

kom. Dominika Słapczyńska

ekspert w Zakładzie Badań Dokumentów i Technik Audiowizualnych CLKP

## Ocena zapisów testów porównań strefowych UTAH przy użyciu Empirycznego Systemu Oceny (ESS)

### Wstęp

Badania poligraficzne nie należą do nowych metod wykorzystywanych w kryminalistyce, choć problematyka ich wiarygodności nadal stanowi pole do dyskusji. W ostatnim czasie w literaturze fachowej niejednokrotnie poruszano problem ich trafności<sup>1</sup>. Kwestia ta jest niezwykle istotna w praktyce sądowej, gdy opinia z badania poligraficznego wliczana jest w poczet dowodów obciążających. W tej sytuacji rodzi się pytanie o poziom dokładności uzyskanych wyników ekspertyzy.

Ostatnie pięć lat przyniosło dynamiczny rozwój badań poligraficznych, zapewniając jej specjalistom narzędzia w postaci wykazu zwalidowanych technik badawczych, sprawdzonych metod analizy danych psychofizjologicznych oraz wspomagających ten proces algorytmów komputerowych.

Wyniki badań naukowych przeprowadzonych nad analizą i interpretacją danych psychofizjologicznych zarejestrowanych w trakcie badań poligraficznych pozwalają obecnie oszacować poziom prawdopodobieństwa błędu konkretnego rozstrzygnięcia. Biegły nie musi już unikać na rozprawie sądowej odpowiedzi o procentowe określenie dokładności badań poligraficznych, gdyż prowadzenie czynności badawczych zgodnie z wytycznymi wybranej techniki pozwala na podstawie danych normatywnych podać w przybliżeniu poziom błędu, którym obarczony jest wynik ekspertyzy.

Dobór właściwej techniki lub technik badawczych to jeden z podstawowych czynników decydujących o metodologicznej poprawności badania i późniejszych analiz. Z poszczególnymi technikami uzyskiwania danych psychofizjologicznych wiążą się określone procedury ich analizy i interpretacji. Niniejszy artykuł stanowi propozycję sposobu analizy i oceny danych psychofizjologicznych uzyskanych w kryminalistycznych badaniach poligraficznych opartych na teście porównań strefowych Utah, którego wartość diagnostyczna sięga powyżej 90% bez wyników rozstrzygniętych, co stanowi wymóg dla metod dowodowych w procedurze amerykańskiej (ASTM<sup>2</sup>).

Zapisy reakcji, zgodnie z metodologią techniki CQT, pierwotnie oceniane są jakościowo, a następnie przy użyciu zaawansowanych metod numerycznych. Wśród tych ostatnich znajduje się Empiryczny System Oceny (Empirical Scoring System – ESS), którego zastosowanie pozwala wyznaczyć poziom wiarygodności konkretnego badania.

### Technika pytań porównawczych

Badania poligraficzne opierają się na dwóch grupach testów. Pierwsza z nich obejmuje testy weryfikacyjne, a wśród nich testy pytań porównawczych (Comparison Question Test – CQT). Druga grupa to grupa testów rozpoznawczych, do której należy test informacji ukrytej (Concealed Information Test – CIT). Testy pytań porównawczych są powszechnie stosowane zarówno w badaniach diagnostycznych, jak i w procedurze skryningowej<sup>3</sup>.

Koncepcja testów CQT zakłada, iż badany niemający związku z danym przestępstwem zareaguje większym pobudzeniem psychofizjologicznym na pytania kontrolne, które dotyczą czynów o charakterze przestępczym lub nagannym moralnie, innych niż ten będący przedmiotem postępowania. Pytania krytyczne natomiast spowodują pobudzenie emocjonalne i fizjologiczne osoby, która swój związek z daną sprawą będzie próbowała zataić, czy to z obawy o przykre dla niej konsekwencje, czy to z innych znanych jej powodów, np. z chęci pomocy bliższej osobie.

### Test Utah

Wśród testów CQT tylko nieliczne spełniają kryterium dopuszczalności dowodowej, jakim jest wartość diagnostyczna powyżej 90%. Wyniki metaanalizy<sup>4</sup> przeprowadzonej przez R. Nelsona i współpracowników wskazują, iż trafność testu UTAH plasuje się na poziomie 93% (Utah PLT) oraz 91% (Utah DLT) z wyłączeniem wyników nierozstrzygniętych.

Test Utah można przeprowadzać w dwóch podstawowych wersjach: z trzema lub czterema pytaniami krytycznymi.

W pierwszym przypadku sekwencja pytań testowych będzie przedstawiała się w następujący sposób:

Przykład 1. Sekwencja pytań w teście Utah PLT w wersji *multi-facet*.

1. SY Czy jest Pan przekonany, że zadam tylko te pytania, które dzisiaj omówiliśmy?
2. SR Czy zamierza Pan odpowiadać zgodnie z prawdą na pytania o śmierć J. Kowalskiego?
3. N1 Czy obecnie jest maj?
4. C1 Czy przed 2012 rokiem użył Pan siły fizycznej wobec kogoś?
5. R1 Czy zadał Pan rany J. Kowalskiemu w dniu 12.04.2013 r.?
6. N2 Czy dzisiaj jest poniedziałek?
7. C2 Czy przed 2012 r. był Pan agresywny wobec bliższej osoby?
8. R2 Czy był Pan w domu J. Kowalskiego w dniu 12.04.2013 r.?
9. N3 Czy jest obecnie 2013 r.?
10. C3 Czy przed 2012 r. użył Pan przemocy wobec kogoś?
11. R3 Czy Pan wie kto pozbawił życia J. Kowalskiego w dniu 12.04.2013 r.?

Pytania krytyczne testu Utah mogą dotyczyć jednej kwestii związanej z zachowaniem sprawcy zdarzenia, a same testy określa się wtedy jako jednokwestyjne (*single-issue test*). Pytania krytyczne w teście jednokwestyjnym w sprawie napadu na sklep mogą mieć następującą postać:

R5 Czy dokonał Pan napadu na sklep przy ul. Szanajcy w Warszawie w dniu 4 maja br.?

R8 Czy dokonał Pan napadu na sklep monopolowy na ul. Szanajcy w zeszłym miesiącu?

R11 Czy brał Pan udział w napadzie na sklep monopolowy na Pradze w maju br.?

Pytania krytyczne testu wielokwestyjnego (*multi-facet test*) odnoszą się natomiast do różnych istotnych okoliczności (aspektów) zachowania sprawcy danego zdarzenia. W badaniu mogą one mieć postać pytań krytycznych przytoczonych w powyższym przykładzie (nr 1).

Z praktyki wynika, że rozpoczęcie czynności badawczych od przeprowadzenia testu próbnego i wytlumaczenie na jego podstawie osobie badanej istoty badania podnosi skuteczność ekspertyzy<sup>5</sup>. Zgodnie z wytycznymi techniki Utah<sup>6</sup> w trakcie badań test prezentuje się co najmniej trzykrotnie, przy czym pytania krytyczne rotują względem pytań kontrolnych, zmieniając miejsce w prezentowanej sekwencji. Ponadto zaleca się prowadzenie stymulacji międzytestowej (np. ponowne omówienie z badanym pytań krytycznych i porównawczych) w celu przeciwdziałania procesowi habituacji badanego<sup>7</sup>.

Analiza uzyskanych poligramów odbywa się dwuetapowo. Po ocenie jakościowej, która pozwala zweryfikować jakość uzyskanych zapisów, poszczególne zmiany na

krzywych analizowane są zgodnie z regułami wybranej metody numerycznej. W przypadku testu Utah można wykorzystać system siedmiostopniowej oceny o tej samej nazwie co format testu (system numerycznej oceny Utah) lub wybrać prostszy, ale równie skuteczny, Empiryczny System Oceny.

Analiza numeryczna zapisów uzyskanych przy użyciu techniki Utah zakłada porównanie natężenia zmian zarejestrowanych na pytania krytyczne względem poprzedzających je reakcji na pytania kontrolne.

### Analiza danych z wykorzystaniem Empirycznego Systemu Oceny (Empirical Scoring System – ESS)

Skuteczność ESS została potwierdzona przez jej autorów w drodze kilku niezależnych badań eksperymentalnych<sup>8</sup>. Zgodnie z danymi jej średnia trafność wynosi ok. 90%, a średni poziom testów nierozstrzygniętych plasuje się na poziomie 10%. Analizy mające na celu zestawienie skuteczności prezentowanej metody oceny numerycznej z innymi systemami pokazują, iż trafność ESS oraz metody siedmiostopniowej jest porównywalna<sup>9</sup>.

Podkreślić należy, że ESS nie zalicza się do metod opartych na nowych ustaleniach w obszarze badań psychofizjologicznych<sup>10</sup>. Jego twórcy korzystali z rozwiązań i ustaleń naukowych poczynionych już przez C. Backstera (1963) w zakresie przekształcania danych liczbowych czy Kirchera i Raskina (1988) odnośnie do cech diagnostycznych, które stanowią tu podstawowe kryteria analizy zapisów<sup>11</sup>.

ESS, jako metoda analizy i oceny reakcji fizjologicznych w testach pytań porównawczych, oparta jest na trzech głównych wytycznych.

1. Analizy poligramów należy dokonać na podstawie 3 cech Kirchera<sup>12</sup>, które obejmują długość linii oddechu (stłumienie oddechu, wzrost częstotliwości oddechu, tymczasowe podniesienie linii podstawowej krzywej oddechu), wielkość amplitudy krzywej elektrodermalnej, wielkość amplitudy linii podstawowej krzywej naczyniowo-sercowej. Blok oddechowy uznawany jest za diagnostyczny tylko wówczas, gdy został zarejestrowany po prezentacji pytania krytycznego. Reakcja w tym kanale musi obejmować co najmniej trzy pełne cykle oddechowe.
2. Identyfikacja wyżej wymienionych cech diagnostycznych na poszczególnych krzywych oraz porównanie ich wielkości na pytania krytyczne i kontrolne pozwala przypisać im wartości numeryczne w skali trzystopniowej. W kanale oddechowym oraz naczyniowo-sercowym natężenie zmian oceniane jest poprzez przyznanie im wartości -1, 0, +1, natomiast reakcjom elektrodermalnym -2, 0, +2.
3. ESS wykorzystuje trzy zasady analizy danych fizjologicznych:

zasada 1: porównując reakcje na pytania porównawcze i krytyczne, wartość punktową należy przypisać zmianie o większym natężeniu, którą można zaobserwować wzrokowo. Nie ma natomiast znaczenia, jak duża jest różnica w natężeniu porównywanych reakcji – im reakcja większa, tym lepiej.

zasada 2: identyfikując poszczególne cechy, należy zwrócić uwagę na moment ich pojawienia się na krzywych; ocenie należy poddać tylko te reakcje, które występują w tzw. oknie oceny reakcji<sup>13</sup>; analizie i ocenie nie podlegają te reakcje, które pojawiły się na krzywych przed prezentacją pytania lub kilka sekund po udzieleniu odpowiedzi przez badanego

zasada 3: nie należy oceniać metodą numeryczną niestabilnych, nieregularnych zapisów zawierających artefakty.

Przy definiowaniu reakcji diagnostycznych czynnikiem kluczowym jest wspomniany już moment, w którym pojawiają się one na poszczególnych krzywych. I tak: okno oceny reakcji w kanale oddechowym (P) rozpoczyna się w chwili prezentowania pytania. W kanale naczyniowo-sercowym (CA) analizowane są zmiany, które pojawiają się od momentu prezentowania bodźca do chwili zakończenia reakcji. Reakcja elektrodermalna (EDA) powinna pojawić się pomiędzy 0,5 sekundy od chwili prezentacji pytania do 5 sekund po odpowiedzi badanego na pytanie.

Porównywanie wielkości reakcji na pytania krytyczne i kontrolne w każdym z kanałów prowadzi się wizualnie bez konieczności precyzyjnego określenia tej różnicy. Autorzy ESS przyjęli założenie, iż większa zmiana w przebiegu zarejestrowanych parametrów jest powodowana przez bodziec, który dla badanego jest bardziej znaczący pod względem czynników emocjonalnych, poznawczych oraz behawioralnych, które przyczyniły się do jego powstania. Jeśli ekspert stwierdzi większe natężenie reakcji na konkretne pytanie niż ta w porównywanym pytaniu, przyznaje jej punkt. Natomiast brak wizualnej różnicy pomiędzy porównywanymi reakcjami uprawnia do przyznania wartości 0.

Wartość dodatnia (+1 dla kanału CA i P lub +2 dla EDA) jest przyznawana, jeśli reakcja na pytanie kontrolne jest większa od tej zarejestrowanej na pytanie krytyczne. Wartość ujemną (-1 dla CA i P oraz -2 dla EDA) przypisuje się reakcji, która pojawiła się po prezentacji bodźca krytycznego i jest większa w stosunku do zmiany zarejestrowanej na pytanie kontrolne. Zmianom występującym na obu krzywych oddechowych nadaje się jedną wspólną wartość liczbową.

### Zasady interpretacji wartości punktowych na podstawie danych normatywnych

Na kolejnym etapie analizy punkty nadane poszczególnym zmianom na krzywych P, CA oraz EDA na pytania krytyczne w trzech powtórzeniach testu Utah ulegają

zsumowaniu, dając w ten sposób trzy oceny cząstkowe dla każdego pytania R5, R8, R11. W przypadku testów diagnostycznych (*single-issue Utah, multi-facet Utah*), oceny cząstkowe sumowane są w wynik całkowity (*grand total*). Wynik końcowy daje podstawę do interpretacji badania i ustalenia poziomu istotności rezultatu.

Metodologia CQT na podstawie uzyskanych wyników punktowych dopuszcza 3 typy wnioskowania: Deception Indicated DI (wprowadzenie w błąd), No Deception Indicated NDI (brak wprowadzenia w błąd) oraz Inconclusive INC<sup>14</sup> (wynik nierozstrzygnięty).

Przyporządkowanie wyniku testu do jednej z powyższych kategorii odbywa się na podstawie dwóch progów punktowych, tj.: dla DI i NDI, które należy wyznaczyć w zależności od zakładanej przez eksperta czułości pomiaru.

Oznacza to, że progi punktowe skorelowane są z przyjętym przez badacza poziomem tolerancji błędów ( $\alpha$ ) dla danego pomiaru. W kryminalnych testach diagnostycznych twórcy ESS rekomendują przyjęcie tolerancji błędów na poziomie  $\alpha = 10\%$  (0,1) dla rozstrzygnięcia negatywnego<sup>15</sup> NDI oraz  $\alpha = 5\%$  (0,05) dla rozstrzygnięcia pozytywnego<sup>16</sup> DI.

Wraz z obniżeniem poziomu tolerancji błędów (poniżej 10%) dla NDI wzrasta poziom czułości pomiaru, ale koreluje on dodatkowo z podwyższeniem poziomu wyników nierozstrzygniętych.

Progi punktowe w zależności od przyjętej dokładności badania w ESS wyznacza się na podstawie danych normatywnych<sup>17</sup>.

R. Nelson i współpracownicy, bazując na jednolitym, prostym systemie analizy i oceny danych oraz posługując się zaawansowanymi narzędziami statystycznymi (metoda Monte Carlo)<sup>18</sup>, ustalili poziom prawdopodobieństwa błędów dla każdego wyniku punktowego uzyskanego zarówno w testach diagnostycznych, jak i przesiewowych. W tab. 1 zgromadzono informacje o charakterze empirycznym, które pozwalają określić optymalne progi punktowe w zależności od zakładanego przez eksperta poziomu dokładności badania.

Przyjmując proponowany poziom tolerancji błędów i posługując się tab. 1, ekspert może ustalić dwa progi punktowe dla trzech rodzajów rozstrzygnięć, które wynoszą +2 dla NDI oraz -4 dla DI. Jeśli całkowita suma punktów przyznanych na podstawie reakcji na wszystkie pytania krytyczne testu jest równa lub większa od +2, wynik testu jest klasyfikowany jako brak wprowadzenia w błąd (NDI). Suma punktów równa lub mniejsza od -4 uprawnia do postawienia diagnozy wprowadzenia w błąd (DI). Wynik w przedziale od -4 do +2 uznawany jest za nierozstrzygnięty (INC).

### Wyznaczanie poziomu istotności – *p-value*

Na podstawie wyniku końcowego testu (*grand total*) ekspert może ocenić prawdopodobieństwo wystąpienia błędów (*p-value*) w analizowanym przez niego przypadku.

Tabela 1

Tabela przedstawia rozkład prawdopodobieństwa błędu dla różnych wyników numerycznej analizy danych testowych uzyskanych przy użyciu ESS w badaniach wszystkimi formatami CQT z 3 pytaniami relewantnymi

Table showing the probability of error for various results of test data analysis according to the ESS system in examinations including all CQT tests with three relevant questions

Progi punktowe – wynik NDI		Progi punktowe – wynik DI	
Na podstawie dystrybucji końcowych wyników dla osób klasyfikowanych jako NDI		Na podstawie dystrybucji końcowych wyników dla osób klasyfikowanych jako DI	
Próg punktowy	Wartość p	Próg punktowy	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
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		
16	<0,001		

Źródło: R. Nelson, M. Handler, Empirical Scoring System: NPC Quick Reference, Lafayette Instrument. Za zgodą.

W naukach przyrodniczych błąd nie jest synonimem pomyłki. Poziom istotności oznacza niemożliwą do uniknięcia niepewność nierozzerwalnie związaną ze specyfiką przedmiotu badania – tu z reakcjami psychofizjologicznymi. W tym sensie niepewność nie oznacza pomyłek i nie sposób jej uniknąć, zachowując większą staranność.

*P-value* jest zatem statystyczną miarą pozwalającą na zorientowanie się, na ile wiarygodne są osiągnięte wyniki. Badania naukowe w obszarze badań poligraficznych przeprowadzane są na ograniczonej liczbie osób – tzw. próbie, a uogólnianie obserwowanych reakcji na całą populację jest obarczone błędem (być może dana osoba na skutek zbiegu różnych okoliczności reaguje w specyficzny sposób). *P-value* to właśnie statystyczna miara pozwalająca określić, na ile prawdopodobne jest, że obserwowane zjawisko jest takim właśnie artefaktem, tj. jego powodem jest zbieg okoliczności, a nie rzeczywiście istniejące zjawisko.

W celu ustalenia *p-value* dla konkretnego wyniku badania poligraficznego uzyskany całkowity wynik punktowy

badacz musi odszukać w tab. 1 i wyznaczyć odpowiadający mu poziom prawdopodobieństwa błędu.

Przykład nr 2: rozstrzygnięcie NDI

Pytanie krytyczne: R5	R8	R11
oceny cząstkowe: +1	+1	+1

Suma (*grand total*): +3 ( $p = ,067$ )

Dla obliczonego wyniku, posługując się tab. 1, można teraz ustalić wartość *p-value*, a następnie dokonać jej interpretacji:

**Całkowity wynik testu (+3) pozwala na zakwalifikowanie badanego jako osoby niemającej związku ze zdarzeniem będącym przedmiotem postępowania, a prawdopodobieństwo popełnienia błędu w analizowanym przypadku wynosi 6,7%. Innymi słowy prawdopodobieństwo, iż osoba badana nie ma związku z danym zdarzeniem w zakresie stawianych pytań krytycznych wynosi 93,3%<sup>19</sup>.**

## Przykład nr 3: rozstrzygnięcie DI

Pytanie krytyczne: R5	R8	R11
oceny cząstkowe: -2	-2	-1

suma -5 ( $p = ,032$ )

Przyjmując analogiczną formułę z przykładu 2, można wnioskować, iż:

**Całkowity wynik testu (-5) pozwala na zakwalifikowanie badanego jako osoby mającej związek ze zdarzeniem będącym przedmiotem postępowania, a prawdopodobieństwo popełnienia błędu dla analizowanego przypadku wynosi 3,2%. Innymi słowy prawdopodobieństwo, iż osoba badana ma związek z danym zdarzeniem w zakresie stawianych pytań krytycznych wynosi 96,8%.**

Jeśli natomiast całkowita liczba punktów testu mieści się w przedziale pomiędzy -4 a +2, wynik testu klasyfikowany jest jako nierozstrzygnięty (*inconclusive* – INC). W takiej sytuacji zastosowanie znajduje analiza dwustopniowa na podstawie tzw. reguły Sentera<sup>20</sup>. Jeśli wynik całkowity testu nie daje podstaw do jego rozstrzygnięcia, należy zastosować analizę opartą wyłącznie na ocenach cząstkowych ustalonych dla poszczególnych pytań krytycznych.

Dla testów nierozstrzygniętych autorzy ESS na drugim etapie oceny zastosowali również tzw. poprawkę Bonferroniego<sup>21</sup>, która wprowadziła nowy poziom tolerancji błędu  $\alpha = 1,7\%$  ( $,017$ ) dla DI (zamiast  $\alpha = 0,5$ ), któremu zgodnie z danymi normatywnymi można przypisać wartość -7.

Opierając analizę jedynie na ocenach cząstkowych, ekspert ustala, które pytanie krytyczne uzyskało najmniejszą wartość liczbową. Jeśli  $p < \alpha$  ( $,017$ ), to wynik testu klasyfikowany jest jako DI; taka sytuacja ma miejsce, gdy którakolwiek z ocen cząstkowych jest równa lub mniejsza -7.

## Przykład nr 4: rozstrzygnięcie DI z zastosowaniem reguły Sentera

Pytanie krytyczne: R5	R8	R11
Oceny cząstkowe +3	+3	-9 ( $p = ,008$ )

suma -3 ( $p = ,058$ )

Interpretacja na podstawie danych liczbowych i normatywnych brzmi:

**Wynik testu oparty na ocenie cząstkowej o najniższej wartości (-9) pozwala na zakwalifikowanie badanego jako osoby mającej związek ze zdarzeniem będącym przedmiotem postępowania, a prawdopodobieństwo popełnienia błędu w analizowanym przypadku wynosi  $p = 2,4\%$ <sup>22</sup>. Innymi słowy prawdopodobieństwo, iż osoba badana ma związek z danym zdarzeniem w zakresie stawianych pytań krytycznych wynosi 97,6%.**

Jeśli natomiast wynik cząstkowy o najniższej wartości jest większy od -7, test pozostaje nierozstrzygnięty.

## Przykład nr 5: rozstrzygnięcie INC

Pytanie krytyczne: R5	R8	R11
ocena cząstkowa +1	+2	-4 ( $p = ,043$ )

suma -1 ( $p = ,159$ )

W tym przypadku:

**Test pozostaje nierozstrzygnięty. Poziom prawdopodobieństwa błędu dla tego rozstrzygnięcia wynosi  $p = 15,9\%$ , co przewyższa przyjętą dla danego pomiaru tolerancję błędu  $\alpha = 5\%$ , a poziom prawdopodobieństwa błędu dla oceny cząstkowej o najniższej wartości wynosi  $p = 4,3\%$ , co przewyższa poziom tolerancji błędu  $\alpha = 1,7\%$  z poprawką Bonferroniego.**

Pamiętać należy, iż zastosowanie reguły Sentera nigdy nie daje końcowego wyniku testu w postaci NDI.

Podsumowując, jeśli wynik całkowity testu jest równy lub mniejszy niż -4 lub ocena cząstkowa jednego z pytań krytycznych w teście nierozstrzygniętym jest równa lub mniejsza niż -7, wynik końcowy badania klasyfikowany jest jako wprowadzanie w błąd (DI).

### Analiza testów UTAH z zastosowaniem ESS i danych normatywnych – przykłady

Poniżej przytoczono zapisy z dwóch spraw, w których prowadzono badania z zastosowaniem testów Utah PLT, a ich wyniki zostały potwierdzone kolejnymi ustaleniami poczynionymi przez prowadzących postępowania.

Poligramy oceniono metodą ilościową ESS (w tabelach przedstawiono wartości liczbowe) na podstawie reakcji diagnostycznych wyznaczonych na poszczególnych krzywych. Uzyskane rezultaty zinterpretowano z zastosowaniem danych normatywnych, co pozwoliło na określenie prawdopodobieństwa błędu w obu przypadkach.

## Przypadek 1. Wynik testu NDI

## Numeryczna ocena ESS:

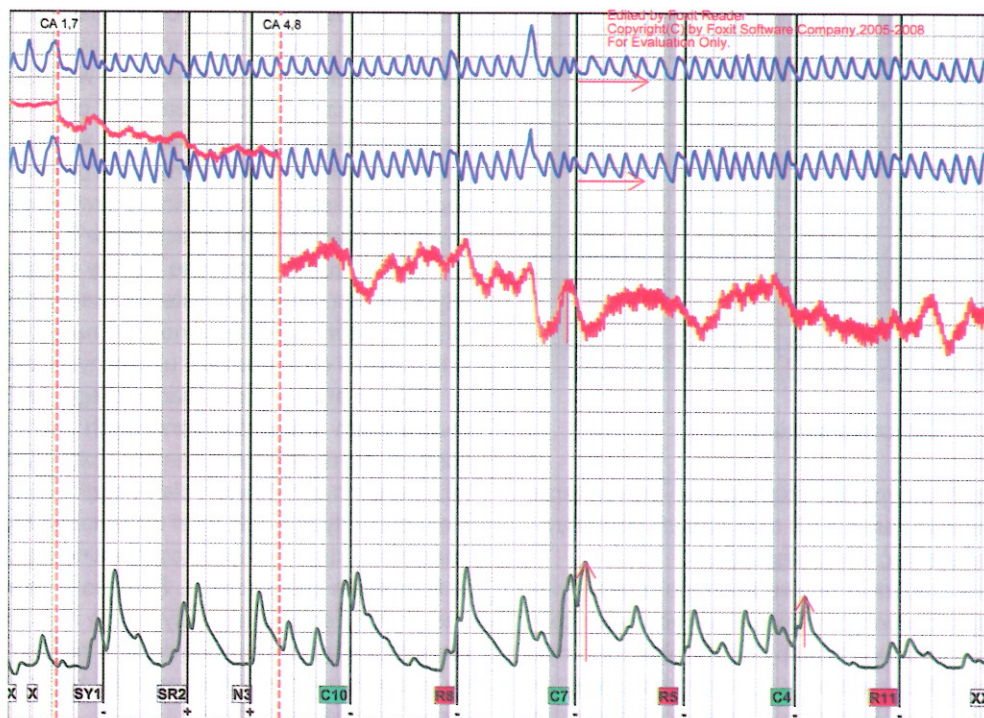
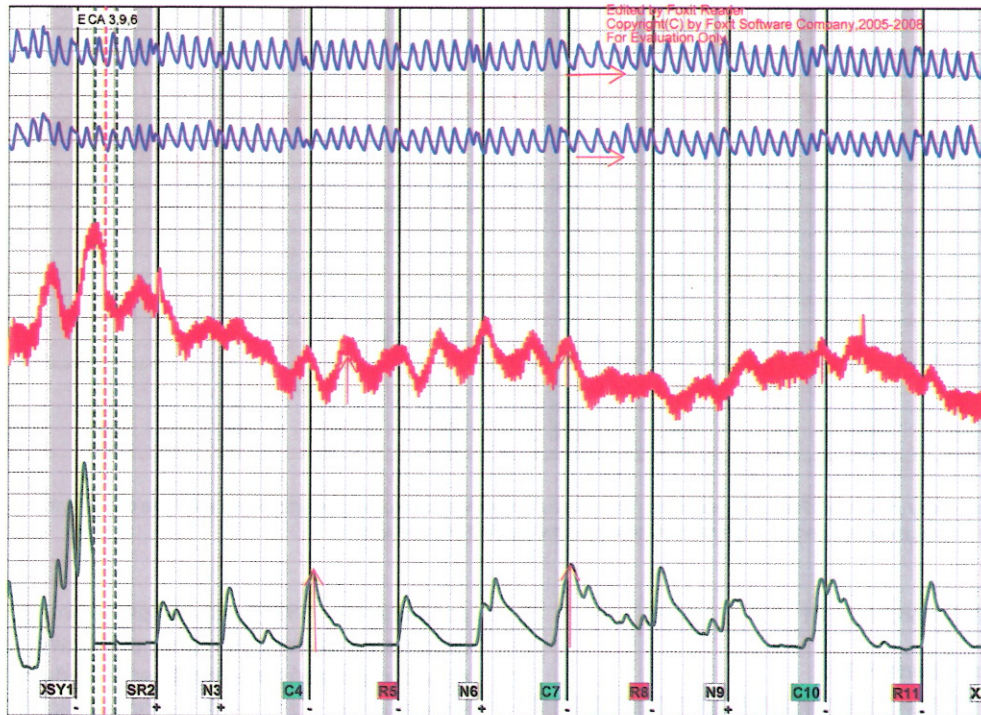
Pytania krytyczne	R5	R8	R11
		Test1	
P	0	+1	0
EDA	+2	+2	0
CA	+1	+1	+1
		Test 2	
P	+1	0	0
EDA	+2	0	+2
CA	+1	-1	-1
		Test 3	
P	+1	0	0
EDA	+2	+2	+2
CA	+1	0	+1
Oceny cząstkowe	11	5	5

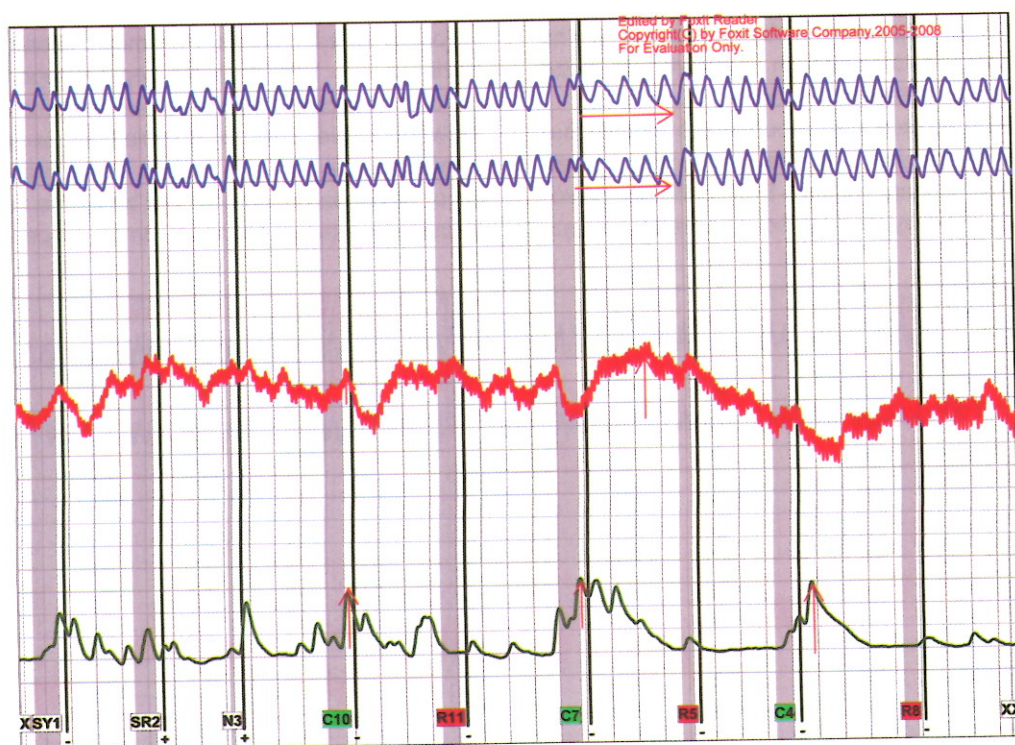
Całkowity wynik testu to +21

W tym przypadku statystyczna interpretacja danych na podstawie zasad ESS oraz danych normatywnych może przedstawiać się w następujący sposób:

Na podstawie naukowo udowodnionego modelu numerycznej analizy danych o nazwie Empiryczny System Oceny (Empirical Scoring System – ESS) obliczono całkowity wynik testu, który wynosi +21, co oznacza brak zwią-

ku osoby ze zdarzeniem. Przyjmując tolerancję błędów na poziomie  $\alpha = 10\%$  (0,1) dla negatywnego rozstrzygnięcia oraz związany z nim próg punktowy na poziomie +2, prawdopodobieństwo popełnienia błędów w tym przypadku wynosi poniżej  $p = 0,1\%$ . Innymi słowy, prawdopodobieństwo, że osoba badana nie jest związana ze zdarzeniem w zakresie pytań krytycznych wynosi więcej niż 99,9%.





Ryc. 1. Trzy poligramy osoby „niewinnej”. Reakcje o większym natężeniu na pytania kontrolne oznaczono strzałkami w kolorze czerwonym  
 Fig. 1. Three polygrams of „innocent” person. High intensity reactions are marked with red arrows

#### Przypadek 2. Wynik testu DI

##### Numeryczna ocena ESS:

Pytania krytyczne	R5	R8	R11
		Test 1	
P	-1	0	0
EDA	+2	-2	-2
CA	-1	0	-1
		Test 2	
P	0	0	0
EDA	-2	-2	-2
CA	-1	-1	-1
		Test 3	
P	0	0	0
EDA	0	-2	-2
CA	-1	-1	-1
Oceny cząstkowe	-4	-8	-9

Całkowita suma -21

W powyższym przypadku statystyczna interpretacja danych na podstawie zasad ESS oraz danych normatywnych może przedstawiać się następująco:

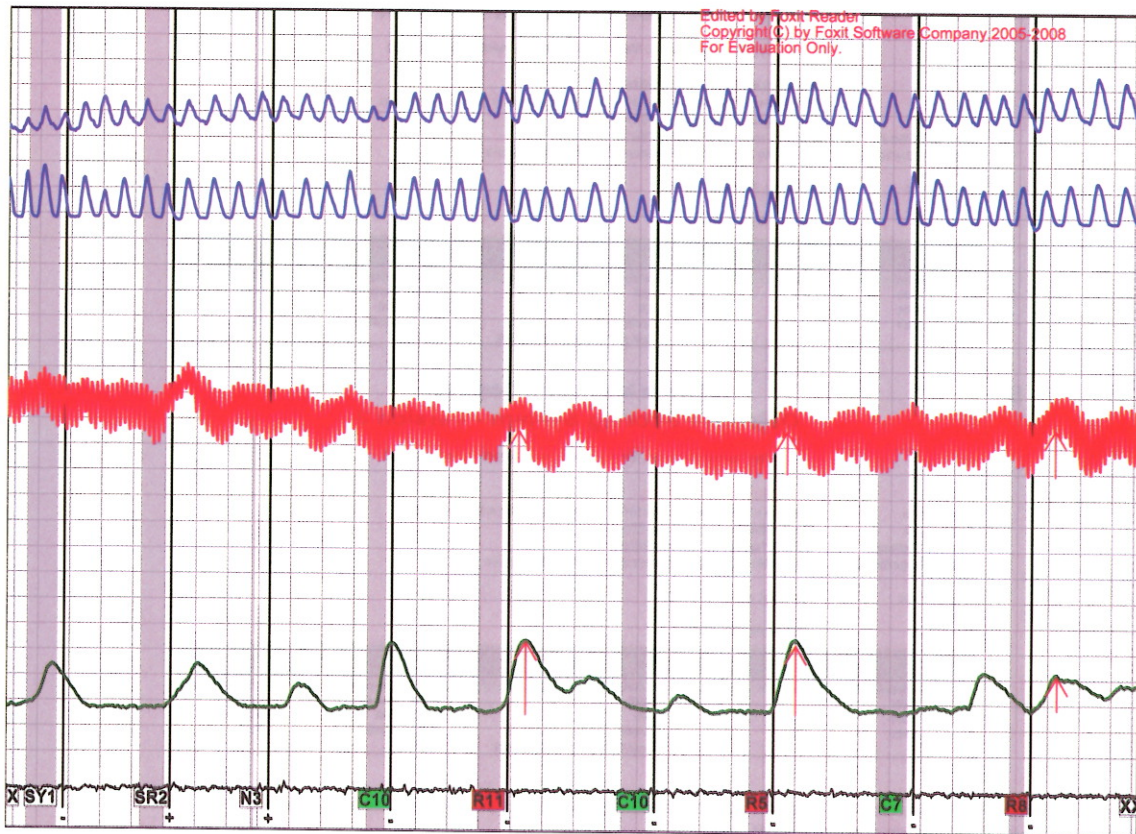
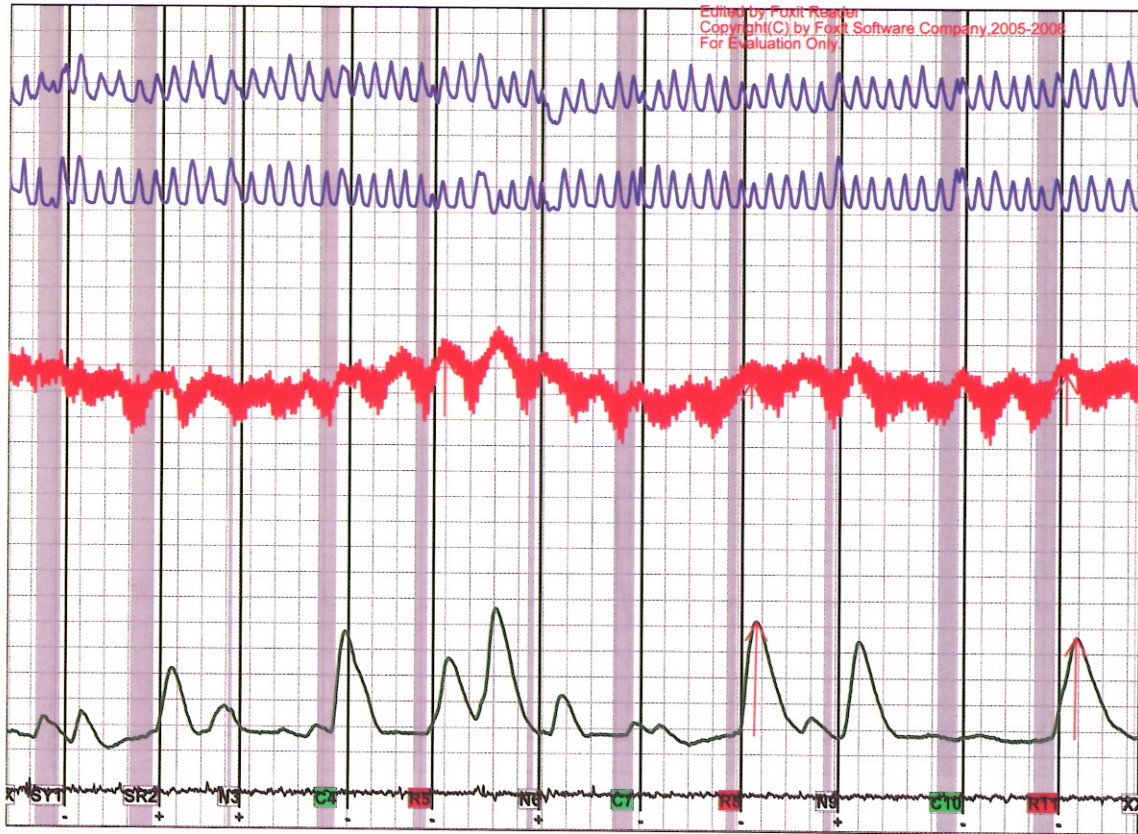
Na podstawie naukowo udowodnionego modelu numerycznej analizy danych o nazwie Empiryczny System Oceny (Empirical Scoring System – ESS) obliczono całkowity wynik testu, który wynosi -21, co oznacza związek osoby ze zdarzeniem. Przyjmując tolerancję błędów na

poziomie  $\alpha = 5\%$  (0,05) dla pozytywnego rozstrzygnięcia oraz związany z nim próg punktowy na poziomie -4, prawdopodobieństwo popełnienia błędów w tym przypadku wynosi mniej niż  $p = 0,1\%$ . Innymi słowy, prawdopodobieństwo, że osoba badana jest związana ze zdarzeniem w zakresie pytań krytycznych wynosi więcej niż 99,9%.

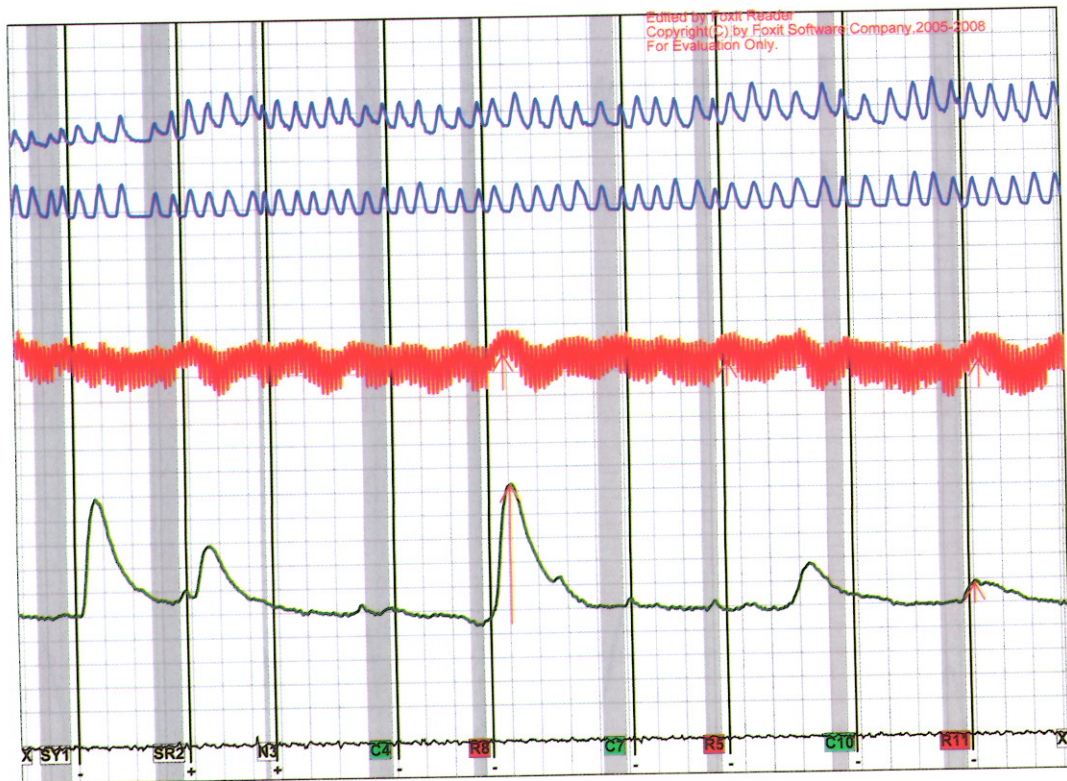
W ocenie zapisów biegli mogą komplementarnie stosować algorytmy komputerowe, które w dużej mierze oparte są na tych samych rozwiązaniach, co systemy ręcznego zliczania reakcji. Z piśmiennictwa wynika, że narzędzia komputerowe do analizy danych psychofizjologicznych cechuje większa dokładność niż tzw. ślepa analizę zapisów dokonywaną przez specjalistów<sup>23</sup>.

Poniżej zaprezentowano raporty wygenerowane przy użyciu komputerowego systemu analizy danych o nazwie OSS-3<sup>24</sup>, który wykorzystano jako narzędzie służące potwierdzeniu i obiektywizacji uzyskanych rezultatów.

Ostatni etap procesu badawczego wiąże się ze sformułowaniem wniosków końcowych opinii, które oparte zostaną na uzyskanych danych liczbowych oraz przyjętych progach punktowych dla decyzji DI oraz NDI. Ustaleniom ekspertyzy biegły musi teraz nadać kryminalistyczne znaczenie, które będzie akceptowalne dla zleceniodawcy w kontekście procesowym. Pomimo wypracowanych zasad analizy reakcji diagnostycznych i ich statystycznej interpretacji przekształcenie danych psychofizjologicznych na procesowo przydatne wnioski na gruncie polskiego prawa wydaje się niełatwe.







Ryc. 2. Trzy poligramy osoby „winnej”  
 Fig. 2. Three polygrams of „guilty” person

A

**Lafayette Instrument Company**  
**Objective Scoring System - Version 3**

By Raymond Nelson, Mark Handler and Donald Krapohl (2007)

**Result** No Significant Reactions  
**Description** There is < 0.1% chance this result was produced by a deceptive 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** Tuesday, June 25, 2013  
**Subject**  
**Examiner** System Administrator

Spot Scores		Decision Alpha (1 tailed)		Components	
ID	p-value	Setting	Value	Component	Weight
R5	< 0.001	NSR	0.100	Pneumo	0.19
R8	< 0.001	SR	0.050	EDA	0.53
R11	< 0.001	Bonferroni corrected alpha	0.017	Cardio	0.28
		Test of Proportions (1 tailed)	0.050		

B

**Lafayette Instrument Company**  
**Objective Scoring System - Version 3**

By Raymond Nelson, Mark Handler and Donald Krapohl (2007)

**Result** Significant Reactions  
**Description** There is < 0.1% chance 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** Tuesday, June 25, 2013  
**Subject**  
**Examiner** System Administrator

Spot Scores		Decision Alpha (1 tailed)		Components	
ID	p-value	Setting	Value	Component	Weight
R5	< 0.001	NSR	0.100	Pneumo	0.19
R8	< 0.001	SR	0.050	EDA	0.53
R11	< 0.001	Bonferroni corrected alpha	0.017	Cardio	0.28
		Test of Proportions (1 tailed)	0.050		

Ryc. 3. Raporty z komputerowej analizy danych fizjologicznych osób badanych; A – osoba „niewinna”, B – osoba „winna”  
 Fig. 3. Computerised evaluation reports of physiological data

## Podsumowanie

Istotnym atutem Empirycznego Systemu Oceny jest prosty w zastosowaniu zbiór reguł oceny zapisów reakcji psychofizjologicznych. Pozwala on na sprawną ocenę poligramów i stwarza możliwość szybkiej konfrontacji wniosków końcowych sformułowanych niezależnie przez różnych ekspertów oceniających te same zapisy.

Dbałość o przestrzeganie wymogów systemu zarządzania jakością, na którym opierają się policyjne laboratoria kryminalistyczne, nakazuje przeprowadzenie weryfikacji rzetelności każdej nowej metody badawczej, w tym Empirycznego Systemu Oceny, przed jej wdrożeniem do powszechnej praktyki. Dlatego prezentowany system oceny powinien stać się przedmiotem kompleksowych badań, które pozwolą ustalić jego wartość diagnostyczną na osobach z polskiego kręgu kulturowego.

Znaczny postęp w analizie zapisów testu CQT, szacowanie wiarygodności wyniku badania na podstawie danych normatywnych, a także rezultaty polskich badań walidacyjnych nad ESS, mogą w przyszłości pozwolić na zmianę istniejącej sytuacji, w której ekspert, aby uwiarygodnić zastosowaną metodę badawczą, może użyć jedynie danych odnośnie do jej wartości diagnostycznej.

Trafnie sformułowane wnioski opinii, wsparte statystyczną analizą, mogą zdecydować o wykorzystaniu opinii z badań poligraficznych w postępowaniu karnym. Określenie stopnia wiarygodności wyniku badania przyczyni się do pozytywnej oceny tego rodzaju dowodu. Dlatego prezentacja rezultatów ekspertyzy na tle współczesnych osiągnięć naukowych zwiększa szansę na jej uznanie w polskim środowisku prawniczym i zmniejsza ryzyko jej odrzucenia przez organy procesowe w toku oceny zgromadzonego w sprawie materiału.

## PRZYPISY

<sup>1</sup> Wójcikiewicz J.: CIT czy CQT, „Problemy Kryminalistyki” 2012, 275(1); Widacki J.: W sprawie wyboru techniki badania poligraficznego. Czy technika oparta na testach GKT (CIT) jest lepsza od techniki opartej na testach CQ?, „Problemy Kryminalistyki” 2011, 273(3).

<sup>2</sup> ASTM E2035-12 Standard Terminology Relating to Forensic Psychophysiology.

<sup>3</sup> Ze względu na cel badania poligraficznego testy dzielimy na diagnostyczne i skringowe. Te pierwsze wykorzystywane są w sprawach kryminalnych i dotyczą zdarzenia będącego przedmiotem prowadzonego postępowania. Testy przesiewowe zaś stosowane są m.in. dla celów kadrowych, dotyczą wielu różnych kwestii istotnych z punktu widzenia interesów danej służby i służą ujawnieniu zagadnień problematycznych, w które może być zaangażowana osoba badana.

<sup>4</sup> Nelson R., Show P., Handker M., Krapohl D.: Meta-Analytic Survey of Criterion Accuracy of Validated Polygraph

Techniques: The Ad-Hoc Committee on Validated Techniques, 2011.

<sup>5</sup> Kircher J.C., Packard T., Bell B.G., Bernhardt P.C.: Effects of Prior Demonstrations of Polygraph Accuracy on Outcomes of Propable-Lie and Directed-Lie Polygraph Test, „Polygraph” 2011, 39(1).

<sup>6</sup> Handler M., Nelson R.: Test Structure and Administration, „European Polygraph” 2008, 2(4).

<sup>7</sup> Honts C.R., Raskin D.C., Kircher, J.C.: Mental and physical countermeasures reduce the accuracy of polygraph tests. „Journal of Applied Psychology”, 1994, 79, 252–259.

<sup>8</sup> Nelson R., Handler M., Shaw P., Gougler M., Blalock B., Russell C., Cushman B., Oelrich M.: Using the Empirical Scoring System, „Polygraph” 2011,40(2).

<sup>9</sup> Krapohl D.J.: Polygraph Principles: A Literature Review, „Polygraph” 2013, 42(1).

<sup>10</sup> Nelson R., Blalock B., Oelrich M., Cushman B.: Reliability of the Empirical Scoring System with Expert Examiners, „Polygraph” 2011, 40(3).

<sup>11</sup> Ibidem.

<sup>12</sup> Cechy Kirchera to opracowany przez naukowców z Uniwersytetu Utah zestaw cech diagnostycznych o potwierdzonej empirycznie korelacji z pobudzeniem psychofizjologicznym rejestrowanym w trakcie badania poligraficznego. Wyznaczenie owych cech umożliwiło standaryzację procesu analizy zmian fizjologicznych oraz pozwoliło na ustalenie prawdopodobieństwa błędu dla testów CQT dla poszczególnych wyników punktowych. Cechy Kirchera posłużyły do opracowania komputerowych algorytmów zliczających wspomagających proces analizy danych – OSS 3 oraz CPS.

<sup>13</sup> Okno oceny reakcji – przedział czasowy, w którym musi pojawić się reakcja, aby mogła zostać uznana za diagnostyczną i podlegać ocenie numerycznej.

<sup>14</sup> Na potrzeby tego opracowania autorka będzie posługiwać się zwrotami DI, NDI i INC wyłącznie w rozumieniu technicznym dla uproszczenia sposobu prezentowania zagadnienia.

<sup>15</sup> Negatywne rozstrzygnięcie – NDI, stwierdzenie braku związku badanego ze sprawą.

<sup>16</sup> Pozytywne rozstrzygnięcie – DI, stwierdzenie związku badanego ze sprawą.

<sup>17</sup> Dane normatywne przedstawiają rozkład prawdopodobieństwa błędu ustalony osobno w grupie NDI oraz DI, w zależności od wyniku punktowego testu. Według autorów ESS dane normatywne podane w tabeli 1 reprezentatywne są dla każdej populacji i przy uwzględnieniu takiego założenia mogą być wykorzystywane do statystycznej oceny zapisów z badań poligraficznych także w Polsce. Niemniej teza ta wymaga weryfikacji w badaniach prowadzonych na populacji polskiej.

<sup>18</sup> Nelson R., Handler M.: Monte Carlo study of criterion validity of the Directed Lie Screening Test using Empirical Scoring System and the Objective Scoring System Version 3, „Polygraph” 2012, 41(3); Using Normative

Reference Data with Diagnostic Exams and the Empirical Scoring System, „Polygraph” 2012, 41(3).

19 Interpretacja statystyczna uzyskanego rozstrzygnięcia.

20 Reguła Sentera zakłada, że gdy wynik testu jest nierozstrzygnięty, interpretacji dokonuje się na podstawie ocen cząstkowych. Gdy któraś z nich jest mniejsza niż próg punktowy dla DI, całe badanie klasyfikuje się jako DI. Reguła Sentera redukuje poziom rozstrzygnięć INC o 60%. Krapohl D., Strum S.: Terminology Reference for the Science of Psychophysiological Detection of Deception, „Polygraph” 2002, 31(3).

21 Nelson R., Handler M., op.cit.

22 Przy podawaniu poziomu prawdopodobieństwa błędu dla wyniku testu opartego na najniższej ocenie cząstkowej jednego z pytań relewantnych należy odpowiadającą jej wartość  $p$  pomnożyć trzykrotnie, tu  $0,8\% \times 3 = 2,4\%$ . Nelson R., Handler M., op. cit.

23 Kircher J.C., Kristjansson S.D., Gardner M.K., Webb A.: Human and computer decision-making in the psychophysiological detection of deception, Final report to the Department of Defense Polygraph Institute (2005), University of Utah: Salt Lake City.

24 Objective Scoring System version 3 – Obiektywny System Oceny wersja 3, jego skuteczność wynosi 90,2%. Nelson R.: Criterion Validity of the Empirical Scoring System and the Objective Scoring System, version 3 with the USAF Modified General Question Technique, „Polygraph” 2011, 40(3).

## BIBLIOGRAFIA

1. ASTM E2035 – 12 Standard Terminology Relating to Forensic Psychophysiology.
2. Cushman B., Krapohl D.J.: Evidence for technical questions in polygraph techniques. Presentation at the APA annual Seminar, Myrtle Beach. Sc (2010).
3. Honts C.R., Raskin D.C., Kircher J.C.: Mental and physical countermeasures reduce the accuracy of polygraph tests. „Journal of Applied Psychology” 1994, 79, 252–259.
4. Kircher J.C., Raskin D.C.: Human versus computerized evaluations of polygraph data in a laboratory setting. „Journal of Applied Psychology” 1988, 73(2).
5. Kircher J.C., Packard T., Bell B.G., Bernhardt P.C.: Effects of Prior Demonstrations of Polygraph Accuracy on Outcomes of Propable-Lie and Directed-Lie Polygraph Test, „Polygraph” 2001, 39(1).
6. Kircher J.C., Kristjansson S.D., Gardner M.K., Webb A.: Human and computer decision-making in the psychophysiological detection of deception, Final report to the Department of Defense Polygraph Institute. (2005), University of Utah: Salt Lake City.
7. Widacki J.: W sprawie wyboru techniki badania poligraficznego. Czy technika oparta na testach GKT (CIT)

jest lepsza od techniki opartej na testach CQ?, „Problemy Kryminalistyki” 2011, 273(3).

8. Wójcikiewicz J.: CIT czy CQT?, „Problemy Kryminalistyki” 2012, 275(1).

9. Nelson R., Blalock B., Oelrich M., Cushman B.: Reliability of the Empirical Scoring System with Expert Examiners, „Polygraph” 2011, 40(3).

10 Nelson R., Handler M.: Using Normative Reference Data with Diagnostic Exams and the Empirical scoring System, „Polygraph” 41(3) 2012.

11. Nelson R., Handler M.: Empirical Scoring System: NPC Quick Reference, Lafayette Instrument.

12. Nelson R.: Criterion Validity of the Empirical Scoring System and the Objective Scoring System, version 3 with the USAF Modified General Question Technique, „Polygraph” 2011, 40(3).

14. Konieczny J.: Badania poligraficzne. Podręcznik dla zawodowców. Wydawnictwa Akademickie i Profesjonalne, Warszawa 2009.

### Streszczenie

Artykuł stanowi propozycję prezentowania wyników testów pytań porównawczych na podstawie Empirycznego Systemu Oceny oraz danych normatywnych, które pozwalają ustalić poziom wiarygodności danego rozstrzygnięcia. Tę prostą w zastosowaniu metodę oceny reakcji psychofizjologicznych zaprezentowano na zapisach z kryminalistycznych badań poligraficznych.

**Słowa kluczowe:** poligraf, ESS, ocena numeryczna

### Summary

The paper presents the way of presentation of Comparison Questions Test results using Empirical Scoring System and normative data that allow to determine the degree of reliability of the decision. This easy-to-use method for assessing psychophysiological reactions is presented on the basis of polygrams from forensic examinations.

**Keywords:** polygraph, ESS, numerical scoring