przybierają one charakter stały, okresowy lub doraźny (*ad hoc*). Inny podział obejmuje kontrolę merytoryczną i formalną.

Kontrola merytoryczna dotyczy zgodności postępowania poligrafera z regułami sztuki, standardami badań poligraficznych oraz poprawności diagnozowania i formułowania wniosków z badań. Z kolei kontrola formalna odnosi się do aspektów proceduralnych – od momentu przyjęcia zlecenia w oparciu o odpowiednią podstawę prawną, poprzez udokumentowanie przeprowadzonych czynności,

³³ L. Meintjes-van der Walt, *Expert Evidence in the Criminal Justice Process. A Comparative Perspective*, Amsterdam 2011, s. 100–105.

³⁴ European Network of Forensic Science Institutes, *Best Practice Manuals*, enfsi.eu/documents/best-practice-manuals [dostęp: 18.05.2020].

³⁵ Zob. omówienie standardów w: M. Gołaszewski, M. Widacki, *Aktualny standard badań...*, *op. cit.*

skończywszy na sprawdzeniu sprawozdania z badań lub opinii, które powinny spełniać określone wymogi formalne w zależności od rodzaju postępowania, w którego ramach badanie zlecono. Jeżeli powołuje się biegłego indywidualnego – kontrolę formalną sprawuje wówczas zleceniodawca, np. organ procesowy. W przypadku biegłych instytucjonalnych ten rodzaj kontroli jest jeszcze sprawowany przez przełożonych poligrafera.

Nie ulega wątpliwości, że kontrolę merytoryczną powinny sprawować osoby posiadające wiadomości specjalne w zakresie badań poligraficznych. Nie jest to problematyczne, gdy istnieją oficjalne procedury kontroli jakości i wyznaczone do tego profesjonalne podmioty. Trudności mogą pojawiać się natomiast, gdy kontrola merytoryczna pozostaje z mocy prawa w kompetencjach podmiotów, które stosownych wiadomości specjalnych nie posiadają. Na przykład w postępowaniach jurysdykcyjnych zadanie merytorycznej kontroli opinii biegłego spoczywa na sędziwie. Sąd powołuje biegłego, ponieważ nie posiada wiadomości specjalnych do oceny danego problemu, a jednocześnie musi dokonać oceny merytorycznej dowodu, jaki od biegłego otrzyma, mimo to wydaje się, że sędziowie z racji wykształcenia prawniczego powinni mieć przynajmniej podstawowe informacje o istocie badań poligraficznych, ale pytanie – czy tak rzeczywiście jest? W razie wątpliwości dopuszczalne jest powołanie innego biegłego. Podobnie – strony postępowania – również mogą posilkować się wsparciem innych poligraferów, choćby prywatnie. A co jeśli badanie poligraficzne jest przeprowadzane na użytek wewnętrzny w danej instytucji – np. w celach przedzatrudnieniowych lub w ramach kontroli pracownika, funkcjonariusza czy żołnierza? Zwykle nadzór merytoryczny w komórce organizacyjnej sprawuje bezpośredni przełożony, a nie jest przy tym regułą, że w komórkach zajmujących się badaniami poligraficznymi funkcje te sprawują osoby posiadające wiadomości specjalne. Pozostaje wówczas retoryczne pytanie – czy poligrafer podlega wówczas jakiegokolwiek kontroli merytorycznej, jeśli nie istnieje procedura, do której zaangażowani są profesjonaliści? A gdyby takowa istniała, np. okresowa – to czy byłoby to wystarczające? – to ostatnie pytanie pozostawiam otwarte.

Kontrolę jakości można prowadzić zarówno w ramach danej jednostki organizacyjnej, gdzie wykonuje się badania poligraficzne, a obok tego – również przez podmioty zewnętrzne. Kontrola wewnętrzna może przybrać formę sformalizowaną – na podstawie oficjalnych procedur nadzoru, a obok tego może odbywać się poprzez nieformalne konsultacje

merytoryczne z udziałem innych poligraferów. Aby możliwe były bieżące konsultacje między poligraferami – wskazane jest, aby ich praca była organizowana w ramach zespołów, a więc w przynajmniej dwie osoby. O zaletach pracy zespołowej świadczą m.in. ustalenia G. Barlanda. Gdy prowadził badania nad dokładnością i rzetelnością testów wykonanych techniką Backstera, zauważył, że zsumowane oceny ewaluatorów poligramów dawały wyższy procent dokładności wyniku niż średnia dla pojedynczego poligrafera (odpowiednio 86% wobec 81,7%)³⁶.

Kontrola merytoryczna może odbywać się w czasie rzeczywistym – podczas badania poligraficznego, a także *post factum*. *Na żywo* (współcześnie – z wykorzystaniem środków technicznych umożliwiających transmisję i komunikację z poligraferem zlokalizowanym w odrębnym pomieszczeniu, a w przeszłości po prostu z udziałem innego poligrafera towarzyszącego badającemu w pracowni) realizowane jest zazwyczaj wsparcie dla osób szkolących się lub odbywających staż przystosowujący do samodzielnej praktyki. Niekiedy, przy sprawach szczególnej wagi, jest to wskazane na zasadzie konsultacji zespołowej. Obserwacja przebiegu badania może być też elementem nadzoru.

Kontrola badań już wykonanych jest natomiast oczywiście przeprowadzana na podstawie sporządzonej przez poligrafera dokumentacji. Według wytycznych APA taka dokumentacja powinna obejmować w szczególności: raport z badania, zarejestrowane dane testowe oraz zapis audio-video. Recenzję kontrolną należy sporządzać na piśmie, z uwzględnieniem danych dotyczących: stron wnioskujących o kontrolę, osoby badanej (fakultatywnie), daty badania, identyfikacji badającego, typu badania (diagnostyczne lub przesiewowe), wykazu materiałów przedkładanych do sprawdzenia, oceny otoczenia badawczego, wywiadu przedtestowego, sformułowania pytań testowych, zebrania danych testowych i ich analizy.

Przy ocenie przedstawionych badanemu informacji w czasie wywiadu przedtestowego bierze się pod uwagę to, czy badający zweryfikował tożsamość osoby badanej, objaśnił cel i procedurę badania, uzyskał zgodę rozmówcy na poddanie się testom z wykorzystaniem poligrafu, odpowiednio omówił zagadnienia będące przedmiotem badania oraz pytania testowe, zanim zostały zadane w czasie testów. Istotne jest również to, czy wywiad nie był prowadzony w sposób oskarżycielski, krytyczne pytania testowe zostały skonstruowane

³⁶ G.H. Barland, *The Reliability of Polygraph...*, *op. cit.*, s. 202.

w sposób jednoznaczny i opisują problemy istotne dla przedmiotu postępowania. Ponadto – czy pytania kontrolne są spójne z typem, który miał być zastosowany i czy generalnie możliwa była odpowiedź tak lub nie, albo inna wynikająca z instrukcji badającego.

Ocena fazy testów odnosi się do tego, czy poligrafer przeprowadził test demonstracyjny, wykorzystał uznany format testu; czy czujniki działały prawidłowo (włącznie z czujnikiem ruchu); czy tempo zadawania pytań było właściwe i czy odbywało się to w neutralnym tonie; czy poprawnie naniesiono przypisy na poligramach. Jeżeli chodzi o zarejestrowane dane testowe – to sprawdza się, czy ich jakość jest odpowiednia; czy badający zastosował uznaną metodę analizy; czy pominął dane zakłócone.

Recenzja powinna wskazywać na ewentualne źródła rozbieżności między oceną badającego a wnioskami kontrolerów, wyszczególnić wszelkie odstępstwa od standardów badawczych, ale bez wypowiedzania się na temat kompetencji poligrafera. Nie należy zmierzać do zmiany opinii, o ile nie doszło do rażącego naruszenia reguł sztuki. Konkluzje kontrolerów mają ograniczyć się do tego, czy oceniane badanie wspierają, czy też nie, albo że potrzeba dodatkowych danych³⁷.

T. Shurany *et al.* uważają, że kontrola jakości badań poligraficznych powinna przebiegać w trzech etapach. Na początku – *ślepa* ocena poligramów. Po jej zakończeniu – zapoznanie się ze szczegółami sprawy, aby sprawdzić poprawność sformułowania pytań testowych. Z kolei w trzecim etapie – analiza zapisu audiowizualnego w celu upewnienia się, że poligrafer nie naruszył standardów prowadzenia badań³⁸.

Zajmujący się zjawiskiem tendencyjnego opiniowania S. Kassin *et al.* uważają za konieczne wprowadzenie procedury tzw. *ślepego oceniania* – i to na tyle rygorystycznego, by oceniający w ogóle nie wiedział, czyj materiał badawczy otrzymał, a ekspert podstawowy nie miał żadnego wpływu na wybór kontrolera. Nieodzowne są również szkolenia, w których należałoby uwzględniać podstawy psychologii, w tym obejmującej zagadnienia percepcji, osądów i podejmowania decyzji³⁹.

³⁷ APA, *APA Model Policy for Quality Assurance*, www.polygraph.org/assets/docs/Misc.Docs/QA%20Model%20Policy%20%28approved%207-20-2017%29.pdf [dostęp: 29.03.2021].

³⁸ T. Shurany *et al.*, *Influence of Case Facts...*, *op. cit.*, s. 138.

³⁹ S.M. Kassin, I.E. Dror, J. Kukucka, *The Forensic Confirmation Bias. Problems, Perspectives, and Proposed Solutions*, „Journal of Applied Research in Memory and Cognition”, 2(1), 2013, s. 49–50.

Stosowanie ślepej oceny powinno ograniczać m.in. wpływ wcześniejszych oczekiwań na interpretację danych testowych. W wielu instytucjach w USA i w Izraelu współcześnie jest to procedura rutynowa, a nie zawsze tak było. W 1950 r. C.M. Wilson, ówczesny przewodniczący International Society for the Detection of Deception (ISDD) – zapytany o to, czy wykresy powinny być udostępniane po badaniu – przekonywał, że przyniosłoby to jedynie niepotrzebny zamęt, zwłaszcza gdy otrzymałaby je osoba niewykształcona. Jego zdaniem wykresy nic nie mówią temu, kto osobiście nie przeprowadził badania. Dlatego nigdy nikomu nie pokazał swoich wykresów⁴⁰. W tamtych latach kryteria diagnostyczne były jednak bardzo nieprecyzyjne i niejednolicie stosowane. Jak przy stanowisku wyrażanym przez C.M. Wilsona można było mówić o rzetelności testów poligraficznych? Jak przeciwdziałać błędom w sztuce i zwyczajnym nadużyciom, kiedy nikt nie miał możliwości skontrolowania danych, na podstawie których poligrafier wydał określoną opinię?

Tak jak nie ma pewności co do jakości pracy poligrafera bez udostępnienia poligramów, tak również nie można usunąć ewentualnych wątpliwości w sprawie przebiegu badania bez zapisu audiowizualnego. Jak pisze L. Stępka: „Nagranie jest podstawą weryfikacji czynności przez organ procesowy, a także przez eksperta w ramach meta opinii”⁴¹. Nagranie może ułatwić ocenę badania i rzetelność uzyskanego dowodu. Sąd samodzielnie może stwierdzić np., czy nie doszło do naruszenia zakazu dowodowego powiązania badania poligraficznego z przesłuchaniem (art. 171 § 5 k.p.k.). Natomiast inny biegły jest wówczas w stanie odnieść się do kwestii metodologicznych. O ile przepisy prawa nie stanowią inaczej – APA wymaga w swoich wytycznych przechowywania dokumentacji z badań przez co najmniej trzy lata⁴², choć w ogólnym standardzie jest mowa tylko o rocznym okresie⁴³. Zapis audio-video powinien przebiegać w sposób ciągły (obejmować każdy etap badania), a jakiegokolwiek przerwy należy wyjaśnić w trakcie nagrywania⁴⁴.

⁴⁰ C.M. Wilson, *Should Graphs be Released or Shown after Test?*, „ISDD Bulletin”, 3(3), 1950; za: N. Ansley, *Development of Deception Criteria Prior to 1950*, „Polygraph”, 28(1), 1999, s. 28.

⁴¹ L. Stępka, *Rejestracja i ocena poprawności ekspertyzy...*, *op. cit.*, s. 99.

⁴² APA, *Model Policy for Release...*, *op. cit.*

⁴³ APA, *Standards of Practice...*, *op. cit.*

⁴⁴ Zob. *ibidem*; ponadto: APA, *Model Policy for law enforcement...*, *op. cit.*

APA rekomenduje, aby wycinek pracy każdego poligrafera w instytucjach publicznych podlegał regularnej niezależnej ocenie⁴⁵. Poza kontrolą indywidualną ocenia się także ogół procedur realizowanych w poszczególnych instytucjach, a także przestrzeganie warunków licencyjnych (o ile na danym obszarze zostały ustanowione) przez biegłych indywidualnych. W USA Departament Obrony wydał wytyczne, z których wynika, że każda instytucja federalna, w której przeprowadza się badania poligraficzne, powinna ustanowić program kontroli jakości, ewentualnie podpisać porozumienie o współpracy z inną agencją rządową, która taki program posiada. Zaleca się, aby była to procedura scentralizowana, z udziałem doświadczonych poligraferów o potwierdzonych kwalifikacjach.

Kontrole są zapowiedziane, organizowane planowo raz na dwa lata. Na poziomie ogólnym ocenia się wewnętrzne zarządzenia, sprawozdania z realizacji programu badań, dane statystyczne. Natomiast szczegółowej kontroli poddawane są wytypowane badania na zasadzie próby reprezentatywnej. Wyniki badań poligraficznych zyskują miano ostatecznych dopiero po takiej kontroli. Procedurę wieńczy raport końcowy z wnioskami i rekomendacjami. Kierownik danej jednostki musiał się do tego ustosunkować w odpowiedzi do dyrektora Instytutu Badań Poligraficznych Departamentu Obrony (obecnie adresatem jest Narodowe Centrum Oceny Wiarygodności)⁴⁶.

W 2011 r. powstał raport – opracowany przez Biuro Podsekretarza Obrony USA do spraw Wywiadu – na temat oceny praktyki, stosowanej metodologii badań poligraficznych oraz przestrzegania procedur obowiązujących w jednostkach podległych Departamentowi Obrony⁴⁷. Analiza objęła badania wykonane od 1 maja 2010 r. do 30 kwietnia 2011 r. – łącznie 43 434 (41 057 – badania przesiewowe; 1537 – w śledztwach kryminalnych, a 840 – w sprawach kontrwywiadowczych). Dzięki badaniom przesiewowym uzyskano w raportowanym okresie prawie 4 tys. przyznań się do rozmaitych przewinień – od naruszeń zasad bezpieczeństwa i niewykazania kontaktów

⁴⁵ APA, *Model Policy for law enforcement...*, *op. cit.*

⁴⁶ Department of Defense Polygraph Institute, *Federal Psychophysiological Detection...*, *op. cit.*

⁴⁷ Office of the Under Secretary of Defense for Intelligence, *Department of Defense Polygraph Program Process and Compliance Study*, 19.12.2011, fas. org/sgp/othergov/polygraph/dod-poly.pdf [dostęp: 19.10.2012].

zagranicznych po problemy natury kontrwywiadowczej oraz przestępstwa kryminalne⁴⁸.

W przedmiotowym raporcie zwrócono ponadto uwagę m.in. na niejednolite postępowanie różnych instytucji w sytuacji, gdy przesiewowe badanie funkcjonariusza zakończy się niekorzystnym wynikiem (przy czym za taki uznawało się nie tylko wskazanie na znaczące reakcje przy pytaniach krytycznych, ale również wynik nierozstrzygnięty), a jednocześnie badany nie przyzna się do niczego. Zazwyczaj zlecano powtórne badania, a równolegle – osoba badana podlegała sprawdzeniu przez personel innych jednostek. Jeśli takie dodatkowe dochodzenie niczego nie wykazywało, funkcjonariusz pozostawał z warunkowym dostępem do informacji niejawnych albo miał ten dostęp czasowo wstrzymany, lecz wyniki badań poligraficznych nie stanowiły (przynajmniej oficjalnie) podstawy do podejmowania jakichkolwiek niekorzystnych decyzji kadrowych⁴⁹.

W Polsce próżno szukać jakichkolwiek form zewnętrznej kontroli nad wykonywaniem badań poligraficznych w instytucjach państwowych. Z kolei ewentualne procedury wewnętrzne, o ile gdziekolwiek istnieją, nie są podane do publicznej wiadomości. Nie ma także indywidualnych wymogów licencyjnych stawianych poligraferom. Przykładowo w USA, w stanie Teksas, aby otrzymać licencję, kandydat winien posiadać wyższe wykształcenie (tytuł licencjata) lub legitymować się co najmniej 5-letnim doświadczeniem w pracy śledczej w okresie poprzedzającym aplikację, zdać pisemny i praktyczny egzamin, pomyślnie przejść sprawdzenia w kartotece kryminalnej oraz przedstawić dowód ubezpieczenia lub uiścić kaucję gwarancyjną w wysokości 5 tys. dolarów na wypadek stwierdzenia naruszenia przepisów w związku z przeprowadzaniem badań poligraficznych⁵⁰. Polskie Towarzystwo Badań Poligraficznych poza – wspomnianą już wcześniej przy omawianiu kwalifikacji poligraferów – procedurą poświadczania rekomendacji eksperckich przewiduje możliwość przeprowadzenia kontroli jakości przez komitet do spraw metodyki badań poligraficznych, w tym rozpatrzenie ewentualnej skargi, ale musi to dotyczyć

⁴⁸ *Ibidem*, s. 28.

⁴⁹ *Ibidem*, s. 31–32.

⁵⁰ *POLYGRAPH EXAMINERS Administrative Rules of the Texas Department of Licensing and Regulation 16 Texas Administrative Code, Chapter 88 (effective September 1, 2015)*, www.tdlr.texas.gov/polygraph/polygraphrules090115.pdf [dostęp: 29.03.2021].

osoby należącej do Towarzystwa. Nie ma właściwie żadnego instrumentu nadzoru na tymi, którzy pozostają poza strukturami organizacji i świadczą swoje usługi na rynku prywatnym, a nierzadko znajdują się na listach biegłych sądowych.

Podsumowanie

W nawiązaniu do tego, co zostało już powiedziane na wstępie – badania poligraficzne mają współcześnie dość szerokie zastosowanie – nie tylko kryminalistyczne, ale także przy realizacji zadań operacyjnych służb, procedurach kadrowych, a nawet terapeutycznych. Były tu rozpatrywane przede wszystkim jako badania przeprowadzane przez biegłego w formie ekspertyzy na różnych etapach procesu karnego, wobec świadków, podejrzanych czy oskarżonych – w celach śledczych i dowodowych. Takie ekspertyzy prowadzą do uzyskania dowodu pośredniego – czyli dopiero do ustaleń, w których następstwie można w drodze rozumowania redukcyjnego wyciągać wnioski na temat faktu głównego. Tak samo jest z wieloma badaniami identyfikacyjnymi także tymi postrzeganymi za najpewniejsze – jak dowód z badań genetycznych.

Badanie poligraficzne należy traktować jako skuteczny sposób ochrony praw jednostki, w szczególności prawa do obrony. Przeprowadzane zawsze za zgodą osoby badanej, nie jest w żadnej mierze zagrożeniem tych praw – tym bardziej że inne metody, np. związane z pracą operacyjną organów ścigania, budzą znacznie poważniejsze zastrzeżenia, a rozmaite czynności procesowe mogą być zdecydowanie bardziej dolegliwe¹. Co więcej – badanie poligraficzne nie dość, że nie narusza godności człowieka, to niejednokrotnie pozwoliło odzyskać ową godność ludziom niewinnym, niesłusznie podejrzewanym o popełnienie przestępstwa. Ekspertyza poligraficzna może dostarczyć dowodu zarówno obciążającego, jak i odciążającego – w szczególności tego drugiego, jak podkreśla np. R. Jaworski².

Opinie biegłych z zakresu badań poligraficznych należą do kategorii tzw. dowodów naukowych. To za sprawą tych właśnie badań, na

¹ Por. R. Jaworski, *Opinia z ekspertyzy poligraficznej...*, *op. cit.*, s. 7.

² *Ibidem*, s. 9.

gruncie orzecznictwa amerykańskiego, kształtowały się zasady dopuszczalności tego typu dowodów w procesie karnym – od powszechnej akceptacji (reguła Frye'a z 1923 r.) – po regułę Dauberta z 1993 r., zakładającą możliwość empirycznej weryfikacji danej metody, znany poziom błędu i stopień akceptacji w środowisku naukowym oraz ustanowione standardy praktyki.

Wartość każdego dowodu podlega ocenie, w tym swobodnej ocenie sądu. Zaczynając od podstaw teoretycznych, poprzez procedury badawcze, a skończywszy na prawidłowości ich zastosowania w analizowanym jednostkowym przypadku, badanie poligraficzne nie ogranicza się do prostego pomiaru zmian reakcji fizjologicznych osoby badanej. Biegły – poligrafer – odpowiada nie tylko za obsługę przyrządu pomiarowego (poligrafu). Spośród jego zadań należy wymienić m.in.: przeprowadzenie wstępnego wywiadu, opracowanie poprawnych metodologicznie pytań testowych, interpretację zarejestrowanych danych testowych i sporządzenie opinii.

Każdy pomiar obarczony jest niepewnością, która może wynikać z niedoskonałości przyrządu, wpływu warunków panujących w otoczeniu, umiejętności badającego czy problemów z reprezentatywnością próby badawczej. W środowisku badań poligraficznych konieczne jest ograniczenie liczby bodźców zewnętrznych (wszelkich dystraktorów) do minimum. Na poprawność pomiaru w badaniach poligraficznych rzutuje również stan poligrafu i oprogramowanie przetwarzające dane rejestrowane przez urządzenie (współcześnie nie wykorzystuje się już praktycznie aparatów analogowych). Wizualną reprezentacją danych surowych są zapisy na poligramach. Dane surowe podlegają filtrowaniu. Autor niniejszej monografii wskazał przykłady takiego filtrowania danych z kanału reakcji elektrodermalnych, które mogłyby wypaczać prawidłową ocenę. Zaleca przy tym stosowanie trybu manualnego bez filtrów lub ich uruchamianie jedynie w niezbędnym zakresie.

Poza krzywymi reakcji organizmu osoby badanej na poligramie muszą zostać oznaczone również inne dane, i jest to rola badającego. Mowa tu m.in. o początku wprowadzenia danego bodźca testowego, czasie jego prezentacji, momencie i rodzaju udzielonej odpowiedzi, a także przypisach odnoszących się do okoliczności niewynikających z zadawanych pytań i naturalnych reakcji na te pytania (względnie inne rodzaje bodźców – np. fotografie).

Prawidłowe wykonanie czynności związanych z obsługą poligrafu oraz innych, składających się na procedurę badania poligraficznego, wymaga ze strony poligrafera posiadania wiadomości specjalnych oraz odpowiednich właściwości osobistych. Poligrafer pozyskuje wiadomości specjalne poprzez specjalistyczne przeszkolenie według wytycznych profesjonalnych organizacji zrzeszających praktyków i naukowców zajmujących się badaniami poligraficznymi. Fakt odbycia stosownego przeszkolenia może być potwierdzony przez te organizacje, alternatywnie przez urzędy państwowe, w których przeprowadza się takie badania, a wtórnie – również przez wpis na listę biegłych sądowych. Nie jest przy tym konieczne, aby poligrafer legitymował się sztywno określonym wykształceniem akademickim – z zakresu psychologii, prawa czy biologii. Szkolenia mają charakter interdyscyplinarny, uwzględniają wybrane aspekty z każdej tych dyscyplin.

Jeżeli chodzi o (ważny przy selekcji kandydatów na szkolenie) optymalny profil psychologiczny *statystycznego* poligrafera – można podsumować na podstawie ustaleń nauki, że jest to osoba z dominującymi cechami ekstrawersji, wykształconymi potrzebami znaczenia i przynależności; o jak najwyższej inteligencji językowej, logicznej i emocjonalnej, a niskim z kolei poziomem neurotyczności; poligrafer powinien być względnie ugodowy, a jednocześnie dostatecznie asertywny; otwarty na nowe doświadczenia oraz ciągłe doskonalenie zawodowe; skrupulatny i zorientowany na zadanie; prezentujący styl poznawczy niezależny od pola percepcji, refleksyjny, wykazujący zdolność do abstrakcyjnego myślenia.

Na proces myślowy każdego biegłego, a więc również poligrafera, oddziałują m.in.: stan emocjonalny, wstępne wyobrażenia na temat przedmiotu badań oraz otoczenie, w jakim biegły funkcjonuje, w szczególności presja ze strony zlecniodawców ekspertyz. Wielość informacji do przetworzenia sprawia, że umysł badającego stosuje uproszczone reguły myślenia, usprawniające podejmowanie decyzji, ale stwarzające przy tym ryzyko błędów poznawczych. W tym kontekście omówiono przykłady występowania heurystyk: dostępności, reprezentatywności, zasygnalizowano występowanie błędów atrybucji w postaci efektów aureoli, diabła, ignorowania różnic indywidualnych czy błędu Otella.

Biegły ocenia znaczenie identyfikacyjne poszczególnych cech, ocenia pewność samej identyfikacji. Doświadczenie i praktyka pozwalają

biegłemu ocenić, czy dana cecha ma duże prawdopodobieństwo wystąpienia. Wprawdzie prawdopodobieństwo jest kategorią obiektywną, jednak wynika z tego, że jego ocena nie jest pozbawiona elementów subiektywizmu³. Nie inaczej jest u poligraferów. Przestrzeń do zaistnienia subiektywizmu istnieje, odkąd zmiany reakcji fizjologicznych osoby badanej trzeba nie tylko zmierzyć, ale również je potem zinterpretować na podstawie zbioru kryteriów przyjętych dla danej metody analitycznej. W zależności od metody biegły ma większy lub mniejszy margines swobody w interpretacji, kiedy nie jest związany obiektywnymi, precyzyjnymi kryteriami.

Przeprowadzone badania empiryczne potwierdzają przyjętą hipotezę, że informacje na temat osoby poddawanej badaniu poligraficznemu, z którymi biegły zapoznaje się przed przeprowadzeniem analizy zarejestrowanych przez poligraf danych, mogą wpływać na przebieg późniejszej interpretacji zapisów. Niezależnie od posiadanego doświadczenia zawodowego biegli sugerują się wcześniej przekazanymi informacjami (kształtującymi wstępny obraz osoby winnej lub niewinnej). Stwarza to ryzyko błędu konfirmacji i rzutuje na oceny numeryczne, lecz tylko wówczas, gdy różnice w zapisach zmian reakcji nie są zbyt oczywiste, wymagają skomplikowanych analiz, a prawidłowy wynik końcowy testu oscyluje na granicy rozstrzygnięcia. W realnych warunkach nie są to jednak przypadki dominujące. Z kolei jednoznaczne dane testowe nie pozostawiają biegłym swobody interpretacji i nie dają możliwości potwierdzenia hipotezy wynikającej z wcześniejszych oczekiwań. Trzeba też zaznaczyć, że opisane zjawisko sugerowania się informacjami przekazanymi przed ekspertyzą nie ujawnia się u poligraferów w jakiś szczególny sposób. Występuje również w innych badaniach kryminalistycznych – w tym tych, które wydają się funkcjonować w powszechnym odbiorze jako bardziej rzetelne (np. badania daktyloskopijne czy analiza mieszaniny DNA⁴).

Dokonaną w 2019 r. nowelizacją polskiego Kodeksu postępowania karnego wprowadzono zmiany w art. 198, który w nowym brzmieniu wyraźnie wskazuje, że organ procesowy ma zdecydować, z jakimi dokumentami zapoznać biegłego, a wcześniejsze opinie biegłych tej samej specjalności należy udostępnić tylko w wyjątkowym wypadku.

³ Por. T. Tomaszewski, *Wartość niekategorycznych opinii biegłych*, „Nowe Prawo”, 9, 1970, s. 71–72.

⁴ Zob. I.E. Dror, G. Hampikian, *Subjectivity and Bias in Forensic DNA Mixture Interpretation*, „Science and Justice”, 51, 2011, s. 204–208.

Zmiany te miały na celu zwiększenie obiektywizmu przy opiniowaniu. Dla biegłego z zakresu badań poligraficznych najważniejsze jest skrupulatne zapoznanie się z okolicznościami rozpatrywanego zdarzenia, wersją przedstawianą przez osobę badaną i inne przesłuchane osoby. Nie ma natomiast potrzeby analizowania i świadomego lub nieświadomego sugerowania się opiniami większości biegłych innych specjalności. Przykładowo w sprawie o pedofilię trudno byłoby zignorować zupełnie stwierdzenie przez biegłego seksuologa zaburzeń preferencji seksualnych u podejrzanego.

Bardziej skomplikowana jest kwestia zapoznania się z ewentualnymi poprzednimi opiniami biegłych tej samej specjalności. Z jednej strony – przekonanie o wysokich lub niskich kwalifikacjach, albo osobista sympatia lub antypatia wobec poprzednika mogą z góry ukierunkowywać postawę solidarności zawodowej bądź wrogości. Z drugiej zaś strony – w przypadku przygotowań do badań powtórnych – nie bez znaczenia jest, jakie pytania były zadawane i jakie techniki badawcze wcześniej zastosowano. W tym kontekście opinia biegłego tej samej specjalności jest więc z zasady przydatna, z wyjątkiem sytuacji, gdy kolejne badanie poligraficzne byłoby wykonywane właśnie z tego względu, że wydana przez poprzedniego biegłego opinia okazała się niepełna, niejasna lub zachodzi sprzeczność w niej samej lub między innymi opiniami w tej samej sprawie. Wówczas, zgodnie z obecnym brzmieniem przywołanego przepisu, takiej opinii nowemu biegłemu słusznie nie udostępnia się.

Optymalnym rozwiązaniem wydaje się zapoznavanie biegłego z zakresu badań poligraficznych z takimi dokumentami, które są niezbędne do wydania opinii, czyli mają związek z celami ekspertyzy – z wyłączeniem opinii biegłych innych specjalności. Wyjątki należałoby uczynić dla ekspertyz medycznych (w szczególności psychiatrycznych) i psychologicznych, ponieważ ewentualne choroby i stan psychiczny mają znaczenie przy podejmowaniu przez poligrafera decyzji o zakwalifikowaniu danej osoby do badania. Z kolei wiedza o cechach osobowości może być przydatna choćby przy doborze odpowiedniego sposobu prowadzenia wywiadu z osobą badaną. Przede wszystkim poligrafer potrzebuje dokumentów opisujących przedmiot sprawy. To mogą być notatki z ustaleniami na temat zaistniałego zdarzenia, protokoły oględzin. Do tego protokoły przesłuchań i niezależnie od fazy postępowania – postanowienie o przedstawieniu zarzutów, aktu oskarżenia. Problem w tym, że selekcja materiałów przez organ

procesowy – nie dość że czasochłonna, też może być obciążona subiektywizmem. W konsekwencji ciężar subiektywizmu jest poniekąd przeniesiony z biegłego na podmiot zlecający ekspertyzę.

W badaniach poligraficznych można wyodrębnić dwie zasadnicze grupy metod analizy danych: jakościowe (czyli globalnej interpretacji zapisów na przestrzeni całego wykresu – bez dodatkowych narzędzi pomiarowych i ilościowego szacowania różnic w wielkości zmian reakcji) oraz metody jakościowo-ilościowe (szeregowanie istotności reakcji według rang lub stosowanie ocen liczbowych oraz reguł decyzyjnych, które prowadzą do opinii będącej konsekwencją obiektywnych obliczeń matematycznych). Te ostatnie, nazywane metodami numerycznymi, dzielą się na skale 3- i 7-pozycyjne. Analiza spoczywa na poligraferze, który przeprowadził badanie, a uzupełniając (np. w ramach konsultacji zespołowych lub procedur kontroli jakości) zadanie to powierza się innym poligraferom, bez udostępniania jakichkolwiek danych poza zapisami reakcji fizjologicznych – nazywamy to wówczas „ślepa interpretacja”. Czynności analityczne wykonuje się manualnie lub automatycznie, komputerowo. Algorytmy całościowo analizujące dane rejestrowane przez poligraf (jak np. *polyscore* czy OSS – *objective scoring system*) zapewniają perfekcyjną rzetelność, lecz nie gwarantują jeszcze takiej dokładności jak w przypadku testów ocenianych przez człowieka. Poza tym, nawet przy ewaluacji komputerowej, część czynności musi zostać zrealizowana manualnie – np. oznaczenie artefaktów na poligramach czy wybór właściwych reguł decyzyjnych. Dlatego algorytmy komputerowe traktowane są jako narzędzia wspomagające proces opiniowania przez biegłego.

Autor niniejszej monografii przedstawił wszystkie najważniejsze metody interpretacji danych w badaniach poligraficznych. Ich ewolucja przebiegała od metod jakościowych w kierunku ilościowych (liczbowego szacowania różnic w reakcjach na pytania krytyczne i kontrolne). Badania empiryczne nie wykazały istotnych statystycznie różnic między nimi, jeżeli chodzi o kryterium dokładności. Nie potwierdziła się zatem hipoteza mówiąca o tym, że metody numeryczne prowadzą do dokładniejszych rezultatów (poprawnych diagnoz). Jeżeli chodzi o kryterium subiektywizmu – odzwierciedlone w stopniu zgodności między oceniającymi – założenie, że metody numeryczne pozostawiają mniejszy margines swobody interpretacji niż metody jakościowe, okazało się trafne tylko w części – w odniesieniu

do metody numerycznej w skali 3-pozycyjnej pod nazwą Empiryczny System Oceniania (ESS). Przy metodzie globalnej analizy zapisów na poligramach zaistniała większa rozbieżność diagnoz, ale podobnie było przy numerycznej skali 7-pozycyjnej systemu Utah (w przypadku tych dwóch ostatnich metod współczynnik zgodności – alfa Krippendorpha – okazał się jednakowy). W tym kontekście zasadne byłoby jeszcze zbadanie w przyszłości znaczenia doświadczenia zawodowego. Intuicyjnie wydaje się, że doświadczenie może być szczególnie istotne przy metodzie jakościowej – im większa liczba przypadków, z którymi zetknie się biegły (licniejszy zasób w pamięci wzorcowych zapisów reakcji) – tym większa szansa na skuteczną identyfikację.

Pozytywnie zweryfikowana została hipoteza zakładająca, że spośród uznanych metod analizy numerycznej nie ma istotnych różnic pod względem końcowych rezultatów uzyskiwanych w testach z pytaniami porównawczymi (typu CQT), przy czym ogółem trafniej identyfikowano osoby niewinne (prawdomówne). Udowodniona została także hipoteza wskazująca, że metody numerycznej analizy zapisów na poligramach o węższej skali (3-pozycyjnej) są bardziej obiektywne niż te o skalach szerszych (7-pozycyjnych). Kiedy poligraferzy dysponowali szerszymi możliwościami punktowania ocen zmian reakcji fizjologicznych przy pytaniach krytycznych – rzetelność testu spadała.

Również prawdziwa okazała się hipoteza, że to oceniający metodą *ślepej interpretacji* diagnozują z większą dokładnością niż eksperci, którzy osobiście przeprowadzili badania poligraficzne. Wylimowanie subiektywnych okoliczności, wynikających przede wszystkim z bezpośredniej interakcji z osobą badaną oraz wpływu wcześniejszych informacji o tej osobie i przedmiocie badania, wydaje się zatem korzystne dla procesu analitycznego. W konsekwencji być może należałoby rozważyć wprowadzenie ślepej interpretacji jako obowiązkowego uzupełnienia procedury badawczej, a więc opiniowania zespołowego, z pozostawieniem prawa do ostatecznej decyzji ekspertowi wiodącemu, odpowiadającemu za całokształt przeprowadzonych czynności, w tym tych z udziałem osoby badanej.

Większość z dotychczas opisywanych w literaturze badań nad dokładnością testów pytań porównawczych w rzeczywistych sprawach (poza laboratorium eksperymentalnym) sugerowało, że to poligraferzy osobiście przeprowadzający testy mogą uzyskiwać lepsze rezultaty. Dokładniejsi okazywali się zwłaszcza w identyfikacji osób niewinnych.

Trzeba jednak zaznaczyć, że poglądy te kształtowały się jeszcze w czasach, kiedy szeroko stosowano jakościowe metody interpretacji zapisów na poligramach. Współcześnie zaś dominują metody numeryczne i trend najwyraźniej uległ odwróceniu.

Metody numeryczne, w szczególności system ESS i jego najnowsza wersja (ESS-M), umożliwiają statystyczne zaprezentowanie konkretnego wyniku liczbowego testu – poprzez posłużenie się wartością p lub kategorią szansy. Dzięki temu wiadomo, jaki stopień pewności należy przypisać określonej diagnozie, jaki margines błędu istnieje (z powodu czynników subiektywnych i innych przyczyn zakłóceń miarodajności testów). Jest to niewątpliwa zaleta i trudno przy tym zrozumieć – dyskryminujące względem innych metod – oczekiwanie absolutnej dokładności testów poligraficznych, wyrażane przez część środowiska prawniczego. Nie można wobec jakiegokolwiek ekspertyzy stawiać tak wygórowanych wymagań jak wobec całokształtu czynności składających się na postępowanie dowodowe⁵.

Wśród sposobów na ograniczenie marginesu subiektywizmu w badaniach poligraficznych autor tejże monografii wskazuje na walidację i standaryzację testów, ujednoczenie procedur badawczych, doprecyzowanie kryteriów diagnostycznych oraz politykę zapewnienia jakości. Omówiono znaczenie badań walidacyjnych, potwierdzających trafność i rzetelność metod wykorzystywanych w badaniach poligraficznych. Rozważano kilka rodzajów trafności testu (fasadową, konstrukcji, kryterialną, wewnętrzną i zewnętrzną). Uogólniając – trafność wskazuje, czy test mierzy to, co ma mierzyć. Z kolei rzetelność świadczy o tym, czy test konsekwentnie mierzy tożsame właściwości; na ile powtarzalne są wyniki badań tej samej osoby i jak zgodni są ze sobą biegli analizujący jednakowy materiał badawczy. To właśnie ten wskaźnik sygnalizuje skalę subiektywizmu w metodzie.

Margines swobody interpretacyjnej w badaniach poligraficznych z pewnością byłby węższy, gdyby maksymalnie doprecyzowano i ujednoczono cechy diagnostyczne wykorzystywane przy analizie zapisów na poligramach oraz kryteria przyznawania ocen numerycznych. W początkowej fazie rozwoju badań poligraficznych poszukiwano jak największej liczby przesłanek diagnozowania. Potem zbiór ten ulegał systematycznemu ograniczeniu w wyniku odrzucenia tych cech, które nie zostały odpowiednio potwierdzone empirycznie. Inne

⁵ Por. T. Widła, *Ocena dowodu...*, *op. cit.*, s. 47.

upraszczano – kilka cech kumulowano w jedną (jak np. różne formy tłumienia oddychania). Mimo to jeszcze w pierwszych latach XXI wieku w systemie analizy stosowanym przez amerykańskie instytucje federalne obowiązywało ponad 20 różnych diagnostycznych cech zapisów. Współcześnie w trzech głównych systemach (ESS, Utah, federalny) mamy ich od 6 do 12. Odpowiedź na nasuwające się przy tym pytanie – o potencjalne konsekwencje zmian w regułach analitycznych – wymagałaby właściwie dalszych, odrębnych badań. Po pierwsze jednak – pamiętajmy, że badanie poligraficzne dostarcza dowodu pośredniego i samoistnie nie przesądza o winie lub niewinności. Istnieją ponadto przesłanki wspierające taką hipotezę, która wskazywałaby, że teoretycznie można sobie wyobrazić rozbieżność rezultatów przy porównaniu różnych zestawów cech diagnostycznych i reguł decyzyjnych, lecz w praktyce wydaje się to mało prawdopodobne. Wartość diagnostyczna badań przeprowadzanych w celach dowodowych, przy różnych metodach na przestrzeni lat, nie odbiegała od poziomu oscylującego wokół 90%. Końcowy wynik testu jest sumą wielu składowych. Podobnie jak w przypadku błędu konfirmacji ryzyko nieprawidłowego wyniku przy zmianie pojedynczych kryteriów oceny zapisów występuje realnie tylko w tych przypadkach, kiedy wynik ten jest bliski progu oznaczającego brak rozstrzygnięcia. Ponadto zmiany dotyczyły głównie oddechowego kanału zapisu, którego wkład w ogólną wartość decyzyjną jest relatywnie najniższy względem pozostałych parametrów.

Także w innych obszarach kryminalistyki sięga się po nowocześniejsze metody przy badaniu tych samych śladów. Znane są przypadki, kiedy np. ślad biologiczny – badany już przed laty bez pożądanego efektu – był wysyłany do ponownej analizy, która współcześnie pozwoliła już na skuteczną identyfikację. Wreszcie – jak w każdej innej dziedzinie – ci, którzy chcą posługiwać się metodą naukową, ograniczeni są tym, na co pozwala aktualny stan nauki. Przykładem może być zmienność zaleceń dotyczących resuscytacji krążeniowo-oddechowej. Czy świadomość tych korekt miałaby spowodować w nas zawahanie przed zastosowaniem tej techniki w razie potrzeby?

Czy zmiany w metodologii badań mogą uzasadniać wzruszenie wyroków w sprawach, w których takie badania były wykorzystywane dowodowo? W ocenie autora niniejszej monografii – tylko w przypadku diametralnej zmiany standardów danego typu badań i tylko, gdy

uzyskany na ich podstawie dowód miał kluczowe znaczenie w sprawie. W badaniach poligraficznych zmiany w przesłankach diagnozowania nie miały charakteru rewolucyjnego, a status opinii biegłego tej specjalności na pewno nie był taki, że ten jeden dowód zdecydował o wyroku, choć mógł być języczkiem u wagi – dlatego każda sprawa wymagałaby indywidualnego rozpatrzenia.

Ważne, aby poligrafer postępował ściśle według reguł obranej metody badawczej. Nie powinien wybiórczo czerpać z różnych systemów analizy danych ani stawiać własnych intuicyjnych ustaleń ponad standaryzowane procedury. Aby zapewnić postępowanie biegłych według standardów profesjonalnych, autor niniejszej monografii zwrócił uwagę na potrzebę stosowania mechanizmów kontroli jakości (m.in. poprzez wspomnianą już wcześniej ślepą interpretację zapisów na poligramach, sprawdzenie w dokumentacji poprawności sformułowania pytań testowych, obserwację przebiegu badania dzięki zapisom audio-wizualnym) oraz rozwiązań na poziomie instytucjonalnym, włącznie z licencjonowaniem. Niezbędny wydaje się również urzędowy wykaz dokumentów potwierdzających kwalifikacje poligraferów, wydany np. w formie rozporządzenia właściwego ministra, a uzgodniony z przedstawicielami profesji i ośrodków kształcenia.

Procesy w kierunku standaryzacji, zapewnienia jakości nieustannie postępują, ale są i tacy poligraferzy (np. wybitny przedstawiciel profesji N. Gordon), którzy przekonują, że dążenie części środowiska biegłych do pełnego unaukowania metod badawczych jest nierealne, ponieważ badania poligraficzne były i pozostaną sztuką. N. Gordon pisze:

Jako profesja powinniśmy kontynuować badania naukowe, aby potwierdzić sztukę, którą praktykujemy. Nie możemy pozwolić poprowadzić się ścieżką, która mylnie sprawi, że uwierzymy, iż to, co robimy jest czystą nauką, co w końcu będzie wymagało, aby posiadać doktoraty. Umiejętności w sztuce są tak samo ważne jak naukowe aspekty związane z poligrafem⁶.

Trudno nie zgodzić się z tym poglądem w tym sensie, że nie jest możliwa pełna standaryzacja procedury badania poligraficznego (jak i innych, których nieodłączny element stanowi interakcja badającego

⁶ N.J. Gordon, *A Field Polygraph Examination. Science or Art?*, „European Polygraph”, 10, 3(37), 2016, s. 109 [tłum. M.G.].

z osobą badaną – tak jest w badaniach poligraficznych i np. psychologicznych, jeśli obejmują wywiad bezpośredni, nawet jeśli byłby on ustrukturyzowany). Dzieje się tak po części ze względu na różnice indywidualne między osobami badanymi, np. w poziomie inteligencji, w kompetencjach językowych. Nie można mówić w ten sam sposób do każdego. Poza rutynowymi procedurami kadrowymi badania w konkretnych sprawach, np. kryminalnych, dotyczą nieidentycznych zdarzeń, a zatem pytania testowe muszą być indywidualnie dostosowywane. Co więcej – sposób reagowania na bodźce testowe może być u danej jednostki nietypowy, a mimo to badający jest w stanie zaobserwować pewne prawidłowości, symptomatyczne reakcje organizmu.

Dopóki na powodzenie tego rodzaju ekspertyzy wpływ będą miały m.in. umiejętności badającego w zakresie prowadzenia wywiadu i formułowania zindywidualizowanych pytań testowych, zdolności percepcyjne w interpretacji zapisów na poligramach czy sposób, w jaki poligrafer radzi sobie z napływem informacji towarzyszących powierzonymu zadaniu – dopóty badanie poligraficzne rzeczywiście będzie zawierało elementy sztuki i nieodłączny im subiektywizm. Jednocześnie wcale nie przekreśla to idei *scientific evidence*. W obecnym wieku obserwuje się stały trend w kierunku standaryzacji jak największej części procedury badania poligraficznego, a warunek naukowego potwierdzenia stosowanych metod ma charakter konieczny, a już na pewno gdy badanie odbywa się na potrzeby dowodowe. Niegdyś mówiło się, że badanie poligraficzne to 50 procent nauki i 50 procent sztuki. Obecnie proporcje te zmieniały się zdecydowanie na korzyść nauki.

Bibliografia

- Adam C., *Forensic Evidence in Court. Evaluation and Scientific Opinion*, Oxford 2016.
- Amsel T.T., *Examiner Approach and its Impact on Polygraph Results*, „European Polygraph”, 10, 1(35), 2016, s. 7–10.
- Amsel T.T., *Practicing Polygraph. Best Practice Guide*, North Charleston 2017.
- Ana Montes. *Cuban Spy*, www.fbi.gov/history/famous-cases/ana-montes-cuba-spy [dostęp: 4.06.2020].
- Ansley N., *Development of Deception Criteria Prior to 1950*, „Polygraph”, 28(1), 1999, s. 28–36.
- APA [American Polygraph Association], *APA Continuing Education Hours Policy*, www.polygraph.org/policy-for-apa-continuing-education-hours [dostęp: 10.06.2020].
- APA, *APA Model Policy for Quality Assurance*, www.polygraph.org/assets/docs/Misc.Docs/QA%20Model%20Policy%20%28approved%207-20-2017%29.pdf [dostęp: 29.03.2021].
- APA, *By-laws*, 27.08.2019, www.polygraph.org/assets/docs/Misc.Docs/BYLAWS%20adopted%208-27-19.pdf [dostęp: 10.06.2020].
- APA, *Education Accreditation Committee. Accreditation Standards (version 2.05)*, 23.09.2019, www.apapolygraph.org/accredited-programs [dostęp: 10.06.2020].
- APA, *Meta-Analytic Survey of Criterion Accuracy of Validated Techniques*, „Polygraph”, 40(4), 2011, s. 193–305.
- APA, *Model Policy for Law Enforcement/Public-service Pre-employment Polygraph Screening examinations*, apoa.memberclicks.net/assets/docs/le%20screening%20model.pdf [dostęp: 9.06.2020].
- APA, *Model Policy for Release and Management of Polygraph Reports and Polygraph Data*, apoa.memberclicks.net/assets/docs/records%20and%20data%20model%20policy.pdf [dostęp: 9.06.2020].

- APA, *PCSOT Committee Operational Policy*, 04.2016, www.polygraph.org/assets/docs/Misc.Docs/pcsot%20operational%20policy%20adopted%205-17-16.pdf [dostęp: 8.06.2020].
- APA, *Standards of Practice*, 23.08.2019, www.polygraph.org/assets/docs/Misc.Docs/standard/adopted%208-23-19%20APA%20STANDARDS%20OF%20PRACTICE%20.pdf [dostęp: 9.06.2020].
- ASTM International, *Standard Guide for Instrumentation, Sensors and Operating Software Used in Forensic Psychophysiological Detection of Deception (Polygraph) Examinations: E2439-09(2016)*, www.astm.org/Standards/E2439.htm [dostęp: 9.06.2020].
- ASTM International, *Standard Guide for Minimum Continuing Education and Training of Individuals Involved in the Detection of Deception (PDD). E2064-00(2014)*, www.astm.org/Standards/E2064.htm [dostęp: 10.06.2020].
- ASTM International, *Standard Guide for PDD Examination Standards of Practice. E2062-07*, www.astm.org/Standards/E2062.htm [dostęp: 9.06.2020].
- ASTM International, *Standard Practice for Calibration and Functionality Checks Used in Forensic Psychophysiological Detection of Deception (Polygraph) Examinations. E2063-12(2017)*, www.astm.org/Standards/E2063.htm [dostęp: 9.06.2020].
- Backster C., *Standardized Polygraph Notepack and Technique Guide*, New York 1969.
- Baesen H.V., Chung C., Yang C., *A Lie Detector Experiment*, „Journal of Criminal Law, Criminology, and Police Science”, 39, 1948–1949, s. 532–537.
- Barland G.H., *Detection of Deception in Criminal Suspects. A Field Validation Study*, Salt Lake City, 1975.
- Barland G.H., *The Reliability of Polygraph Chart Evaluations*, „Polygraph”, 1(4), 1972, s. 192–206.
- Barland G.H., Raskin D.C., *An Experimental Study of Field Techniques in „Lie Detection”*, „Psychophysiology”, 9, 1972, s. 275.
- Barrick M.R., Mount M.K., *The Big Five Personality Dimensions and Job Performance. A Meta-Analysis*, „Personnel Psychology”, 44, 1991, s. 1–26.
- Bell B.G., Grubin D., *Functional Magnetic Resonance Imaging May Promote Theoretical Understanding of the Polygraph Test*, „The Journal of Forensic Psychiatry & Psychology”, 21(1), 2010, s. 52–65.

- Bell B.G., Raskin D.C., Honts C.R., Kircher J.C., *The Utah Numerical Scoring System*, „Polygraph”, 28(1), 1999, s. 1–9.
- Bell S., *A Beginner's Guide to Uncertainty of Measurement*, Teddington 2001.
- Bell S., *Measurement Uncertainty in Forensic Sciences. A Practical Guide*, London–New York 2017.
- Benussi V., *Die Atmung Asymptome der Lüge (The Respiratory Symptoms of Lying)*, „Archiv für die gesamte Psychologie”, 11, 1914, s. 244–273.
- Ben-Shakhar G., Bar-Hillel M., Lieblich I., *Trial by Polygraph. Scientific and Juridical Issues in Lie Detection*, „Behavioral Sciences”, 4(4), 1986, s. 459–479.
- Bersh P., *A Validation Study of Polygraph Examiner Judgments*, „Journal of Applied Psychology”, 53, 1969, s. 399–403.
- Blackwell N.J., *PolyScore 3.3 and Psychophysiological Detection of Deception Examiner Rates of Accuracy When Scoring Examinations from Actual Criminal Investigations*, „Polygraph”, 28(2), 1999, s. 149–175.
- Blackwell N.J., Department of Defense Polygraph Institute, *An Evaluation of the Effectiveness of the Polygraph Automated Scoring system (PASS) in Detecting Deception in a Mock Crime Analog Study (Report No. DoDPI94-R0003)*, 1994, apps.dtic.mil/sti/pdfs/ADA305755.pdf [dostęp: 9.06.2020].
- Blackwell N.J., Department of Defense Polygraph Institute, *Polyscore. A Comparison of Accuracy (Report No. DoDPI94-R0003)*, Fort McClellan 1996.
- Blalock B., Cushman B., Nelson R., *A Replication and Validation Study on an Empirically Based Manual Scoring System*, „Polygraph”, 38(4), 2009, s. 281–286.
- Bocheński J., *Sto zabobonów. Krótki filozoficzny słownik zabobonów*, Kraków 1994.
- Boratyńska K.T., *Dział V. Dowody*, [w:] A. Sakowicz (red.), *Kodeks postępowania karnego. Komentarz*, wyd. 7, Warszawa 2016, s. 468–650.
- Congress of the United States: Office of Technology Assessment, *Scientific Validity of Polygraph Testing. A research review and evaluation – a technical memorandum*, Washington 1983.
- Cook Ch.M., *The Role and Rights of the Expert Witness*, „Journal of Forensic Sciences”, 9(4), 1964, s. 456–460.
- Crewson P.E., *A Comparative Analysis of Polygraph with Other Screening and Diagnostic Tools*, „Polygraph”, 32(2), 2003, s. 57–85.

- Defense Academy for Credibility Assessment, *Assessment of Optimal Interrogation Approaches*, 2007, apps.dtic.mil/sti/pdfs/ADA476369.pdf [dostęp: 27.06.2020].
- Defense Academy for Credibility Assessment, *Assessment of Psychophysiological Detection of Deception (PDD) Pretest*, 2007, apps.dtic.mil/dtic/tr/fulltext/u2/a483331.pdf [dostęp: 27.06.2020].
- Department of Defense Polygraph Institute, *DoDPI General Questions Test (GQT)*, 1995.
- Department of Defense Polygraph Institute, *Federal Psychophysiological Detection of Deception Examiner Handbook*, [2006].
- Department of Defense Polygraph Institute, *Test Data Analysis. DoDPI Numerical Evaluation Scoring System*, 2004.
- Dollins A.B., Krapohl D.J., Dutton D.W., *Computer Algorithm Comparison*, „Polygraph”, 29(3), 2000, s. 237–247.
- Dror I.E., Charlton D., Pèron A.E., *Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications*, „Forensic Science International”, 156, 2006, s. 74–78.
- Dror I.E., Hampikian G., *Subjectivity and Bias in Forensic DNA Mixture Interpretation*, „Science and Justice”, 51, 2011, s. 204–208.
- Dror I.E., Rosenthal R., *Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts*, „Journal of Forensic Sciences”, 53(4), 2008, s. 900–903.
- Edel E.C., Jacoby J., *Examiner Reliability in Polygraph Chart Analysis. Identification of Physiological Responses*, „Journal of Applied Psychology”, 60(5), 1975, s. 632–634.
- Eggleston R., *Sixth Wilfred Fullagar Memorial Lecture. Beyond Reasonable Doubt*, „Monash University Law Review”, 4(1), 1977, s. 1–22.
- Ekman P., *Kłamstwo i jego wykrywanie w biznesie, polityce i małżeństwie*, tłum. S.E. Draheim, M. Kowalczyk, Warszawa 2015.
- Elaad E., Ginton A., Ben-Shakhar G., *The Effects of Prior Expectations and Outcome Knowledge on Polygraph Examiners' Decisions*, „Journal of Behavioral Decision Making”, 7, 1994, s. 279–292.
- Elaad E., Ginton A., Ben-Shakhar G., *The Role of Prior Expectations in Polygraph Examiners Decisions*, „Psychology, Crime & Law”, 4, 1998, s. 1–16.
- Elfenbein H.A., Marsh A.A., Ambady N., *Emotional Intelligence and the Recognition of Emotion from Facial Expressions*, [w:] L.F. Barrett, P. Salovey (eds), *The Wisdom in Feeling. Psychological Processes in Emotional Intelligence*, New York 2002, s. 37–59.

- Enos F., Benus S., Cautin R.L., Graciarena M., Hirschberg J.B., Shriberg E., *Personality Factors in Human Deception Detection. Comparing Human to Machine Performance*, „Interspeech”, 2006, s. 1–4.
- Faigman D.L., Porter E., Saks M.J., *Check Your Crystal Ball at the Courthouse Door, Please. Exploring the Past, Understanding the Present, and Worrying About the Future of Scientific Evidence*, „Cardozo Law Review”, 15, 1994, s. 1799–1835.
- Fakt naukowy* [hasło], sjp.pwn.pl/sjp/fakt-naukowy;2458324.html [dostęp: 13.05.2020].
- Fita S., *Słownik metrologiczny*, www.metrologia.pwr.edu.pl/pliki/slownik_metrologiczny.pdf [dostęp: 18.05.2020].
- Försterling F., *Atrybucje. Podstawowe teorie, badania i zastosowanie*, tłum. J. Radzicki, Gdańsk 2005.
- Fundacja Europejskie Centrum Inicjatyw w Naukach Sądowych, *Ocena kompetencji biegłych sądowych. Oczekiwania i rekomendacje*, Warszawa 2015, forensicwatch.pl/web/pliki/baza-wiedzy/Opracowania/Ocena-Kompetencji-Bieglych-Sadowych.pdf [dostęp: 18.05.2020].
- Gastwirth J.L. (ed.), *Statistics for Social Science and Public Policy*, New York 2000.
- Ginton A., *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, 2019, s. 1–6.
- Ginton A., *The Importance of the Consistency Factor in CQT and Other Polygraph Tests*, „Polygraph”, 42(3), 2013, s. 146–162.
- Godby R.E., *Polygraph Examinations With Interpreters*, „Polygraph”, 21(2), 1992, s. 83–91.
- Gołaszewski M., *Badania poligraficzne w polskim procesie karnym po Postanowieniu Sądu Najwyższego z dnia 29 stycznia 2015 r.*, „Przegląd Bezpieczeństwa Wewnętrznego”, 13(7), 2015, s. 139–150.
- Gołaszewski M., *Badania poligraficzne w śledztwach – fakty i mity*, „Przegląd Policyjny”, 2(118), 2015, s. 135–153.
- Gołaszewski M., *Przesłuchanie biegłego z zakresu badań poligraficznych (wariograficznych) w postępowaniu jurysdykcyjnym – tryb, najczęściej zadawane pytania i prawidłowe odpowiedzi*, [w:] K. Banasik, A. Kargol, A. Kubiak-Cyruł, M. Lubelski, E. Plebanek, A. Strzelec (red.), *W poszukiwaniu prawdy. Rozważania o prawie, historii i sprawiedliwości*, Kraków 2018, s. 269–298.

- Gołaszewski M., *Psychologiczne podstawy efektywności badań poligraficznych – nowe podejście w XXI w.*, „Studia Prawnicze. Rozprawy i Materiały”, 2(11), 2012, s. 101–112.
- Gołaszewski M., *Użyteczność badań poligraficznych w ograniczaniu recydywy wśród sprawców przestępstw przeciwko wolności seksualnej*, „Prokuratura i Prawo”, 12, 2016, s. 5–22.
- Gołaszewski M., *Validated Techniques and Scoring Models for PDD Test Data Analysis – Conclusions from the 2011 APA Report*, „European Polygraph”, 6, 4(22), 2012, s. 227–240.
- Gołaszewski M. (red.), *Współczesne standardy badań poligraficznych*, Warszawa 2013.
- Gołaszewski M. et al., *Metodyka przeprowadzania testu Utah ZCT*, [w:] J. Widacki (red.), *Kierunki rozwoju instrumentalnej i nieinstrumentalnej detekcji kłamstwa. Problemy kryminalistyczne, etyczne i prawne*, Kraków 2018, s. 70–74.
- Gołaszewski M., Markowicz P., Jastrzębska J., *Postrzeżenie badań poligraficznych w Polsce oraz ich użyteczność w służbach i innych obszarach działalności*, „Przegląd Bezpieczeństwa Wewnętrznego”, 18, 2018, s. 151–167.
- Gołaszewski M., Widacki M., *Aktualny standard badań poligraficznych a praktyka polska*, [w:] J. Widacki (red.), *Badania poligraficzne w Polsce*, Kraków 2014, s. 151–218.
- Gołaszewski M., Zając P., Widacki J., *Thermal Vision as a Method of Detection of Deception. A Review of Experiences*, „European Polygraph”, 9, 1(31), 2015, s. 5–24.
- Goodenough D.R., *Field Dependence*, [w:] H. London, J.E. Exner (eds), *Dimensions of Personality*, New York 1978, s. 165–216.
- Gordon N.J., *A Field Polygraph Examination. Science or Art?*, „European Polygraph”, 10, 3(37), 2016, s. 103–109.
- Gordon N.J., Cochetti C., *The Horizontal Scoring System*, „Polygraph”, 16(2), 1987, s. 116–125.
- Gougler M., Nelson R., Handler M., Krapohl D., Shaw P., Bierman L., *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), 2011, s. 194–305.
- Góralczyk G., *Ekonomiczna analiza prawa jako wyraz zmiany tradycyjnego paradygmatu*, [w:] A. Samonek, *Teoria prawa. Między nowoczesnością a ponowoczesnością*, Kraków 2013, s. 161–172.

- Grubin D., Kamenskov M., Dwyer R., Stephenson T., *Post-conviction Polygraph Testing of Sex Offenders*, „International Review of Psychiatry”, 31(2), 2019, s. 141–148.
- Gruza E., *Granice opiniowania biegłego*, [w:] Z. Kegel (red.), *Problemy dowodu z dokumentu*, Wrocław 1999, s. 75–81.
- Handler M., *Utah Probable Lie Comparison Test*, „Polygraph”, 35(3), 2006, s. 139–148.
- Handler M., Nelson R., *Utah Approach to Comparison Question Polygraph Testing*, „Polygraph”, 38(1), 2009, s. 15–30.
- Handler M., Nelson R., Goodson W., Hicks M., *Empirical Scoring System. A Cross-cultural Replication and Extension Study of Manual Scoring and Decision Policies*, „Polygraph”, 39(4), 2010, s. 200–215.
- Hasel L.E., Kassin S.M., *On the Presumption of Evidentiary Independence. Can Confessions Corrupt Eyewitness Identifications?*, „Psychological Science”, 20, 2009, s. 122–126.
- Holmes W.D., *The Degree of Objectivity in Chart Interpretation*, [w:] V.A. Leonard, *Academy Lectures on Lie-detection*, vol. 2, Springfield 1957, s. 67–70.
- Honts C.R., *Criterion Development and Validity of the Control Question Test in Field Application*, „Journal of General Psychology”, 123, 1996, s. 309–324.
- Honts C.R., Driscoll L.N., *A Field Validity Study of the Rank Order Scoring System (ROSS) in Multiple Issue Control Question Tests*, „Polygraph”, 17(1), 1988, s. 1–15.
- Honts C.R., Driscoll L.N., *An Evaluation of the Reliability and Validity of Rank Order and Standard Numerical Scoring of Polygraph Charts*, „Polygraph”, 16(4), 1987, s. 241–257.
- Honts C.R., Handler M., Shaw P., Gougler M., *The Vasomotor Response in the Comparison Question Test*, „Polygraph”, 44(1), 2015, s. 62–71.
- Honts C.R., Raskin D.C., *A Field Study of the Directed Lie Control Question*, „Journal of Police Science and Administration”, 16(1), 1988, s. 56–61.
- Honts C.R., Raskin D.C., Kircher J.C., *Effects of Physical Countermeasures and their Electromyographic Detection during Polygraph Tests for Deception*, „Psychophysiology”, 1, 1987, s. 241–247.
- Honts C.R., Raskin D.C., Kircher J.C., Hodes R.L., *Effects of Spontaneous Countermeasures on the Physiological Detection of Deception*, „Journal of Police Science and Administration”, 16(2), 1988, s. 91–94.

- Honts C.R., Reavy R., *The Comparison Question Polygraph Test. A Contrast of Methods and Scoring*, „Physiology & Behavior”, 143, 2015, s. 15–26.
- Horvath F.S., *The Effect of Selected Variables on Interpretation of Polygraph Records*, „Journal of Applied Psychology”, 62(2), 1977, s. 127–136.
- Horvath F.S., *Verbal and Non-verbal Clues to Truth and Deception during Polygraph Examinations*, „Journal of Police Science and Administration”, 1(2), 1973, s. 138–152.
- Horvath F.S., *Verbal and Non-verbal Clues to Truth and Deception during Polygraph Examinations* [praca magisterska obroniona na Michigan State University], 1972, d.lib.msu.edu/etd/15283/datastream/OBJ/download/Verbal_and_nonverbal_clues_to_truth_and_deception_during_polygraph_examinations.pdf [dostęp: 11.01.2021].
- Horvath F.S., Reid J., *The Reliability of Polygraph Examiner Diagnosis of Truth and Deception*, „Journal of Criminal Law, Criminology and Police Science”, 62(1), 1971, s. 276–281.
- Hunter F.L., Ash P., *The Accuracy and Consistency of Polygraph Examiners' Diagnoses*, „Journal of Police Science and Administration”, 1(3), 1973, s. 370–371.
- Ibek A., Gołaszewski M., *Formułowanie opinii z badań poligraficznych i tworzenie argumentów dowodowych*, [w:] M. Gołaszewski (red.), *Współczesne standardy badań poligraficznych*, Warszawa 2013, s. 77–87.
- Inbau F.E., *Lie Detection and Criminal Interrogation*, Baltimore 1942.
- Inbau F.E., *Lie Detection and Criminal Interrogation*, ed. 2, Baltimore 1948.
- Inbau F.E., *Scientific Evidence in Criminal Cases*, „Journal of Criminal Law and Criminology”, 24(6), 1934, s. 1140–1158.
- Inbau F.E., *Some Avoidable Lie-Detector Mistakes*, „Journal of Criminal Law and Criminology”, 40(6), 1950, s. 791–795.
- Jaegermann K., *Opiniowanie sądowo-lekarskie. Eseje o teorii*, Warszawa 1991.
- Jaegermann K., Marek Z., *Rozbieżności w opiniach sądowo-lekarskich*, „Archiwum Medycyny Sądowej i Kryminologii”, 29(4), 1979, s. 249–257.
- Janecki J., *Rozrzut rocznych wyników badania glukozy z 21 laboratoriów polskich i wynikające z niego wnioski dotyczące wdrożenia*

- ogólnokrajowych norm glukozy we krwi, „Diagnostyka Laboratoryjna”, 47(1), 2011, s. 77–83.
- Jaworski R., *Opinia z ekspertyzy poligraficznej jako dowód odciążający*, Wrocław 1999.
- Jayne B.C., *Contributions of Physiological Recordings in the Polygraph Technique*, „Polygraph”, 19(2), 1990, s. 105–117.
- Joling R.J., *Legal Commentaries on Forensic Scientist in Court*, „Journal of Forensic Sciences”, 8(3), 1963, s. 339–353.
- Jordan H.W., *Evidence – Admissibility of Deception („Lie Detector”) Tests*, „Journal of Criminal Law and Criminology”, 29, 1938, s. 287–291.
- Kahn J., Nelson R., Handler M., *An Exploration of Emotion and Cognition during Polygraph Testing*, „Polygraph”, 38(3), 2009, s. 184–197.
- Kassin S.M., Dror I.E., Kukucka J., *The Forensic Confirmation Bias. Problems, Perspectives, and Proposed Solutions*, „Journal of Applied Research in Memory and Cognition”, 2(1), 2013, s. 42–52.
- Kazimierska D., *Wartość naukowa metod wykorzystywanych w identyfikacji*, „Prokuratura i Prawo”, 11, 2016, s. 111–129.
- Keeler L., *A Method for Detecting Deception*, „American Journal of Police Science”, 1(1), 1930, s. 38–51.
- Kenny A., *The Psychiatric Expert in Court*, „Psychological Medicine”, 14(2), 1984, s. 291–302.
- Kircher J.C., Raskin D.C., *Computer Methods for the Psychophysiological Detection of Deception*, [w:] M. Kleiner (ed.), *Handbook of Polygraph Testing*, San Diego 2002, s. 287–326.
- Kircher J.C., Raskin D.C., *Human versus Computerized Evaluations of Polygraph Data in a Laboratory Setting*, „Journal of Applied Psychology”, 73, 1988, s. 291–302.
- Kircher J.C., Kristjansson S.D., Gardner M.K., Webb A., *Human and Computer Decision-making in the Psychophysiological Detection of Deception (Final Report. DoDPI02-R-0001)*, University of Utah, 2005.
- Kleiner M., *Handbook of Polygraph Testing*, San Diego 2002.
- Knopp K.A., *Czy kobiety naprawdę są bardziej inteligentne emocjonalnie niż mężczyźni? O różnicach międzypłciowych w zakresie zdolności emocjonalnych*, „Fides et Ratio”, 4(12), 2012, s. 95–112.
- Konieczny J., *Badania poligraficzne. Podręcznik dla zawodowców*, Warszawa 2009.
- Kopeć M., Terelak J.F., *Style poznawcze a style radzenia sobie ze stresem u osób z przewlekłym bólem kręgosłupa*, „Studia Psychologica”, 9, 2009, s. 57–74.

- Kopparumsolan V., *The Effects of Examiner Personality Variables and Training on the Accuracy of Detection of Deception*, Singapore 2018, dr.ntu.edu.sg/bitstream/10356/73352/1/FINAL%20.pdf [dostęp: 28.06.2020].
- Krapohl D.J., *Short Report Update for the Objective Scoring System*, „Polygraph”, 31(4), 2002, s. 298–302.
- Krapohl D.J., Dutton D.W., *Believing Is Seeing. The Influence of Expectations on Blind Scoring of Polygraph Data*, „Polygraph”, 47(2), 2018, s. 91–107.
- Krapohl D.J., Dutton D.W., Ryan A.H., *The Rank Order Scoring System, Replication and Extension with Field Data*, „Polygraph”, 30(3), 2001, s. 172–181.
- Krapohl D., Goodson W., *Decision Accuracy for the Relevant-Irrelevant Screening Test. Influence of an Algorithm on Human Decision-Making*, „European Polygraph”, 9(34), 2015, s. 189–208.
- Krapohl D., McCloughan J.B., Senter S.M., *How to Use the Concealed Information Test*, „Polygraph”, 38(1), 2009, s. 34–49.
- Krapohl D., McManus B., *An Objective Method for Manually Scoring Polygraph Data*, „Polygraph”, 28(3), 1999, s. 209–222.
- Krapohl D.J., Nix K., *Evidence and Common Sense. Suggestions for Scoring Electrodermal Responses*, „APA Magazine”, 52(1), 2019, s. 55–71.
- Krapohl D., Rosales T., *Decision Accuracy for the Relevant-Irrelevant Screening Test. A Partial Replication*, „Polygraph”, 43(1), 2014, s. 20–29.
- Krapohl D.J., Senter S.M., Stern B.A., *An Exploration of Methods for the Analysis of Multiple-Issue Relevant/Irrelevant Screening Data*, „Polygraph”, 34(1), 2005, s. 47–62.
- Krapohl D.J., Shaw P.K., *Fundamentals of Polygraph Practice*, San Diego 2015.
- Krzyścin A., *Ocena wyników badania poligraficznego* [niepublikowany skrypt w dyspozycji Agencji Bezpieczeństwa Wewnętrznego].
- Kulicki M., *Ekspertyza wariograficzna*, [w:] J. Wójcikiewicz, *Ekspertyza sądowa. Zagadnienia wybrane*, Warszawa 2007, s. 267–280.
- Larson J.A., *Modification of the Marston Deception Test*, „Journal of Criminal Law & Criminology”, 12(3), 1922, s. 390–399.
- Larson J.A., *The Cardio-pneumo-psychogram and Its Use in the Study of the Emotions, with Practical Applications*, „Journal of Experimental Psychology”, 5(5), 1923, s. 323–328.

- Larson J.A., *The Cardio-pneumo-psychogram in Deception*, „Journal of Experimental Psychology”, 6(6), 1923, s. 420–454.
- Lebrija E., *Evidence-based Best Scoring Practices using ESS* [prezentacja podczas Seminarium Polskiego Towarzystwa Badań Poligraficznych], Warszawa, 5.11.2022.
- Lee G.P., Meador K.J., *Neural Substrates of Emotion as Revealed by Functional Magnetic Resonance Imaging*, „Cognitive and Behavioral Neurology”, 17(1), 2004, s. 9–17.
- Leśniak B., Leśniak M., *Subiektywność w opiniach biegłych sądowych – prawne, etyczne i psychologiczne aspekty*, „Zeszyty Naukowe Politechniki Śląskiej”, 131, 2018, s. 287–297.
- Leszczyńska A., *Użyteczność badań poligraficznych w terapii i kontroli osób skazanych za przestępstwa przeciwko wolności seksualnej*, „Sexuologia Polska”, 11(2), 2013, s. 48–55.
- Leśniak M., *Wartość dowodowa opinii pismoznawczej*, Pińczów 2012.
- Leśniak M., Zubańska M., *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)*, „European Polygraph”, 4, 2(12), 2010, s. 81–86.
- Lipka M., *Pierwszy w Polsce przypadek zastosowania „wykrywacza kłamstw” w postępowaniu karnym*, „Problemy Kryminalistyki”, 48, 1964, s. 252–254.
- Lombroso C., *Crime. Its Causes and Remedies*, Boston 1911.
- Lykken D.T., *Psychology and the Lie Detector Industry*, University of Minnesota 1974.
- Lykken D.T., *The GSR in the Detection of Guilt*, „Journal of Applied Psychology”, 43(6), 1959, s. 385–388.
- Markiewicz J., Nedoma J., Serda H., *Z badań nad dokładnością oznaczeń alkoholu we krwi*, „Problemy Kryminalistyki”, 61–62, 1996, s. 472–478.
- Marston W.M., *Systolic Blood Pressure Symptoms of Deception*, „Journal of Experimental Psychology”, 2(2), 1917, s. 117–163.
- Marston W.M., *The Lie Detector Test*, New York 1938.
- Maczak A., *Style poznawcze*, [w:] J. Strelau (red.), *Psychologia. Podręcznik akademicki. Psychologia ogólna*, t. 2, Gdańsk 2006, s. 761–782.
- Materska M., *Procesy ewaluacji*, [w:] M. Materska, T. Tyszka (red.), *Psychologia i poznanie*, Warszawa 1992.
- Matte J.A., *A Field Study of the Backster Zone Comparison Technique's Either-Or Rule and Scoring System Versus Two Other Scoring Systems*

- When Relevant Question Elicits Strong Response*, „European Polygraph”, 4, 2(12), 2010, s. 53–69.
- Matte J.A., *Examination and Cross-Examination of Experts in Forensic Psychophysiology. Using The Polygraph*, New York 2000.
- Matte J.A., *Forensic Psychophysiology. Using The Polygraph. Scientific Truth Verification – Lie Detection*, New York 1996.
- Matte J.A., *The Art and Science of the Polygraph Technique*, Springfield 1980.
- Meijer E.H., Verschuere B., Merckelbach H.L.G., Crombez G., *Sex Offender Management Using the Polygraph. A Critical Review*, „International Journal of Law and Psychiatry”, 31(5), 2008, s. 423–429.
- Meintjes-van der Walt L., *Expert Evidence in the Criminal Justice Process. A Comparative Perspective*, Amsterdam 2011.
- Meiron E., Krapohl D., Kircher J.C., *An Assessment of the Backster „Either-Or” Rule in Polygraph Scoring*, „Polygraph”, 37(4), 2008, s. 240–249.
- Meyjes F., *Scientific Criminal Investigation Techniques under Dutch Law*, „Journal of Criminal Law and Criminology”, 51(6), 1961, s. 653–660.
- Miller F.M., *The Forensic Scientist in Court*, „Journal of Forensic Sciences”, 8(3), 1963, s. 315–323.
- Molchanov A.Y., Molchanova N.A., *Atlas poligramm*, Jaroslav 2007.
- Moszczyński J., *Subiektywizm w badaniach kryminalistycznych*, Olsztyn 2011.
- Najder K., *Schematy poznawcze*, [w:] M. Materska, T. Tyszka (red.), *Psychologia i poznanie*, Warszawa 1992, s. 38–60.
- National Academy of Sciences, *Strengthening Forensic Science in the United States. A Path Forward*, Washington 2009.
- National Research Council – National Academies Press, *The Polygraph and Lie Detection*, Washington 2003.
- Nelken J., *Dowód poszlakowy w procesie karnym*, Warszawa 1970.
- Nelson R., *Accuracy Effects for ESS and Three-Position Scores of Federal ZCT Exams Using the Grand Total Rule with Traditional/Federal and Multinomial Cutscores*, „Polygraph & Forensic Credibility Assessment”, 49(2), 2020, s. 157–175.
- Nelson R., *Explainer for the ESS-M Narrative Summary*, „APA Magazine”, 52(6), 2019, s. 30–48.
- Nelson R., *Literature Survey of Structural Weighting of Polygraph Signals. Why Double the EDA?*, „Polygraph & Forensic Credibility Assessment”, 48(2), 2019, s. 105–112.

- Nelson R., *Multinomial Cutscores for Bayesian Analysis with ESS and Three-Position Scores of Comparison Question Polygraph Tests*, „Polygraph & Forensic Credibility Assessment”, 49(1), 2020, s. 61–72.
- Nelson R., *Multinomial Reference Distributions for the Empirical Scoring System*, „Polygraph & Forensic Credibility Assessment”, 46(2), 2017, s. 81–115.
- Nelson R., *Tech Talk. Frequently Asked Questions about the LXSoftware PLE Scoring Tool*, „Lafayette Instrument”, 19.08.2016, www.researchgate.net/publication/306292694_Tech_Talk_Frequently_asked_questions_about_the_LXSoftware_PLE_scoring_tool [dostęp: 10.02.2021].
- Nelson R., *Using the Lafayette Respiration Line Excursion Tool*, „Polygraph Magazine”, 49(4), 2016, s. 59–62.
- Nelson R., Blalock B., Oelrich M., Cushman B., *Reliability of the Empirical Scoring System with Expert Examiners*, „Polygraph”, 40(3), 2011, s. 131–139.
- Nelson R., Handler M., *Empirical Scoring System. NPC Quick Reference*, Lafayette Instrument, 2010.
- Nelson R., Handler M., Coffey T., Prado R., Blalock B., *How To. A Step-by-Step Worksheet for the Multinomial ESS*, www.researchgate.net/publication/332240538_poligrafia_polygraph_How_To_A_Step-by-Step_Worksheet_for_the_Multinomial_ESS [dostęp: 9.03.2021].
- Nelson R., Handler M., Krapohl D., *Development and Validation of the Objective Scoring System, version 3*, www.oss3.info/poster1.html [dostęp: 27.02.2021].
- Nelson R., Handler M., Krapohl D., *Introduction to the Objective Scoring System – version 3*, www.oss3.info/index.html [dostęp: 27.02.2021].
- Nelson R., Handler M., Shaw P., Gougler M., Blalock B., Russell C., Cushman B., Oelrich M., *Using the Empirical Scoring System*, „Polygraph”, 40(2), 2011, s. 67–78.
- Nelson R., Krapohl D., *Criterion Validity of the Empirical Scoring System with Experienced Examiners. Comparison with the Seven-Position Evidentiary Model Using the Federal Zone Comparison Technique*, „Polygraph”, 40(2), 2011, s. 79–85.
- Nelson R., Krapohl D.J., Handler M., *Brute Force Comparison. A Monte Carlo Study of the Objective Scoring System Version 3 (OSS-3) and Human Polygraph Scorers*, „Polygraph”, 37(3), 2008, s. 185–215.
- Niedbała M., *Próby stwierdzenia istnienia w pamięci śladów za pomocą hipnozy i ich wykorzystanie w procesie*, [w:] D. Moskał, P. Wąsik

- (red.), *Ślady pamięciowe. Aspekty psychologiczne i kryminalistyczne*, Kraków 2013, s. 17–23.
- Niziołek K., Gołębiewska E., *The Perceived Validity of the Polygraph in Polish Criminal Courts*, „Polygraph & Forensic Credibility Assessment”, 47(1), 2018, s. 26–36.
- Office of the Under Secretary of Defense for Intelligence, *Department of Defense Polygraph Program Process and Compliance Study*, 19.12.2011, fas.org/sgp/othergov/polygraph/dod-poly.pdf [dostęp: 19.10.2012].
- Orne M.T., *Implications of Laboratory Research for the Detection of Deception*, „Polygraph”, 2(3), 1973, s. 169–199.
- Otter-Henderson K.D., Honts C.R., *Spontaneous Countermeasures during Polygraph Examinations. An Apparent Exercise in Futility*, „Modern Psychological Studies”, 8(1), 2001, s. 11–15.
- Patrick C.J., Iacono W.G., *Validity of the Control Question Polygraph Test. The Problem of Sampling Bias*, „Journal of Applied Psychology”, 76(2), 1991, s. 229–238.
- Pennigton N., Hastie R., *Juror Decision-Making Models. The Generalization Gap*, „Psychological Bulletin”, 89(2), 1981, s. 246–287.
- Pierce R.W., *The Peak of Tension Test*, „ISDD Bulletin”, 3(3), 1950, s. 1, 4–5.
- Podlesny J.A., *Is the Guilty Knowledge Polygraph Technique Applicable in Criminal Investigations? A Review of FBI Case Records*, „Crime Laboratory Digest”, 20, 1993, s. 57–61.
- Polski Komitet Normalizacyjny, *Ogólne wymagania dotyczące kompetencji laboratoriów badawczych i wzorcujących*, Warszawa 2005.
- Polskie Stowarzyszenie Poligraferów, *Certyfikowani poligraferzy Polskiego Stowarzyszenia Poligraferów*, poligrafer.com/certyfikowani-poligraferzy.html [dostęp: 10.06.2020].
- Polskie Towarzystwo Badań Poligraficznych, *Lista rekomendowanych ekspertów*, ptbp.pl/lista-ekspertow [dostęp: 10.06.2020].
- Polskie Towarzystwo Badań Poligraficznych, *Uchwała Nr 6/2013 Zebrania Założycielskiego Polskiego Towarzystwa Badań Poligraficznych z dnia 24 listopada 2013 r. w sprawie procedury wpisu na listę rekomendowanych przez Towarzystwo ekspertów z zakresu badań poligraficznych*, ptbp.pl/wp-content/uploads/Dokumenty/Uchwała_w_sprawie_procedury_wpisu_na_liste_rekomendowanych_przez_Towarzystwo_ekspertow.pdf [dostęp: 7.06.2023].

- Polskie Towarzystwo Badań Poligraficznych, *Uchwała nr 7/2016 Zarządu Polskiego Towarzystwa Badań Poligraficznych z dnia 20.07.2016 r. Standard kursów i szkoleń z zakresu badań poligraficznych*, ptbp.pl/wp-content/uploads/Dokumenty/PTBP-standard_kursow_i_szkolen_-_tekst_jednolity_2022.pdf [dostęp: 7.06.2023].
- Polygraph Examiners Administrative Rules of the Texas Department of Licensing and Regulation 16 Texas Administrative Code, Chapter 88 (effective September 1, 2015)*, www.tdlr.texas.gov/polygraph/polygraphrules090115.pdf [dostęp: 29.03.2021].
- Popper K.R., *The Logic of Scientific Discovery*, London 2002.
- Porges S.W., *Teoria poliowagalna*, tłum. A. Gomola, Kraków 2020.
- Raskin D.C., Barland G.H., Podlesny J.A., *Validity and Reliability of Detection of Deception*, „Polygraph”, 6(1), 1977, s. 1–39.
- Raskin D.C., Honts C.R., Kircher J.C., *Credibility Assessment. Scientific Research and Applications*, San Diego 2014.
- Raskin D.C., Kircher J.C., Honts C.R., Horowitz S.W., *A Study of The Validity of Polygraph Examinations in Criminal Investigations. Final Report to the National Institute of Justice*, Salt Lake City 1988.
- Reid J., Arther R., *Behavior Symptoms of Lie-Detector Subjects*, „Journal of Criminal Law and Criminology and Police Science”, 44, 1953, s. 104–110.
- Reid J., Inbau F., *Truth and Deception. The Polygraph (Lie-detector) Technique*, Baltimore 1977.
- Risinger D.M., Saks M.J., Thompson W.C., Rosenthal R., *The Daubert / Kumho Implications of Observer Effects in Forensic Science. Hidden Problems of Expectation and Suggestions*, „California Law Review”, 90(1), 2002, s. 1–56.
- Robertson B., Vignaux G.A., Berger Ch., *Interpreting Evidence. Evaluating Forensic Sciences in the Courtroom*, 2nd edition, Oxford 2016.
- Sakowicz A. (red.), *Kodeks postępowania karnego. Komentarz*, wyd. 7, Warszawa 2016.
- Salovey P., Caruso D., Mayer J.D., *Emotional Intelligence Meets Traditional Standards for an Intelligence*, „Intelligence”, 27, 1999, s. 267–298.
- Senate Select Committee on Intelligence, *An Assessment of the Aldrich H. Ames Espionage Case and Its Implications for U.S. Intelligence*, 1.11.1994, fas.org/irp/congress/1994_rpt/ssci_ames.htm [dostęp: 4.06.2020].

- Senter S., Weatherman D., Krapohl D., Horvath F., *Psychological Set or Differential Salience. A Proposal for Reconciling Theory and Terminology in Polygraph Testing*, „Polygraph”, 39(2), 2010, s. 109–117.
- Senter S.M., Dollins A.B., Krapohl D.J., *A Comparison of Polygraph Data Evaluation Conventions Used at the University of Utah and the Department of Defense Polygraph Institute*, „Polygraph”, 33(4), 2004, s. 214–222.
- Sevilla C., *The Polygraph Examiner as a Witness in Court*, „Polygraph”, 2(2), 1973, s. 122–132.
- Shurany T., Matte J.A., Stein E., *Influence of Case Facts on Blind Scorers of Polygraph Tests*, „European Polygraph”, 3–4(9–10), 2009, s. 133–139.
- Shurany T., Ravid I., *Evaluation of Polygraph Charts. Formats, Criteria and Scoring*, T. 1, Publications, Israel 2005.
- Slowik S.M., *Global Evaluation. An Inductive Approach to Case Resolution*, „Polygraph”, 11(3), 1982, s. 215–224.
- Slowik S.M., Buckley J.P., *Relative Accuracy of Polygraph Examiner Diagnosis of Respiration, Blood Pressure and GSR Recordings*, „Journal of Police Science and Administration”, 3(3), 1975, s. 305–310.
- Snyder M., *Self-monitoring of Expressive Behavior*, „Journal of Personality and Social Psychology”, 30(4), 1974, s. 526–537.
- Standard for Polygraphic Examinations in Criminal Cases*, „European Polygraph”, 1(1), 2007, s. 65–77.
- Stern B.A., Krapohl D.J., *The Infamous James Alphonso Frye*, „Polygraph”, 32(3), 2003, s. 188–199.
- Stępką L., *Rejestracja i ocena poprawności ekspertyzy wariograficznej (poligraficznej)*, [w:] V. Kwiatkowska-Wójcikiewicz, D. Wilk, J. Wójcikiewicz (red.), *Kryminalistyka a nowoczesne technologie*, Kraków 2019, s. 99–117.
- Stępnik A., *Reprezentacje umysłowe, rodzaje pamięci a wiedza*, „Przełęcz Filozoficzno-Literacki”, 2(39), 2014, s. 167–187.
- Stryszowski P., Bieńkuński J., *Wykorzystanie i znaczenie ekspertyz wariograficznych. Studium przypadku*, „Prokuratura i Prawo”, 7–8, 2014, s. 171–201.
- Suzuki A., Watanabe S., Ohnishi K., Arusuma N., *The Objective Analysis of GSR in the Detection of Deception. An Analysis of GSR Amplitudes in Terms of Rank Scores*, „Polygraph”, 8(1), 1979, s. 53–63.
- Suzuki A., Watanabe S., Ohnishi K., Matsuno K., Arasuna M., *Polygraph examiners' judgements in chart interpretation – reliability of judgement*, „Kagaku Keisatsu Kenkyunsho” (Police Science Report), 26(1), 1973.

- Swinford J., *Manually Scoring Polygraph Charts Utilizing the Seven-Position Numerical Analysis Scale at the Department of Defense Polygraph Institute*, „Polygraph”, 28(1), 1999, s. 10–27.
- Timm W.H., *Analyzing Deception from Respiration Patterns*, „Journal of Police Science & Administration”, 10(1), 1982, s. 47–51.
- Tomaszewski T., *Wartość niekategorycznych opinii biegłych*, „Nowe Prawo”, 9, 1970, s. 64–77.
- Trovillo P.V., *Deception Test Criteria. How One Can Determine Truth and Falsehood from Polygraphic Records*, „Journal of Criminal Law and Criminology”, 33(4), 1942, s. 338–358.
- Tyszka T., *Psychologiczne pułapki podejmowania decyzji*, Gdańsk 2000.
- U.S. Department of Defense, *Present Status of DoD Research on the Polygraph. Report of the DoD Joint Services Group On a Coordinated R&D Program of Lie Detection Research*, 28.08.1968.
- Vendemia J.M.C., Buzan R.F., Simon-Dack S.L., *Reaction Time of Motor Responses in Two-stimulus Paradigms Involving Deception and Congruity with Varying Levels of Difficulty*, „Behavioral Neurology”, 16, 2005, s. 25–36.
- Vrij A., Van der Steen J., Koppelaar L., *The Effects of Street Noise and Field Independence on Police Officers' Shooting Behavior*, „Journal of Applied Social Psychology”, 25(19), 1995, s. 1714–1725.
- Waltoś S., *Opinie biegłych – na ostrym zakręcie w polskim procesie karnym*, [w:] J. Czapska, A. Okrasa (red.), *Bezpieczeństwo – policja – kryminalistyka. W poszukiwaniu wiedzy przydatnej w praktyce*, Kraków 2015, s. 29–37.
- Weaver R.S., *Effects of Differing Numerical Chart Evaluation Systems on Polygraph Examination Results*, „Polygraph”, 14(1), 1985, s. 34–42.
- Weaver R.S., *The Numerical Evaluation of Polygraph Charts. Evolution and Comparison of Three Major Systems*, „Polygraph”, 9(2), 1980, s. 94–108.
- Webb A.K., Handler M.D., Krapohl D.J., Kircher J.C., *A Comparison of the Objective Scoring System and Probability Analysis*, „Polygraph”, 37(4), 2008, s. 250–255.
- Weber M., *Obiektywność poznania w naukach społecznych*, [w:] *Problemy socjologii wiedzy*, oprac. A. Chmielecki et al., Warszawa 1985.
- Wells G.L., Memon A., Penrod S.D., *Eyewitness Evidence. Improving its Probative Value*, „Psychological Science in the Public Interest”, 7(2), 2006, s. 45–75.

- What Were the Greatest Inventions of All Time?*, www.edinformatics.com/inventions_inventors [dostęp: 19.05.2020].
- Wicklender D.E., Hunter F.L., *The Influence of Auxiliary Sources of Information in Polygraph Diagnosis*, „Journal of Police Science and Administration”, 3(4), 1975, s. 405–409.
- Widacki J., *Analiza przesłanek diagnozowania w badaniach poligraficznych*, Katowice 1982.
- Widacki J. (red.), *Badania poligraficzne w Polsce*, Kraków 2014.
- Widacki J., *Glosa do wyroku Sądu Apelacyjnego w Katowicach z 7 czerwca 2017 r., II AKA 167/17*, „Palestra”, 10, 2017, s. 74–77.
- Widacki J., *Historia badań poligraficznych*, Kraków 2017.
- Widacki J., *Instytucja naukowa lub specjalistyczna w rozumieniu art. 193 § 2 k.p.k.*, „Państwo i Prawo”, 9, 2013, s. 45–53.
- Widacki J. (red.), *Kierunki rozwoju instrumentalnej i nieinstrumentalnej detekcji kłamstwa. Problemy kryminalistyczne, etyczne i prawne*, Kraków 2018.
- Widacki J. (red.), *Kryminalistyka*, wyd. 4, Warszawa 2018.
- Widacki J., *Obrońca wobec opinii biegłego w procesie karnym*, „Studia Prawnicze. Rozprawy i Materiały”, 1(16), 2015, s. 5–20.
- Widacki J., *Opinia z badań poligraficznych w procesie karnym*, „Palestra”, 7–8, 2017, s. 5–10.
- Widacki J., *Poligraf czy wariograf*, „Problemy Kryminalistyki”, 294(4), 2016, s. 35–39.
- Widacki J., *Results of Polygraph Examinations. Direct or Circumstantial Evidence?*, „European Polygraph”, 8, 2(28), 2014, s. 61–67.
- Widacki J., *W sprawie wyboru techniki badania poligraficznego. Czy technika oparta na testach GKT (CIT) jest lepsza od techniki opartej na testach CQT?*, „Problemy Kryminalistyki”, 273, 2011, s. 5–10.
- Widacki J., *Wartość diagnostyczna badania poligraficznego i jej znaczenie kryminalistyczne*, Kraków 1977.
- Widacki J., *Wprowadzenie do problematyki badań poligraficznych*, Warszawa 1981.
- Widacki J., Gołaszewski M., *Subiektywizm w badaniach poligraficznych*, „Studia Prawnicze. Rozprawy i Materiały”, 2(21), 2017, s. 189–216.
- Widacki J., Horvath F., *An Experimental Investigation of the Relative Validity and Utility of the Polygraph Technique and Three Other Common Methods of Criminal Identification*, „Journal of Forensic Sciences”, 23(3), 1978, s. 596–601.

- Widacki J., Mikrut Z., Widacki M., Antos J., *Próba wykorzystania zmian temperatury twarzy jako wskaźnika zmian emocjonalnych przy detekcji kłamstwa*, [w:] J. Widacki (red.), *Kierunki rozwoju instrumentalnej i nieinstrumentalnej detekcji kłamstwa. Problemy kryminalistyczne, etyczne i prawne*, Kraków 2018, s. 131–148.
- Widła T., *Ocena dowodu z opinii biegłego*, Katowice 1992.
- Wilcox D.T., *The Use of the Polygraph in Assessing, Treating and Supervising Sex Offenders. A Practitioner's Guide*, Chichester 2009.
- Wilcox D.T., Gray R., *The Use of the Polygraph with Sex Offenders in the UK*, „European Polygraph”, 6, 1(19), 2012, s. 55–67.
- Winter J.E., *A Comparison of the Cardio-pneumo-psychograph and Association Methods in the Detection of Lying in Cases of theft Among College Students*, „Journal of Applied Psychology”, 20(2), 1936, s. 243–248.
- Wojciechowski J., *Podstawy teoretyczne powstawania reakcji w czasie badań poligraficznych*, „Problemy Kryminalistyki”, 276, 2012, s. 19–31.
- Wojciszke B., *Psychologia społeczna*, Warszawa 2011.
- Wolańska-Nowak P., *Interpretacja wyników ekspertyzy*, [w:] J. Wójcikiewicz (red.), *Ekspertyza sądowa. Zagadnienia wybrane*, Warszawa 2007, s. 576–579.
- Wójcikiewicz J. (red.), *Ekspertyza sądowa. Zagadnienia wybrane*, Warszawa 2007.
- Życiński J., *Problem wyjaśniania w naukach nomotetycznych a poznanie faktu jednostkowego*, [w:] J. Widacki (red.), *Z zagadnień teorii opinii biegłego*, Materiały IV Sympozjum Metodologii Kryminalistyki i Nauk Pokrewnych, Chęciny 24–25 VI 1983, Katowice 1983, s. 59–71.

Orzecznictwo

- 113 S Ct 2786 (1993), casetext.com/case/daubert-v-merrell-dow-pharmaceuticals-inc? [dostęp: 21.05.2020].
- 293 F. 1013 (D.C. Cir. 1923), casetext.com/case/frye-v-united-states-7 [dostęp: 20.05.2020].
- 885 F.2d 1529 (11th Cir. 1989), casetext.com/case/us-v-piccinonna? [dostęp: 21.05.2020].
- Deskovic v. City of Peekskill*, 894 F. Supp. 2d 443 (S.D.N.Y. 2012), casetext.com/case/deskovic-v-city-of-peekskill-3 [dostęp: 8.06.2020].

People v. Kenny, 167 Misc. 51, 3 N.Y.S.2d 348 (1938), casetext.com/case/people-v-kenny-19 [dostęp: 20.05.2020].

Postanowienie SN z dnia 29 stycznia 2015 r., I KZP 25/14.

Wyrok SN z dnia 12 listopada 1973 r., II KR 285/72.

Wyrok SO w Białymstoku z dnia 5 stycznia 2016 r., III K 92/15.

Książka poświęcona została jednemu z najważniejszych aspektów badań poligraficznych, związanemu ze sposobem ich prowadzenia, a zwłaszcza z oceną uzyskiwanych wyników [...]

Niebagatelne znaczenie miało doświadczenie zawodowe i własna praktyka ekspercka Autora, który potrafił te atuty w pełni wykorzystać. W widoczny sposób wpłynęło to na treść opracowania oraz wartość podanych w nim ustaleń i wniosków, jak również podniosło jego walor naukowy i praktyczny [...]

Książka powinna być obowiązkowym źródłem informacji dla biegłych i innych specjalistów zajmujących się badaniami poligraficznymi, ale będzie też ciekawą lekturą dla biegłych innych specjalności i osób zainteresowanych od strony naukowej i praktycznej kryminalistyki, jak również dla studentów różnych kierunków studiów (np. prawniczych, bezpieczeństwa wewnętrznego, nauk policyjnych).

*fragment recenzji sporządzonej
przez prof. dr. hab. Tadeusza Tomaszewskiego
z Uniwersytetu Warszawskiego*

