

Redakcja:
Bogdan Zawadzki
Krzysztof Fronczyk
Kamilla Bargiel-Matusiewicz

WYBRANE PROBLEMY
WSPÓŁCZESNEJ DIAGNOSTYKI KLINICZNEJ
W PSYCHOLOGII



Tom 4 Serii wydawniczej Wydziału Psychologii Uniwersytetu
Warszawskiego pt. *Psychological Currents. Theory and Research*

WYBRANE PROBLEMY
WSPÓŁCZESNEJ DIAGNOSTYKI KLINICZNEJ
W PSYCHOLOGII

Redakcja:
Bogdan Zawadzki
Krzysztof Fronczyk
Kamilla Bargiel-Matusiewicz

WYBRANE PROBLEMY
WSPÓŁCZESNEJ DIAGNOSTYKI KLINICZNEJ
W PSYCHOLOGII



Seria wydawnicza Wydziału Psychologii Uniwersytetu Warszawskiego
Psychological Currents. Theory and Research.

Tom 4

Wybrane problemy współczesnej diagnostyki klinicznej w psychologii
Redakcja naukowa: Bogdan Zawadzki, Krzysztof Fronczyk, Kamilla Bargiel-
-Matusiewicz

Recenzenci:

prof. dr hab. Jerzy Brzeziński (Uniwersytet im. Adama Mickiewicza w Poznaniu)
dr hab. Włodzimierz Strus, prof. uczelni (Uniwersytet Kardynała Stefana
Wyszyńskiego w Warszawie)

Korekta językowa:

Karolina Rymarczyk

Projekt okładki:

Karolina Rymarczyk

Ilustracja na okładce:

shutterstock.com; autor: Alexander Supertramp

Skład i łamanie:

Studio DTP Academicon | Patrycja Waleszczak, Adam Dorot
dtp@academicon.pl, dtp.academicon.pl



Publikacja jest udostępniona na licencji Creative Commons Uznanie autorstwa 4.0
Treść licencji jest dostępna na stronie: <https://creativecommons.org/licenses/by/4.0/>

Publikacja finansowana ze środków Wydziału Psychologii Uniwersytetu Warszawskiego
przyznanych przez MNiSW w formie subwencji na utrzymanie i rozwój potencjału
badawczego w roku 2024 (501-D125-01-1250000 zlec. 5011000632).

Warszawa: Wydawnictwo Liberi Libri, 2024

www.LiberiLibri.pl

ISBN: 978-83-63487-67-6

DOI: 10.47943/lib.9788363487676

SKRÓCONY SPIS TREŚCI

BOGDAN ZAWADZKI, KRZYSZTOF FRONCZYK, KAMILA BARGIEL-MATUSIEWICZ	
Wstęp	13
 CZĘŚĆ I: WYBRANE ZAGADNIENIA ORAZ WSPÓŁCZESNE WYZWANIA DIAGNOSTYKI KLINICZNEJ	 17
EMILIA ŁOJEK	
Modernizacja metod i procedur badania w praktyce klinicznej wyzwaniem dla współczesnej neuropsychologii	 19
AGNIESZKA E. ŁYŚ, KAMILA BARGIEL-MATUSIEWICZ	
Diagnoza psychologiczna związana z postępowaniem dowodowym w sytuacjach przemocy seksualnej – wyzwanie dla współczesnej psychologii sądowej	 47
ANNA JĘDRASZKIEWICZ	
Zastosowanie modeli psychiatrii obliczeniowej do badań nad PTSD	65
MICHAŁ CHRUSZCZEWSKI	
Nuda: opis zjawiska, konotacje kliniczne i pomiar	101
 CZĘŚĆ II: PROBLEMY DIAGNOZY KLINICZNEJ I PRZEGLĄD NARZĘDZI BADAWCZYCH W ZAKRESIE NEUROPSYCHOLOGII ORAZ PSYCHOFIZJOLOGII	 123
DOMINIKA ŻARNECKA, EMILIA ŁOJEK	
Zaburzenia węchu w perspektywie neuropsychologicznej: znaczenie kliniczne i przegląd metod badania węchu	 125

MARTA SIEPSIAK, WOJCIECH ŁUKASZ DRAGAN, NORA UGLIK-MARUCHA

Diagnoza objawów mizofonii w badaniach naukowych i praktyce klinicznej

– analiza porównawcza dwóch kwestionariuszy dostępnych w języku polskim

– *MisoQuest* i *S-Five* 155

CZĘŚĆ III: PROBLEMY PSYCHOMETRYCZNE POMIARU KLINICZNEGO

W ODNIESIENIU DO ZABURZENIA STRESOWEGO POURAZOWEGO (PTSD) 177

MARIA CYNIAK-CIECIURA, AGNIESZKA POPIEL, PATRYK BIELAK, VICTORIA CAMINO, EUN JUNG CHA, YUNKYUNG CHO, JULIE K. CREMEANS-SMITH, DAVID B. FRUEHSTORFER, MARINA GALARREGUI, ROCÍO GOLDFARB, MYOUNG HO HYUN, ZHANNA KALININA, EDUARDO KEEGAN, ALIYA MAMBETALINA, LOUISE MCHUGH, MARIANA MIRACCO, ATSUSHI OSHIO, CHOWON PARK, ANDRÉS PARTARRIEU, LORENA DE ROSA, RAIKHAN SABIROVA, ADIL SAMEKIN, EMILIANO SÁNCHEZ, MARIA SARNO, CECILIA TARRUELLA, GULMIRA TUYAKOVNA TOPANOVA, GULMIRA M. TULEKOVA, BOGDAN ZAWADZKI

Czy pomiar objawów PTSD za pomocą *PTSD Checklist dla DSM-5* (PCL-5)

jest równoważny kulturowo? Raport z badania przeprowadzonego w ośmiu kulturach

w czasie pandemii COVID-19 179

BOGDAN ZAWADZKI, SZYMON BARCIŚ, MARIA CYNIAK-CIECIURA

Struktura objawów PTSD – analiza równoważności wersji inwentarza PCL-5

o stałej i losowej kolejności pozycji 217

SZCZEGÓŁOWY SPIS TREŚCI

BOGDAN ZAWADZKI, KRZYSZTOF FRONCZYK, KAMILLA BARGIEL-MATUSIEWICZ

Wstęp	13
 CZĘŚĆ I: WYBRANE ZAGADNIENIA ORAZ WSPÓŁCZESNE WYZWANIA DIAGNOSTYKI KLINICZNEJ	 17
 EMILIA ŁOJEK	
Modernizacja metod i procedur badania w praktyce klinicznej wyzwaniem dla współczesnej neuropsychologii	19
Streszczenie	20
Abstract	20
Wprowadzenie	21
Różne podejścia do celu badania i narzędzi neuropsychologicznych	22
Podejście kliniczno-eksperymentalne	23
Podejście psychometryczne	25
Praktyka kliniczna vs badania naukowe	26
Przyczyny konserwatyzmu	30
Od odkrycia naukowego do praktyki klinicznej	30
Standardy psychometryczne	31
Ścieżki rozwoju	33
Integracja wiedzy naukowej	33
Nowoczesna psychometria	35
Monitorowanie sprawności w życiu codziennym	40
Podsumowanie i wnioski	41
Bibliografia	42

AGNIESZKA E. ŁYŚ, KAMILLA BARGIEL-MATUSIEWICZ

Diagnoza psychologiczna związana z postępowaniem dowodowym w sytuacjach przemocy seksualnej – wyzwanie dla współczesnej psychologii sądowej	47
Streszczenie	48
Abstract	48
Ocena psychologicznej wiarygodności zeznań – możliwości i ograniczenia	50
Narzędzia do diagnozowania przemocy seksualnej	53
Osoby z niepełnosprawnościami jako ofiary przemocy seksualnej	55
Inne trudności związane z diagnozą przemocy seksualnej	56
Podsumowanie	58
Bibliografia	59

ANNA JĘDRASZKIEWICZ

Zastosowanie modeli psychiatrii obliczeniowej do badań nad PTSD	65
Streszczenie	66
Abstract	66
Psychiatria obliczeniowa – wstępna charakterystyka	68
Zarys podejścia	68
Uczenie ze wzmacnianiem	71
Bayesowskie podejścia do poznania	73
Wartość dodana	75
Wstępne wyniki badań i propozycje teoretyczne na przykładzie PTSD	77
Różnice neuroobliczeniowe w uczeniu	78
<i>Cel i wstępne informacje o badaniu</i>	78
<i>Specyfika modelowania</i>	80
<i>Wyniki badania</i>	84
PTSD z perspektywy przetwarzania predykcyjnego	86
<i>Podstawy teoretyczne</i>	86
<i>Dysfunkcjonalność przetwarzania informacji w PTSD</i>	87
<i>Kontekst dotychczasowych badań</i>	88
Komentarze krytyczne	89
Funkcje modeli obliczeniowych	89
Nieścisłość terminologiczna	90
Wnioski końcowe	92
Podziękowania	94
Bibliografia	94

MICHAŁ CHRUSZCZEWSKI

Nuda: opis zjawiska, konotacje kliniczne i pomiar	101
Streszczenie	102
Abstract	102
Próby określenia znaczenia pojęcia <i>nuda</i>	103
Kliniczny i subkliniczny kontekst nudy	105
Rozpowszechnienie nudy i jej pomiar	107
Wykaz kwestionariuszowych miar nudy	108
Przegląd wybranych kwestionariuszy mierzących nudę	110
Perspektywy dalszych badań nad nudą	115
Bibliografia	116

<p>CZĘŚĆ II: PROBLEMY DIAGNOZY KLINICZNEJ I PRZEGLĄD NARZĘDZI BADAWCZYCH W ZAKRESIE NEUROPSYCHOLOGII ORAZ PSYCHOFIZJOLOGII</p>	123
--	-----

DOMINIKA ŻARNECKA, EMILIA ŁOJEK

<p>Zaburzenia węchu w perspektywie neuropsychologicznej: znaczenie kliniczne i przegląd metod badania węchu</p>	125
Streszczenie	126
Abstract	126
Znaczenie węchu w perspektywie neuropsychologicznej	127
Dysfunkcje węchu w zaburzeniach neurodegeneracyjnych	129
Dysfunkcje węchu w zaburzeniach psychicznych	130
Depresja	131
Schizofrenia	132
Przegląd metod badania węchu	133
Metody samoopisowe	134
Questionnaire of Olfactory Disorders (<i>QOD</i>)	134
The Multi-Clinic Smell and Taste Questionnaire (<i>MCSTQ</i>)	135
Importance of Olfaction Questionnaire (<i>IOQ</i>)	135
The Relational Scale of Olfaction (<i>EROL</i>)	136
The Self-administered Odor Questionnaire (<i>SAOQ</i>)	136
Hyposmia Rating Scale (<i>HRS</i>)	137
The Affective Impact of Odor Scale (<i>AIO</i>)	137
The Odor Awareness Scale (<i>OAS</i>)	138
Kwestionariusze służące ocenie wrażliwości na nieprzyjemne zapachy	138
<i>Zapachy chemiczne</i>	138

<i>Zapachy wydzielane przez człowieka</i>	139
Narzędzia przesiewowe	140
<i>Skrócona wersja Questionnaire of Olfactory Disorders (QOD-NS)</i>	140
Self-reported Mini Olfactory Questionnaire (<i>Self-MOQ</i>)	140
Metody testowe	141
<i>SST</i>	141
<i>SST – wersja przesiewowa</i>	142
<i>UPSIT</i>	142
Brief Smell Identification Test (<i>BSIT</i>) i Pocket Smell Test (<i>PST</i>) – wersje przesiewowe	143
Pozostałe narzędzia testowe	143
Badania węchu retronosowego	144
Metody fizjologiczne	145
Podsumowanie	145
Bibliografia	147

MARTA SIEPSIAK, WOJCIECH ŁUKASZ DRAGAN, NORA UGLIK-MARUCHA

Diagnoza objawów mizofonii w badaniach naukowych i praktyce klinicznej – analiza porównawcza dwóch kwestionariuszy dostępnych w języku polskim – <i>MisoQuest</i> i <i>S-Five</i>	155
Streszczenie	156
Abstract	157
Definicja i krótka historia rozwoju rozumienia mizofonii	158
Charakterystyka dwóch narzędzi do pomiaru mizofonii dostępnych obecnie w języku polskim	161
Tworzenie, charakterystyka i znaczenie <i>MisoQuest</i>	161
Tworzenie, charakterystyka i znaczenie <i>S-Five</i>	166
<i>MisoQuest</i> i <i>S-Five</i> w badaniach naukowych i praktyce klinicznej – podsumowanie	171
Bibliografia	172

CZĘŚĆ III: PROBLEMY PSYCHOMETRYCZNE POMIARU KLINICZNEGO W ODNIESIENIU DO ZABURZENIA STRESOWEGO POURAZOWEGO (PTSD)

MARIA CYNIAK-CIECIURA, AGNIESZKA POPIEL, PATRYK BIELAK, VICTORIA CAMINO, EUN JUNG CHA, YUNKYUNG CHO, JULIE K. CREMEANS-SMITH, DAVID B. FRUEHSTORFER, MARINA GALARREGUI, ROCÍO GOLDFARB, MYOUNG HO HYUN, ZHANNA KALININA, EDUARDO KEEGAN, ALIYA MAMBETALINA, LOUISE MCHUGH, MARIANA MIRACCO, ATSUSHI OSHIO, CHOWON PARK, ANDRÉS PARTARRIEU, LORENA DE ROSA, RAIKHAN SABIROVA, ADIL SAMEKIN, EMILIANO SÁNCHEZ, MARIA SARNO, CECILIA TARRUELLA, GULMIRA TUYAKOVNA TOPANOVA, GULMIRA M. TULEKOVA, BOGDAN ZAWADZKI

Czy pomiar objawów PTSD za pomocą <i>PTSD Checklist dla DSM-5 (PCL-5)</i> jest równoważny kulturowo? Raport z badania przeprowadzonego w ośmiu kulturach w czasie pandemii COVID-19	179
---	-----

Streszczenie	182
Abstract	183
Struktura objawów zaburzenia stresowego pourazowego	184
<i>PTSD Checklist dla DSM-5</i>	187
Równoważność pomiarowa	188
Cele i hipotezy badawcze	190
Metoda	191
Procedura	191
Uczestnicy	191
Narzędzia	194
Procedura analizy danych	195
Wyniki	196
Dyskusja	203
Bibliografia	205
Załączniki	211

BOGDAN ZAWADZKI, SZYMON BARCIŚ, MARIA CYNIAK-CIECIURA

Struktura objawów PTSD – analiza równoważności wersji inwentarza PCL-5 o stałej i losowej kolejności pozycji	217
Streszczenie	218
Abstract	218
Problem badań, pytania i hipotezy badawcze	224
Metoda	225
Zastosowane narzędzia badawcze	225
<i>Inwentarz PCL-5</i>	225
<i>Dodatkowe narzędzia badawcze</i>	228
Analizy wstępne	228
Badania właściwe	231
Wyniki	238
Dyskusja	245
Bibliografia	250

Wstęp

Zapraszamy Czytelników do zapoznania się z kolejnym tomem z serii wydawniczej Wydziału Psychologii Uniwersytetu Warszawskiego pt. *Psychological Currents. Theory and Research*. W bieżącym tomie Czytelnicy znajdą osiem prac poświęconych wybranym zagadnieniom współczesnej psychologicznej diagnostyki klinicznej. Książka ta powstała w celu przekazania wiedzy na temat aktualnych wyzwań, jakie napotykają klinicyści oraz naukowcy w obszarze diagnostyki psychologicznej zarówno na poziomie specyficznych kontekstów praktyki zawodowej, jak i na poziomie ogólnych tendencji i możliwości związanych z postępowaniem naukowym, ze zmianami społeczno-kulturowymi i technologicznymi oraz z refleksami na poziomie etycznym. Obszary, w obrębie których prezentowane są poszczególne zagadnienia związane z diagnozą, to: neuropsychologia, psychofizjologia, psychologia sądowa, psychiatria, psychologia emocji, osobowości i różnic indywidualnych oraz procesów poznawczych.

Część pierwsza tomu poświęcona jest wyzwaniom diagnostyki klinicznej w obszarach neuropsychologii, psychologii sądowej oraz psychiatrii. Rozdział autorstwa Emilii Łojek zatytułowany *Modernizacja metod i procedur badania w praktyce klinicznej wyzwaniem dla współczesnej neuropsychologii* prezentuje dyskusję, która toczy się współcześnie w środowisku neuropsychologów na temat modernizacji metod badania stosowanych w praktyce klinicznej. Autorka analizuje przyczyny zjawiska stosunkowo wolno postępującego wprowadzania do praktyki klinicznej nowych narzędzi. Wskazuje procesy umożliwiające modernizację i tworzenie nowych metod oraz ich ograniczenia. Praca Agnieszki E. Łyś i Kamilli Bargiel-Matusiewicz zatytułowana *Diagnoza psychologiczna związana z postępowaniem dowodowym w sytuacjach przemocy seksualnej – wyzwanie dla współczesnej psychologii sądowej* dotyczy zagadnień diagnozy psychologicznej w kontekście perspektywy prawnej oraz społecznej zjawiska przemocy seksualnej. Analizie zostają poddane możliwości i ograniczenia metod analizy treści zeznań, takich jak *Statement Validity Assessment* (SVA), *Multivariable Adults' Statement Assessment Method* (MASAM), *Psychologically*

Based Credibility Assessment Tool (PBCAT) czy *Global Evaluation System* (GES). W rozdziale opisane są także narzędzia do diagnozowania wystąpienia przemocy seksualnej, jak *Child Sexual Behavior Inventory-2* (CSBI-2) – narzędzie służące do oceny prawdopodobieństwa wystąpienia doświadczenia wykorzystania seksualnego u dzieci i *Pict-o-Stat* – narzędzie służące do oceny prawdopodobieństwa wystąpienia doświadczenia wykorzystania seksualnego u osób z niepełnosprawnościami. Tekst Anny Jędraszekiewicz pt. *Zastosowanie modeli psychiatrii obliczeniowej do badań nad PTSD* dotyczy nowego interdyscyplinarnego podejścia do badania zaburzeń psychicznych opartego na tworzeniu modeli obliczeniowych. Tekst wprowadza podstawowe założenia teoretyczne podejścia, które skupia się na wyjaśnianiu konkretnych dysfunkcji za pomocą precyzyjnych metod matematycznych i symulacji komputerowych. Krytykuje jednocześnie nadmierne koncentrowanie się na opracowywaniu klasyfikacji w systemach psychiatrycznych, podkreślając, że lepsze zrozumienie fundamentalnych zjawisk z perspektywy przetwarzania informacji ułatwi formułowanie trafniejszych diagnoz. W rozdziale omówiono dwie ramy teoretyczne psychiatrii obliczeniowej: uczenie ze wzmocnieniem oraz wnioskowanie bayesowskie. Obie stanowią ogólne zbiory założeń teoretycznych określających relacje między terminami matematycznymi i ich odpowiednikami w postaci konstruktów psychologicznych. Konkretnie zastosowania w badaniach opisano na przykładzie pourazowego zaburzenia stresowego (PTSD).

Część druga tomu poświęcona jest diagnozie klinicznej w obszarze neuropsychologii oraz psychofizjologii. Dominika Żarnecka oraz Emilia Łojek w pracy pt. *Zaburzenia węchu w perspektywie neuropsychologicznej: znaczenie kliniczne i przegląd metod badania węchu* omawiają psychologiczną rolę węchu oraz metody pomiaru jego funkcjonowania. Szczególną rolę zwrócono na przedstawienie specyfiki dysfunkcji węchu w zaburzeniach neurodegeneracyjnych i psychicznych. W artykule, na wybranych przykładach, omówiono także różne metody badania węchu, włączając w to zarówno samoopisowe kwestionariusze, a także testy wykorzystujące rzeczywiste bodźce zapachowe oraz metody fizjologiczne dotyczące pomiaru reakcji organizmu (np. potencjały wywołane fMRI, PET itd.) na rzeczywiste bodźce zapachowe. W artykule przedstawiono zalety i ograniczenia poszczególnych grup metod. Marta Siepsiak, Wojciech Łukasz Dragan oraz Nora Uglik-Marucha w rozdziale zatytułowanym *Diagnoza mizofonii w badaniach naukowych i praktyce klinicznej – analiza porównawcza dwóch kwestionariuszy dostępnych w języku polskim – MisoQuest i S-Five* omawiają dwa narzędzia przeznaczone do diagnozy słabo poznanego zjawiska mizofonii, będącej

nadwrażliwością na powtarzalne dźwięki, zwłaszcza związane z jedzeniem, wywołującą negatywne emocje o ekstremalnym nasileniu, znacząco wpływając na codzienne funkcjonowanie. Autorzy zwracają uwagę na zróżnicowane korelaty psychofizjologiczne i neurobiologiczne mizofonii, podkreślając rolę przetwarzania poznawczego w nasileniu objawów, a jednocześnie omawiają proponowane kryteria diagnostyczne mizofonii, która nie jest jeszcze oficjalnie uwzględniona w żadnych klasyfikacjach diagnostycznych. Przedstawione przez Autorów dwa narzędzia do pomiaru mizofonii, *MisoQuest* i *S-Five*, stanowią więc tu istotny postęp. Krytyczne omówienie zalet i ograniczeń obu narzędzi może przyczynić się do lepszego diagnozowania mizofonii zarówno w badaniach naukowych, jak i praktyce klinicznej.

Część trzecia tomu poświęcona jest kwestii równoważności pomiaru klinicznego w różnych grupach czy zastosowaniach narzędzi diagnostycznych. Prace prezentowane w tej części dotyczą objawów pourazowego zaburzenia stresowego, badanego za pomocą *PTSD Checklist dla DSM-5 (PCL-5)* – jednego z najbardziej obecnie popularnych narzędzi samoopisowych przeznaczonych do badania tego zaburzenia. Pomimo tego zawężenia mają jednak szersze znaczenie, bowiem podnoszą problem równoważności psychometrycznej różnych wersji narzędzi stosowanych do diagnozy czy równoważności międzygrupowej pomiaru. Tekst Marii Cyniak-Cieciury i in. pt. *Czy pomiar objawów PTSD za pomocą PTSD Checklist dla DSM-5 (PCL-5) jest ekwiwalentny kulturowo? Raport z badania przeprowadzonego w ośmiu kulturach w czasie pandemii COVID-19* powstał w ramach badania COVID-TEMP i dotyczy równoważności psychometrycznej PCL-5 w próbach pochodzących z ośmiu krajów. Wyniki analiz międzykulturowych wskazują, że struktura czynnikowa jest równoważna, podobnie jak za równoważne można uznać ładunki czynnikowe, co wskazuje, że w każdej próbie zbadano ten sam konstrukt. Nie potwierdzono natomiast równoważności skalarnej i rezydualnej, co czyni bezprzedmiotowym dokonywanie porównań międzykulturowych. Jak stwierdzają Autorzy, inne (nieliczne!) badania też wskazują na takie ograniczenie pomiarowe narzędzi klinicznych i kwestia ta w każdym przypadku wymaga głębszej refleksji metodologicznej i odpowiednich analiz psychometrycznych. Ostatni tekst przygotowany przez Bogdana Zawadzkiego, Szymona Barcisia i Marię Cyniak-Cieciurę także dotyczył równoważności inwentarza PCL-5, ale tym razem w odniesieniu do wersji o stałej i losowej kolejności pozycji. Popularyzacja badań online i możliwość losowania kolejności narzędzi i/lub nawet pozycji oferowana przez platformy internetowe, np. Qualtrics, przyczyniła się do szerokiego wyko-


rzystywania tych możliwości. Efekt ten był zweryfikowany w prezentowanej pracy z zastosowaniem wersji o stałej (jak dla formy „papier-ołówek”) oraz losowej kolejności pozycji w tej samej grupie 740 osób badanych dwukrotnie w odstępie dwóch tygodni (poprzez panel Ariadna) w planie: wersja stała-losowa albo losowa-stała. Analiza z zastosowaniem bardzo restryktywnych kryteriów równoważności wykazała pełną równoważność obu pomiarów, ale jedynie równoważność konfiguralną dla wersji o odmiennej kolejności pozycji (zbieżność struktury czynnikowej, ale już nierówność ładunków itp.). W konkluzji Autorzy adresują problem weryfikacji równoważności pomiarowej nie tylko narzędzi klinicznych, ale także przeznaczonych do badania odmiennych konstruktów psychologicznych, a także różnych wersji tych samych narzędzi, np. tradycyjnych vs online czy ze stałą vs odmienną kolejnością pozycji (jak też ogólnie sensu praktyki grupowania czy losowania pozycji).

Książka ta jest adresowana zasadniczo do psychologów oraz psychiatrów, neurologów i – ogólniej – badaczy prowadzących dociekania naukowe w obszarze różnic indywidualnych, psychologii klinicznej, zdrowia i jakości życia. Teksty są zaawansowane i ich recepcja wymaga dobrego przygotowania teoretycznego, metodologicznego oraz psychometrycznego. Zachęcamy jednak do zapoznania się z nimi wszystkich Czytelników zainteresowanych zagadnieniami diagnozy psychologicznej, żywiąc nadzieję, że wartość informacyjna prezentowanych tekstów w pełni wynagrodzi trud włożony w ich lekturę.

W zakończeniu chcielibyśmy wyrazić naszą wdzięczność Panom Profesorom: Jerzemu Marianowi Brzezińskiemu i Włodzimierzowi Strusowi za niezwykle wnikliwe recenzje oraz cenne uwagi i sugestie dotyczące zarówno poszczególnych tekstów, jak też tomu jako całości.

Bogdan Zawadzki,
Krzysztof Fronczyk
i Kamilla Bargiel-Matusiewicz

CZĘŚĆ I: WYBRANE ZAGADNIENIA ORAZ WSPÓŁCZESNE WYZWANIA DIAGNOSTYKI KLINICZNEJ

EMILIA ŁOJEK  <https://orcid.org/0000-0002-0809-5601>
Uniwersytet Warszawski, Wydział Psychologii

**Modernizacja metod i procedur badania w praktyce
klinicznej wyzwaniem dla współczesnej neuropsychologii**
*Modernization of assessment methods and procedures
in clinical practice as a challenge for current neuropsychology*

Autor korespondencyjny: Emilia Łojek
e-mail: emilia@psych.uw.edu.pl

Sposób cytowania: Łojek, E. (2024). Modernizacja metod i procedur badania w praktyce klinicznej wyzwaniem dla współczesnej neuropsychologii. W: B. Zawadzki, K. Fronczyk i K. Bargiel-Matusiewicz (red.), *Wybrane problemy współczesnej diagnostyki klinicznej w psychologii* (s. 19–46). Wydawnictwo Liberi Libri. <https://doi.org/10.47943/lib.9788363487676.rozdzial01>

Streszczenie

Niniejszy rozdział prezentuje dyskusję, która toczy się w środowisku neuropsychologów na temat modernizacji metod badania stosowanych w praktyce klinicznej. W odróżnieniu od obszaru badań naukowych, praktykę kliniczną w neuropsychologii charakteryzuje duży opór przed modernizacją narzędzi pomiaru. Korzenie technik badania neuropsychologicznego powszechnie stosowanych w praktyce klinicznej nawet w krajach wysoko rozwiniętych sięgają niekiedy XIX wieku. Istnieje wiele przyczyn tego zjawiska. Jedną z nich jest troska o zachowanie najwyższych standardów psychometrycznych i klinicznych ustalonych dla metod oraz procedur badania diagnostycznego, wypracowanych w toku rozwoju neuropsychologii jako nauki. Modernizacja i tworzenie nowych metod, które będą miały zastosowanie w praktyce klinicznej może dokonywać się: (1) w efekcie integracji neuronauk i nauk o zdrowiu, (2) dzięki wdrażaniu nowoczesnych algorytmów psychometrycznych i matematycznych, w tym sztucznej inteligencji oraz (3) poprzez dane pozyskiwane przez urządzenia techniczne, z których korzysta pacjent w życiu codziennym. Każda z tych potencjalnych dróg rozwoju ma swoje ograniczenia. Akceptując założenie o konieczności modernizacji metod pomiaru służących diagnozie neuropsychologicznej w praktyce klinicznej, nie należy zapominać, że proces ten nie może być celem nadrzędnym.

Słowa kluczowe: neuropsychologia, metody badania diagnostycznego, psychometria

Abstract

This chapter presents the ongoing discussion among neuropsychologists about the modernization of research methods used in clinical practice. Unlike the area of scientific research, the clinical practice of neuropsychology is characterized by great resistance to the modernization of assessment methods. The roots of neuropsychological measurement methods commonly used in clinical practice, even in highly developed

countries, sometimes date back to the 19th century. There are many reasons for this phenomenon. One of them is the concern to maintain the highest psychometric and clinical standards established for diagnostic assessment methods and procedures, established in the course of the development of neuropsychology as a science. Modernization and creation of new methods that will be used in clinical practice may be achieved: (1) as a result of the integration of neurosciences and health sciences, (2) thanks to the implementation of modern psychometric and mathematical algorithms, including artificial intelligence, and (3) through data obtained by technical devices used by the patient in everyday life. Each of these potential development paths has its limitations. While accepting the assumption of the need to modernize measurement methods for neuropsychological diagnosis in clinical practice, we should not forget that this process cannot be the primary goal.

Keywords: neuropsychology, diagnostic assessment methods, psychometrics

WPROWADZENIE

W ciągu kilku ostatnich dziesięcioleci dokonał się olbrzymi postęp w naukach o mózgu oraz jego funkcjach u osób zdrowych i chorych. Co istotne, w znacznym stopniu wiąże się on z wprowadzaniem i szerokim stosowaniem innowacyjnych metod pomiaru, procedur i technologii, co warunkuje zmianę paradygmatów badawczych, pozwala na zbieranie zupełnie nowych danych, a w efekcie prowadzi do weryfikacji dotychczasowych koncepcji teoretycznych. Od drugiej połowy XIX wieku do lat 80. XX wieku głównym źródłem wiedzy o relacji pomiędzy strukturami mózgu a zachowaniem były badania, w których za pomocą testów i prób kliniczno-eksperymentalnych mierzono zaburzenia czynności psychicznych u chorych z uszkodzeniami mózgu (leżjami). Zmiana tego paradygmatu i progres w wiedzy naukowej rozpoczął się wraz z wprowadzeniem technik funkcjonalnego neuroobrazowania, które umożliwiło badanie aktywności ośrodkowego układu nerwowego (OUN) u osób zdrowych i chorych podczas wykonywania określonej czynności. Wyniki tych prac

przyczyniły się do weryfikacji koncepcji relacji mózg–zachowanie opartych na wynikach badań leżyjnych, przenosząc uwagę badaczy z funkcjonalnego znaczenia anatomicznie wyodrębnionych struktur OUN na poziom sieci neuronalnych całego mózgowia, których organizacja czynnościowa okazała się znacznie bardziej złożona niż kiedykolwiek przypuszczano. Funkcjonalne neuroobrazowanie umożliwiło także dostrzeżenie niewykrywalnych wcześniej dyskretnych aberracji pracy OUN, które współwystępują z dysfunkcjami obserwowalnymi na poziomie behawioralnym. Odkrycia te pomogły w lepszym zrozumieniu patomechanizmów zaburzeń czynności psychicznych. W chwili obecnej stoimy przed kolejną rewolucją w wiedzy o człowieku i jego zdrowiu psychicznym. Dokonuje się ona za sprawą postępującej integracji nauk oraz technicznych możliwości pozyskiwania i gromadzenia wielkich, wielopozomowych danych poczynając od poziomu molekularnego, genetycznego, fizjologicznego poprzez sieci neuronalne i struktury OUN, aż do złożonych czynności psychicznych i procesów społecznych (Parsons i Duffield, 2019). Algorytmy analizy tych danych stanowią obecnie duże wyzwanie dla nauki, w czym również wspomaga nas nowoczesna technologia – sztuczna inteligencja, maszyny uczące się (Bilder, 2011).

W jakim stopniu te rewolucyjne zmiany metodologii badań i koncepcji relacji mózg–zachowanie przekładają się na praktykę kliniczną z zakresu neuropsychologii? Od kilku lat w środowisku neuropsychologów trwa dyskusja na ten temat, gdyż przejście od badań naukowych do praktyki klinicznej w neuropsychologii okazuje się nadzwyczaj trudne (Ardila, 2013; Casaletto i Heaton, 2017; Łojek, 2014, 2019; Marcopulos i Łojek, 2016, 2019). W poniższym opracowaniu podjęto próbę opisu kształtowania się tego zjawiska, analizy jego przyczyn oraz znalezienia rozwiązań, które pomogłyby przejść neuropsychologii stosowanej w XXI wiek.

RÓŻNE PODEJŚCIA DO CELU BADANIA I NARZĘDZI NEUROPSYCHOLOGICZNYCH

Aby zrozumieć źródła obecnego kryzysu należy sięgnąć do początków neuropsychologii, czyli pierwszej połowy XX wieku, w której w miarę niezależnie rozwinęły się dwa główne podejścia do badania diagnostycznego pacjentów z uszkodzeniami mózgu: kliniczno-eksperymentalne i psychometryczne.

Podejście kliniczno-eksperymentalne

Twórcą podejścia kliniczno-eksperymentalnego był Aleksander Łuria, autor teorii układów funkcjonalnych mózgu oraz zasad i metod badania neuropsychologicznego opartych na tej teorii (Łuria, 1967). Łuria zakładał, że podstawę dla każdej czynności umysłowej stanowi zintegrowana praca różnych obszarów kory mózgu, tworzących układy funkcjonalne, w których każdy obszar odpowiedzialny jest za określony aspekt danej czynności. Poznanie roli każdego obszaru mózgu w układzie funkcjonalnym wymagało według Łurii analizy syndromologicznej, polegającej na poszukiwaniu wspólnych cech (deficytu podstawowego) w obrębie zespołu zaburzeń. Na przykład zaburzenia fonetyczne jako deficyt podstawowy u pacjenta z afazją po uszkodzeniu okolicy Wernickego w lewym płacie skroniowym, zdaniem Łurii, będą mogły przejawiać się wspólnymi cechami w zaburzeniach wielu różnych, złożonych czynności umysłowych, w których zaangażowane są zdolności rozumienia znaczenia dźwięków mowy (mówieniu, czytaniu, pisaniu, liczeniu, powtarzaniu itd). Rezultaty analizy syndromologicznej, wnioskowanie o deficycie podstawowym, czynnościach zaburzonych i zachowanych dawały podstawę do diagnozy, a następnie rehabilitacji neuropsychologicznej chorego. Ustalanie zbliżonej lokalizacji uszkodzenia OUN u wielu chorych z określonym deficytem podstawowym w ramach zespołu zaburzeń stanowiło wskaźnik roli czynnościowej tego obszaru w danym układzie funkcjonalnym i miało znaczenie dla weryfikacji założeń teoretycznych o relacji mózg–zachowanie.

Z założenia procedura badania oraz próby/zadania opracowane przez Łurię były niewystandaryzowane. Cechowała je plastyczność i indywidualizowane podejście do chorego. Łuria koncentrował się nie tylko na końcowym rezultacie wykonania próby i ogólnym poziomie deficytu, ale również na sposobach dochodzenia do rozwiązania. Ten sposób badania sprzyjał zrozumieniu prawdopodobnych mechanizmów obserwowanych zaburzeń, które Łuria interpretował w odniesieniu do przyjętej teorii organizacji czynnościowej kory mózgu. Dostosowanie badania do poziomu funkcjonowania i potrzeb pacjenta pomagało także ukazać zachowane możliwości chorego, co nie jest łatwe w przypadkach pacjentów z głębokimi deficytami. Wiarygodność (rzetelność) badania neuropsychologicznego opierała się według Łurii na efektywności analizy syndromologicznej, czyli powtarzalnym wykrywaniu deficytu podstawowego w różnych czynnościach umysłowych u danego chorego.

Koncepcja układów funkcjonalnych oraz założenia badania chorych z uszkodzonym mózgiem opracowane przez Łurię stanowiły kamień milowy w rozwoju neuropsychologii; były (są) inspiracją dla wielu jego bezpośrednich uczniów i kontynuatorów w Polsce (por. Kądziaława, 1983, 2002; Klimkowski, 1987; Maruszewski, 1969) i na świecie (por. Christensen i in., 2009; Goldberg, 1990; Mikadze i in., 2019).

Wraz z postępem neuronauk w sposób naturalny weryfikowane są koncepcje teoretyczne Łurii. Należy jednak mocno podkreślić, że nie tracą na aktualności takie postulaty Łurii, jak oparcie badania neuropsychologicznego na aktualnej wiedzy o relacji pomiędzy mózgiem a zachowaniem, poszukiwanie wspólnych mechanizmów zaburzeń, które mogą przejawiać się w różnych czynnościach, dostosowywanie sposobu badania do indywidualnego chorego, analiza dochodzenia pacjenta do rozwiązania, opis mocnych stron i zachowanych możliwości. Podkreśla się, iż zindywidualizowane badanie postulowane przez Łurię może być trafniejsze niż rezultaty rozbudowanych baterii testów i odnoszenie ogólnych wyników badania do norm, np. w badaniu chorych z głębokimi deficytami lub uszkodzeniami OUN mających charakter rozsiany lub rozległy (Guedalia i in., 2000; Mikadze i in., 2019).

Pomimo powyższych zalet, brak standaryzacji metod i procedur oraz normalizacji rezultatów okazał się na tyle dużym ograniczeniem, że metody badania Łurii w oryginalnej postaci nie przyjęły się w krajach Ameryki Północnej i Europy Zachodniej. Próby Łurii zostały zebrane przez Christensen i wydane w postaci wystandaryzowanej Baterii Testów Neuropsychologicznych Luria-Nebraska (*Luria-Nebraska Neuropsychological Battery* [LNNB]; por. Golden i in., 1980), co niewątpliwie zmieniło ich pierwotny charakter.

W Polsce, podobnie jak w wielu innych krajach (np. Finlandii), teorie oraz metody badania Łurii przyjęły się i obowiązywały przez wiele dziesięcioleci. Łucki (1967) pod kierunkiem Maruszewskiego opracował *Zestaw prób do badania zaburzeń wyższych czynności nerwowych u chorych z uszkodzeniami mózgu*, inspirowany koncepcją teoretyczną Łurii i jego podejściem kliniczno-eksperymentalnym. Chociaż brakuje danych o rozpowszechnieniu i sposobach wykorzystywania tego narzędzia w Polsce, to jednak powtórne wydanie *Zeszytów Łuckiego* (Łucki, 1995) wskazuje, że mogą być one nadal w jakimś stopniu stosowane w praktyce klinicznej w naszym kraju.

Podjęcie psychometryczne

W miarę niezależnie od podejścia kliniczno-eksperymentalnego, choć w podobnym czasie, na zachodzie Europy i w krajach Ameryki Północnej zaczęto rozwijać podejście psychometryczne w badaniu chorych z uszkodzeniami OUN. Podejście to cechowało wykorzystanie w praktyce klinicznej wystandaryzowanych narzędzi psychologicznych, służących do oceny zdolności intelektualnych, w oryginalnym zamierzeniu – na potrzeby szkół i edukacji szkolnej (por. Howieson, 2019). Techniki te powstawały w oparciu o dane wielkoskalowe i analizy statystyczne, dostarczające norm dla oceny ogólnego poziomu intelektualnego oraz w wybranych zakresach funkcji poznawczych (np. ilorazy inteligencji w skalach pełnej, słownej i wykonaniowej w *Skali inteligencji Wechslera*). W przeciwieństwie do metod łuriowskich, konstrukcja tych narzędzi nie odnosiła się do wiedzy o relacjach mózg–zachowanie czy zespołach bądź objawach zaburzeń zachowania po uszkodzeniach OUN.

Według Howieson (2019), inicjując stosowanie testów psychologicznych do badania inteligencji w warunkach klinicznych, neuropsychologowie chcieli wzmocnić trafność wnioskowania o ogólnym poziomie deficytu intelektualnego u chorych po uszkodzeniu OUN. Testy te zaczęto oceniać pod względem przydatności do indywidualnej diagnozy neuropsychologicznej oraz badań naukowych na różnych grupach chorych z uszkodzeniami mózgu (por. Strauss i in., 2006). Szczególne miejsce wśród narzędzi neuropsychologicznych zajmuje także *Bateria testów Halsteada-Reitana* (*The Halstead-Reitan Neuropsychological Test Battery* [HRB]), będąca zbiorem różnych metod i prób psychologicznych, które zaczęto kompletować i standaryzować z myślą o badaniu chorych neurologicznie w latach 40. ubiegłego wieku (Kądziaława, 1998; Lezak i in., 2012). Powstanie niektórych testów włączonych później do HRB datuje się na pierwsze dwudziestolecie XX wieku. Do nich należy między innymi *Test łączenia punktów* (*Trail Making Test* [TMT]) opracowany na potrzeby wojska amerykańskiego do oceny ogólnego stanu intelektualnego żołnierzy (por. Łojek i Stańczak, 2012).

Od czasów powstania *Skali inteligencji Wechslera* i HRB opracowano bardzo wiele wystandaryzowanych testów neuropsychologicznych. Lezak i in. (2012) przedstawili dane psychometryczne, wartość kliniczną, wady i zalety stosowania w badaniach i praktyce klinicznej z neuropsychologii ponad 700 takich metod. Umożliwiają one nie tylko wnioskowanie o deficycie w odniesieniu do odpowiednich norm, ale także tworzenie profili funkcjonalnych,

które poszerzają wiedzę o specyficznych zaburzeniach neuropsychologicznych w różnych populacjach pacjentów z dysfunkcjami OUN. Narzędzia te są rozwijane, modyfikowane, doskonalone dla celów badań diagnostycznych i naukowych w neuropsychologii.

PRAKTYKA KLINICZNA VS BADANIA NAUKOWE

Jak zasygnalizowano na wstępie, istnieją dwa odmienne obrazy stanu metod używanych do pomiaru neuropsychologicznego w zależności od celu i kontekstu, w jakim są stosowane: praktyki klinicznej vs badań naukowych. Na powstawanie rozbieżności pomiędzy nauką a praktyką w neuropsychologii zwracali już uwagę Lezak i in. w kolejnych wydaniach podręcznika *Neuropsychological assessment* (por. Lezak, 1995). Obecnie rozdźwięk ten znacznie się pogłębił i zjawisko to charakteryzuje aktualny stan dyscypliny (Marcopulos i Łojek, 2019).

Chociaż wiedza na temat tego, jakie metody pomiaru są faktycznie wykorzystywane w praktyce klinicznej jest ograniczona, to unikatowych informacji dostarczają dwukrotne badania ankietowe zrealizowane wśród neuropsychologów klinicznych w USA i Kanadzie, członków National Academy of Neuropsychology, International Neuropsychological Society oraz American Psychological Society Division 40 (Rabin i in., 2005, 2016). Celem ankiety było zebranie informacji na temat narzędzi neuropsychologicznych, które są stosowane przez tych specjalistów w ich praktyce klinicznej. W pierwszym pomiarze wzięło udział ponad 700 specjalistów, a w drugim ponad 500. Wszyscy mieli wykształcenie wyższe oraz minimum 15-letni staż pracy. Najczęściej prowadzili prywatną praktykę, pracowali w szpitalach, poradniach lub centrach rehabilitacyjnych. Pacjenci kierowani byli do nich na diagnozę neuropsychologiczną przez lekarzy (najczęściej neurologów i psychiatrów), psychologów, przedstawicieli szkół, stron w postępowaniu sądowym i innych.

Rabin i in. zamieścili w obu opracowaniach listę 25 najczęściej stosowanych testów. W tabeli 1 zestawiono pięć narzędzi neuropsychologicznych, które w obu ankietach były podane jako stosowane najczęściej przez minimum 15% badanych. W przypadku pozostałych 20 testów z listy, procent specjalistów, którzy je wykorzystywali w swojej praktyce codziennej wynosił poniżej 10.

Tabela 1

Zestawienie testów neuropsychologicznych najczęściej stosowanych w praktyce klinicznej w USA i Kanadzie (por. Rabin i in., 2005, 2016)

Test neuropsychologiczny	Ankieta I. %	Ankieta II. %	Źródło lub pierwsze wydanie
WAIS-IV lub wcześniejsze	63,1	64,9	1939 ^a
WMS-IV lub wcześniejsze	42,7	27,4	1945 ^a
TMT	17,6	26,4	1938 Ścieżki Paddingtona ^b
CVLT-II lub wcześniejsze	17,3	21,5	XIX w. próby Ebbinghaus ^c
WISC-IV lub wcześniejsze	15,9	20,5	1939 ^a

Adnotacja. Ankieta I. % – procent neuropsychologów stosujących dany test w swojej praktyce klinicznej według wyników ankiety pierwszej (Rabin i in., 2005); Ankieta II. % – procent neuropsychologów stosujących dany test w swojej praktyce klinicznej według wyników powtórnej ankiety (Rabin i in., 2016); WAIS – *Skala inteligencji Wechslera dla dorosłych*; WMS – *Skala pamięci Wechslera*; TMT – *Test łączenia punktów*; CVLT – *Kalifornijski test uczenia się językowego*; WISC – *Skala inteligencji Wechslera dla dzieci*.

^a Lezak i in., 2012; ^b Łojek i Stańczak, 2012; ^c Schmand, 2019.

Zgodnie z tabelą 1 metodami, które są obecnie najczęściej stosowane do badania diagnostycznego w neuropsychologii klinicznej w krajach Ameryki Północnej są wystandaryzowane testy psychologiczne, których, jak wcześniej wyjaśniono, pierwotnym zadaniem była ocena ogólnego poziomu intelektualnego lub pamięci na potrzeby psychologii szkolnej (np. *Skale inteligencji Wechslera*, *Skala pamięci Wechslera*) lub wojska (np. TMT). Do zestawu najczęściej stosowanych metod wchodzi również HRB, LNNB oraz *Bostoński test do diagnozy afazji* (*Boston Diagnostic Aphasia Examination* [BDAE]) oparty na XIX-wiecznym modelu afazji Wernickego-Lichtheima.

Część z omawianych metod jest już dostępna w wersji elektronicznej. Jednak zaledwie 6% specjalistów podało, że korzysta ze skomputeryzowanych wersji testów, większość stosowała te same metody w tradycyjnej formie (Rabin i in., 2016). Powyższe dane nabierają szczególnego znaczenia w obliczu faktu, że neuropsychologia kliniczna w USA i Kanadzie należy do najlepiej rozwiniętych na świecie. Dostępność metod pomiaru neuropsychologicznego w różnych wersjach i technologiach w praktyce klinicznej w tych krajach jest potencjalnie bardzo duża (por. Janzen i Guger, 2016; Sweet i in., 2002).

Neuropsychologowie kliniczni w Polsce mają znacznie mniejszy niż w USA i Kanadzie wybór specjalistycznych testów neuropsychologicznych. Można przypuszczać, że biorąc pod uwagę metody legalnie wydane, wystandaryzowane i znormalizowane, w praktyce neuropsychologicznej w naszym kraju

również najczęściej stosowane są *Skale inteligencji Wechslera* w wersjach dla dorosłych i dzieci (Brzeziński i in., 2004; Stańczak i in., 2020)¹.

W kontraście do wąskiego, archaicznego zestawu metod i procedur najczęściej stosowanych w praktyce klinicznej w neuropsychologii, stoi szeroki wachlarz technik pomiaru używanych i intensywnie rozwijanych w badaniach naukowych. Przegląd tych metod można znaleźć, zaznajamiając się z projektem Kryteria Domen Badawczych (RDoC; *Research Domain Criteria*) amerykańskiego Narodowego Instytutu Zdrowia Psychicznego (<https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc>). Kluczową część tego niezwykle kompendium współczesnej wiedzy o zdrowiu psychicznym i metodach jego badania stanowią dane z neurobiologii, neurofizjologii, neuropsychologii poznawczej i klinicznej. Jak wynika z samej nazwy, kryteria RDoC zostały przygotowane z myślą przede wszystkim o projektach naukowych, chociaż, co znamienne, docelowo mają służyć polepszeniu diagnozy, prewencji, terapii i leczenia zaburzeń funkcji psychicznych u ludzi. Jednak sposoby przejścia od badań naukowych do aplikacji w praktyce klinicznej postulowane w RDoC pozostają niejasne i są na etapie rozważań w środowisku neuropsychologów (por. Geraldo i in., 2019).

Zmniejszaniu rozbieżności w metodach stosowanych w praktyce klinicznej vs badaniach naukowych w neuropsychologii mogłyby służyć platformy internetowe dające możliwość realizacji pomiarów neuropsychologicznych w formie zdalnej, pozwalające na gromadzenie wielkich danych i ich wielodyscyplinarną integrację (Germine i in., 2019). Do znanych platform tego typu należy między innymi amerykańska platforma *Test My Brain* (<https://testmybrain.org/about.html>) polecana przede wszystkim do badań naukowych, ale również do ćwiczeń w zakresie wybranych funkcji poznawczych z indywidualnymi chorymi (por. Schmand, 2019). W badaniach naukowych często stosowany jest także portal z baterią CANTAB (*Cambridge Neuropsychological Test Automated Battery*; www.cambridgecognition.com)². Konstrukcję

¹ Z testów najczęściej używanych w USA i Kanadzie podanych w tabeli 1 polskie adaptacje i normalizacje mają *Kalifornijski test uczenia się językowego* (Łojek i Stańczak, 2010a) oraz *Kolorowy test połączeń* (wersja TMT; Łojek i Stańczak, 2012). W oparciu o *Skalę pamięci Wechslera* Choynowski (1959) opracował *Skalę pamięci*, która do dziś pozostała wyłącznie w formie próby eksperymentalnej. Trudno jednak ocenić, jak często te narzędzia są stosowane w praktyce klinicznej.

² Obie platformy są odpłatne, chociaż trzeba zwrócić uwagę, że korzystanie z formalnie wydanych, wystandaryzowanych testów psychologicznych również nie jest darmowe.

CANTAB rozpoczęto ponad 30 lat temu i od tego czasu jest ona nieustannie modernizowana i udostępniana w różnych językach. Zakres funkcji neuropsychologicznych mierzonych w CANTAB jest obecnie bardzo szeroki. Wiele prób ma swoje źródło w pracach z psychologii eksperymentalnej, neurokognitywistyki czy neuropsychofizjologii (m.in. pomiar czasu reakcji, zadanie „stop-idź”, próby oparte na zjawisku torowania, testy mierzące zachowania hazardowe, rozpoznawanie emocji). Platforma zawiera także cyfrowe wersje wielu kwestionariuszy i skal psychologicznych oraz metod do oceny sprawności w życiu codziennym. CANTAB jest rekomendowany do badań naukowych z udziałem chorych z uszkodzeniami mózgu, ale polecana jest także do indywidualnej pracy z pacjentami.

Ponadto platformy tego typu gromadzą duże dane z wykonania omawianych zadań przez osoby zdrowe. Wyniki te stanowią podstawę do opracowywania norm i ich użycie jest rekomendowane specjalistom dla celów diagnozy pacjentów z dysfunkcjami OUN. Platforma jest także intensywnie wykorzystywana do badań naukowych. Zgodnie z PubMed do grudnia 2023 roku było ponad 66 tysięcy publikacji naukowych opisujących badania z użyciem CANTAB. Jak wynika z pracy Rabin i in. (2016), przynajmniej do 2016 roku, żadna z omawianych platform internetowych nie należała jednak do 25 narzędzi neuropsychologicznych najczęściej stosowanych w praktyce klinicznej w USA i Kanadzie.

Powyższa sytuacja mogła się zmienić w związku z obostrzeniami w bezpośrednich kontaktach międzyludzkich wprowadzonych w czasie pandemii COVID-19. W tym czasie pojawiły się bardzo wyraźne rekomendacje dla wykonywania badań diagnostycznych pacjentów z uszkodzeniami mózgu w formie zdalnej, metodami skomputeryzowanymi, przy wzięciu pod uwagę zasad stosowania tych narzędzi, ich zalet i ograniczeń (Bilder i in., 2020; Cysique i in., 2022). Zakres stosowania tych metod w pandemii pozostaje jednak nieznany, nie ma także sygnałów o tym, że zmiany sposobu diagnozowania chorych w codziennej praktyce neuropsychologicznej, jakie wymusiła pandemia COVID-19, mogą mieć charakter bardziej długotrwały.

PRZYZYNY KONSERWATYZMU

Istnieje wiele przyczyn, dla których neuropsychologowie kliniczni okazują się skrajnie konserwatywni w wyborze metod diagnostycznych. Poniżej omówiono niektóre z nich.

Od odkrycia naukowego do praktyki klinicznej

Historia neuronauk wskazuje, że droga od epokowych odkryć naukowych dotyczących OUN do narzędzia, które byłoby stosowane w praktyce neuropsychologicznej jest zazwyczaj długa. Czas ten potrzebny jest do wielokrotnych replikacji badań źródłowych, weryfikacji założeń koncepcji teoretycznych wynikających z tych badań, drobiazgowego sprawdzania wartości metod pomiaru, a także dostrzeżenia aplikacyjnego znaczenia teorii i metod badania. Dla przykładu ponad sto lat minęło od przełomowego wydarzenia w nauce, jakim był opis objawów afazji ruchowej u chorego z uszkodzeniem lewego płata czołowego przez Brokę w 1861 roku, do pierwszego wydania – w 1972 roku – jednego z najczęściej używanych narzędzi do badania afazji – BDAE (ostatnie wydanie w 2003 roku; por. Lezak i in., 2012)³. Inny przykład to miary względnie stałych cech zachowania w oparciu o wiedzę o układzie nerwowym. Typologia układu nerwowego w oparciu o siłę, równowagę i ruchliwość procesów pobudzenia i hamowania komórek nerwowych kory mózgu Pawłowa powstała na podstawie badań realizowanych na przełomie XIX i XX wieku. Typologia Pawłowa stanowiła podstawę oryginalnych teorii osobowości Eysencka i temperamentu Strelaua, a w dalszej kolejności kwestionariuszy ich autorstwa, opracowywanych i doskonalonych przez wiele dekad, począwszy od przełomu lat 50. i 60. ubiegłego wieku. Ostateczna wersja *Kwestionariusza osobowości Eysencka* ukazała się w 1975 roku⁴. Jedną z ostatnich wersji *Kwestionariusza temperamentu* Strelaua opublikowano w latach 90. (por. Strelau 2008; Teopltitz,

³ Według badania Rabin i in. (2005, 2016) BDAE również należy do najczęściej używanych testów neuropsychologicznych w USA i Kanadzie.

⁴ Autorka opracowania dziękuje prof. Włodzimierzowi Strusowi za zwrócenie uwagi, że pomiar kluczowych w kontekście wspomnianej typologii wymiarów neurotyczności i ekstrawersji był możliwy już za pomocą wcześniejszych kwestionariuszy Eysencka, a więc *Eysenck Personality Inventory* opublikowanego w 1963 roku, a nawet *Maudsley Personality Inventory* z 1959 roku.

1982; Zawadzki i Strelau, 2018)⁵. Należy podkreślić, że powyższe kwestionariusze temperamentu i osobowości należą do wyjątkowych w naukach o mózgu i zachowaniu. Inne, powszechnie znane narzędzia psychologiczne do badania cech osobowości (np. NEO-PI-R czy MMPI) nie powstały już w oparciu o wiedzę neurobiologiczną czy neuropsychologiczną. Tworzono je, analizując dane behawioralne, a dopiero współcześnie łączy się profile behawioralne ze złożonymi danymi o OUN na poziomach strukturalnym, sieciowym i czynnościowym, co owocuje dużą ilością nie zawsze spójnych wyników (por. Butcher i in., 2012; Siuta, 2006). Stąd niektórzy autorzy zwracają uwagę na to, że ważne jest nie tylko wykorzystywanie coraz nowocześniejszych technik pomiaru dla celów diagnostycznych, ale przede wszystkim integracja wiedzy opartej na faktach, pochodzącej z badań na różnych poziomach neurobiologicznej organizacji zachowania, pozyskiwana różnymi metodami, w tym tradycyjnymi neuropsychologicznymi (Parsons i Duffield, 2019; Prince, 2018).

Standardy psychometryczne

Kolejną przyczyną konserwatyzmu neuropsychologów mogą być także wysokie wymagania psychometryczne, które muszą spełniać metody pomiaru stosowane dla celów diagnostycznych w praktyce klinicznej. Zalecane są tu metody, których wskaźniki rzetelności wynoszą minimum 0,70 (por. Lezak i in., 2012). Istnieje przekonanie, że techniczne unowocześnienie, komputeryzacja uznanych metod neuropsychologicznych do badania funkcji poznawczych (np. czasów reakcji, uwagi złożonej czy funkcji wykonawczych) gwarantuje wyższą niż dotychczas dokładność i rzetelność pomiaru (Kessels, 2019). Okazuje się, że pogląd ten nie jest w pełni trafny. Feenstra i in. (2017) w metaanalizie danych dotyczących skomputeryzowanych metod neuropsychologicznych do pomiaru funkcji poznawczych wykazali wskaźniki rzetelności wahające się od 0,17 do 0,87, przy medianie wynoszącej 0,70. Wbrew oczekiwaniom wskaźniki te nie były istotnie wyższe niż te uzyskiwane w przypadku metod w wersji

⁵ Historia powstawania *Kwestionariusza temperamentu* Strelaua obejmuje publikację różnych wersji tego narzędzia. Wstępna wersja została zamieszczona w artykule opublikowanym w 1967 roku w *Przeglądzie Psychologicznym* pt. *Próba zastosowania kwestionariusza do badania temperamentu (w ujęciu pawłowskim)*, a finalnie w 1969 roku w monografii pt. *Temperament i typ układu nerwowego*. Autorka opracowania dziękuje prof. Bogdanowi Zawadzkiemu za powyższą cenną uwagę.

tradycyjnej, a ponadto rzetelność wielu testów czy prób w wersji skomputeryzowanej nie spełniała wymaganych standardów. Feenstra i in. wykazali ponadto, że trafność zbieżna pomiędzy analogicznymi testami w wersjach tradycyjnej i cyfrowej jest umiarkowana (mediana $r = 0,49$). Zmiana formy i modyfikacje proceduralne powodują, że po unowocześnieniu znanej metody klinicysta ma jednak do czynienia z nieco innym narzędziem, które wymaga zbadania i opisu psychometrycznego. Zatem przejście od sprawdzonego, wartościowego psychometrycznie i klinicznie testu w wersji tradycyjnej do skomputeryzowanej formy tego samego testu powinno budzić uzasadnione wątpliwości neuropsychologów praktyków (Schmand, 2019).

Dodatkowe utrudnienia wiążą się z szybkością postępu technicznego i idącą za tym dostępnością coraz nowszych, różnych wersji sprzętu i oprogramowania. Testy diagnostyczne w wersji cyfrowej mogą być dostępne na różnych urządzeniach (starszych vs nowszych komputerach, z różnym oprogramowaniem, ekranami, kartami graficznymi, na laptopach, iPadach, smartfonach), oraz stosowane w różnych warunkach (stacjonarnie lub zdalnie, w obecności osoby badającej lub bez). Przy braku wiedzy o trafności i rzetelności, punktach odcięcia czy normach danego narzędzia jego użyteczność dla celów diagnostycznych może być wątpliwa, a diagnoza obarczona błędem pomiaru. Problem dotyczy także sytuacji, w której specjalista monitoruje stan pacjenta i chory rozwiązuje te same testy poznawcze kilka razy, ale na różnym sprzęcie, w różnych warunkach.

Poważne ograniczenia pomiarów prowadzonych zdalnie, szczególnie bez nadzoru osoby badającej, są dyskutowane od dłuższego czasu (por. Bilder i in., 2020). Klinicysta otrzymuje wyniki, ale nie kontroluje warunków badania, nie monitoruje stanu pacjenta (wysiłku, zaangażowania, zmęczenia, nastroju), nie ma pewności, czy chory właściwie zrozumiał instrukcję, nie obserwuje sposobu rozwiązywania zadań, nie ma możliwości ewentualnego skorygowania procedury badania, tak aby dotrzeć do zachowanych możliwości pacjenta.

Kontrowersje budzi także jakość danych normalizacyjnych zbieranych zdalnie (Bilder i in., 2020). Do zastrzeżeń wymienionych powyżej dochodzą argumenty dotyczące ograniczeń reprezentatywności próby normalizacyjnej oraz trafności pomiaru w przypadku niektórych grup społecznych. Do nich należą osoby mające ograniczony dostęp do nowoczesnych technologii (np. o niskim statusie społeczno-ekonomicznym), niekorzystające z tego typu urządzeń (np. w podeszłym wieku) czy badane w języku, który nie jest ich językiem macierzystym (por. Fernandez, 2019). Ponadto dane normalizacyjne

do metod neuropsychologicznych dostępnych zdalnie gromadzone są niekiedy przez wiele lat – w czasie, gdy narzędzia i procedury ulegają modyfikacjom (np. CANTAB; por. Schmand, 2019). Normy te budzą zatem uzasadnione wątpliwości klinicystów: interpretacja wyniku chorego w odniesieniu do norm o niskiej jakości może być nietrafna, niesie ryzyko błędnej diagnozy i niekorzystnych konsekwencji dla chorego. Normy zbierane w sposób tradycyjny, opisane w podręcznikach do testów, pomimo swoich ograniczeń (np. utraty aktualności i konieczności systematycznych renormalizacji) wciąż stanowią bezpieczniejszą alternatywę dla neuropsychologów praktyków.

ŚCIEŻKI ROZWOJU

Sugerowanych jest kilka, do pewnego stopnia komplementarnych ścieżek rozwoju, które mogą prowadzić do postępu w zakresie metod pomiaru stosowanych w praktyce neuropsychologicznej.

Integracja wiedzy naukowej

Zjawiskiem charakteryzującym obecne neuronauki jest dążenie do integracji wiedzy o mózgu i zachowaniu gromadzonej z zastosowaniem metod pomiaru z różnych dyscyplin. Dane z badań na poziomie behawioralnym (pozyskiwane m.in. z użyciem tradycyjnych metod neuropsychologicznych) są zestawiane z wynikami pomiarów neurobiologicznych oraz mapowania czynności poznawczych w odniesieniu do określonych regionów czy sieci mózgu z pomocą fMRI, PET czy ERP EEG⁶ u osób zdrowych i w różnych populacjach klinicznych (Geraldo i in., 2019).

Powyższe zjawisko ma dalekosiężne znaczenie dla neuropsychologii jako nauki, ale również specjalistycznych metod pomiaru neuropsychologicznego. Jednym z efektów badań integrujących neuronauki jest poszerzanie się wiedzy o trafności, rzetelności, wartości klinicznej i ekologicznej tradycyjnych

6 fMRI (*functional magnetic resonance imaging*) – funkcjonalny rezonans magnetyczny, PET (*positron emission tomography*) – pomiar pozytonowej emisji elektronów, ERP EEG (*event-related potential electroencephalography*) – metoda bioelektrycznej reakcji mózgu na zdarzenie z zastosowaniem elektroencefalografu.

narzędzi pomiaru neuropsychologicznego (Kessels, 2019). Wykazywana jest ich specyfika, komplementarność, unikatowość danych w stosunku do metod neuroobrazowych czy neurofizjologicznych. Dla przykładu tradycyjne techniki pomiaru neuropsychologicznego lepiej określają ogólny stan umysłowy i poziom deficytu w zakresie poszczególnych funkcji poznawczych oraz trafniej przewidują zdolności chorego do radzenia sobie w życiu codziennym niż wyniki badania w skanerze. Z kolei dane z funkcjonalnych pomiarów neuroobrazowych i neurofizjologicznych stanowią ważne dopełnienie metod behawioralnych (niezależnie, czy w formie „papier-ołówek”, czy skomputeryzowanej) w odniesieniu do opisu neurobiologicznej organizacji i reorganizacji czynności mózgu pacjenta. Wykazywanie braku ekwiwalentności porównywalnych metod może stanowić stymulację do tworzenia nowych technik pomiaru funkcji behawioralnych (Geraldo i in., 2019).

Postępująca integracja neuronauk prowadzi także do weryfikacji makro- i mikromodeli relacji mózg–zachowanie, co w konsekwencji może mieć znaczenie rewolucjonizujące dla metod pomiaru neuropsychologicznego. Obserwacja równoległej aktywności rozproszonych sieci neuronalnych całego OUN podczas wykonywania zadań poznawczych doprowadziła do podważenia dotychczasowych założeń o związkach danych funkcji czy zdolności poznawczych ze skupiskami komórek nerwowych mózgu o znanej lokalizacji i połączeniach. Na przykład czynność nazywania przedmiotów może łączyć się nie tylko z lewym płatem skroniowym (uszkodzenie tego obszaru zazwyczaj powoduje zaburzeniami nazywania – anomię) czy siecią brzuszną kory mózgu związaną z odpowiedzią na pytanie „Co widzę?”. W zależności od cech obiektu, indywidualnych doświadczeń życiowych, indywidualnej organizacji funkcjonalnej mózgu czy kontekstu (sytuacji, zadania) nazwa obiektu może uaktualnić się u osoby zdrowej przy zaangażowaniu różnych obszarów mózgu (por. Łojek, 2009). Prowadzi to do rozważań nie tylko nad złożonością neurobiologicznej reprezentacji funkcji psychicznych, ale także ponownego w historii neuronauk zadawania pytania, czym są tzw. wyższe funkcje nerwowe. Podkreślana jest konieczność propozycji nowych ontologii czy konstruktów teoretycznych dla opisu tych funkcji (Prince, 2018). Zakłada się, że dogłębne zrozumienie procesów przetwarzania informacji przez mózg będzie możliwe dopiero wtedy, gdy rozpoznane zostaną wszystkie neuronalne ścieżki rozwiązywania danego zadania u osób zdrowych i z uszkodzonym mózgiem. Ma to w konsekwencji doprowadzić do postępu w badaniu i rehabilitacji chorych z dysfunkcjami mózgu (Parsons i Duffield, 2019; Prince, 2017). Powyższy sposób myślenia

o kierunku, w którym może iść integracja neuronauk, może potencjalnie zawocować gruntownymi zmianami w metodach badania neuropsychologicznego w praktyce klinicznej. Efekty tej debaty nie są jednak jeszcze widoczne w postaci zmiany przyjętych teorii mózg–zachowanie bądź paradygmatów badawczych.

Należy podkreślić, że postulaty integracji wiedzy opartej na faktach naukowych nie dotyczą wyłącznie jednoczenia neuronauk. Wizja integracji obejmuje wszelkie dziedziny zajmujące się zdrowiem człowieka. Zgodnie z tą wizją nowoczesne technologie i uczące się algorytmy powinny pozyskiwać, gromadzić, analizować wielkie dane, obejmujące wyniki wszechstronnych badań specjalistycznych, archiwalnych i aktualnych dotyczących zdrowia pacjenta, w tym: genetycznych, medycznych, laboratoryjnych, a także neuroobrazowych, neuropsychologicznych, społecznych, środowiskowych i innych. Celem tej integracji i użycia nowoczesnych technologii jest walka z chorobami i polepszenie zdrowia ludzkości (Parsons i Duffield, 2019).

Nowoczesna psychometria

Duże nadzieje na rozwój metod pomiaru w neuropsychologii klinicznej wiąże się obecnie z odejściem od klasycznej teorii testów i wdrażaniem nowoczesnych modeli psychometrycznych, a w szczególności IRT (*item response theory*) – teorii odpowiedzi na pozycje testowe. U podstaw IRT leży założenie, że ukryta cecha (*latent trait*) decyduje o udzieleniu określonej odpowiedzi przez osobę badaną w danej pozycji testu psychologicznego (De Champlain, 2010; Kondratek i Pokropek, 2013). Relacje pomiędzy cechą ukrytą (latentną) a odpowiedziami w zadaniu lub na pytanie kwestionariusza (wskaźnikami) są modelowane przy zastosowaniu serii parametryzacji, która ma na celu określenie właściwości wskaźników oraz rozkład cech u respondentów. Rzadziej stosowane jest także nieparametryczne modelowanie IRT. W większości analiz IRT zakłada się, że o odpowiedzi w danej pozycji testu decyduje jedna cecha latentna (modelowanie jednowymiarowe), rozwijane są jednak również paradygmaty dwu- i wielowymiarowe.

Modele IRT wydają się szczególnie pasować do celów i potrzeb pomiarów neuropsychologicznych. Modele dwuwymiarowe IRT z założenia są zbieżne z celami badania diagnostycznego w praktyce klinicznej, które ma prowadzić zarówno do ustalenia ogólnego poziomu zaburzeń, jak i stanu określonych funkcji psychicznych (w tym poznawczych). Wielowymiarowe modelowanie IRT może wspomóc w określaniu ukrytych cech wspólnych dla wielu testów

neuropsychologicznych, które mają mierzyć zdolności z różnych domen poznawczych (np. pamięci roboczej, zdolności językowych i funkcji wykonawczych). Mogą także ustalić cechy latentne wspólne dla testów jednego typu, np. różnorodnych skal przesiewowych do oceny ogólnego stanu intelektualnego potrzebnych do badania chorych z demencją i łagodnymi zaburzeniami poznawczymi. Łączenie metod neuropsychologicznych w ramach analiz IRT oraz wyłanianie wskaźników najbardziej wartościowych klinicznie, pozycji nadmiarowych lub nietrafnych może prowadzić do tworzenia nowych narzędzi o wysokiej wartości psychometrycznej i użyteczności klinicznej.

IRT wzbudza zainteresowanie neuropsychologów również z uwagi na możliwość pomiaru adaptacyjnego. Na podstawie analizy rzetelności wyników z dotychczasowych odpowiedzi osoby badanej skomputeryzowane algorytmy IRT mogą wskazywać, która pozycja testowa czy zadanie powinno być rozwiązywane przez osobę badaną w następnej kolejności. Algorytm pomaga również w podjęciu decyzji o zakończeniu badania w momencie, gdy zostanie osiągnięty zakładany przez klinicystę określony poziom rzetelności wyników. IRT pozwala zatem na dopasowanie zadań do możliwości chorego, skraca test i pomiar przy zachowaniu oczekiwanych standardów psychometrycznych. Potencjalnie krótszy czas badania diagnostycznego jest niezwykle ważny, gdyż pomiary z zastosowaniem wszechstronnych baterii testów neuropsychologicznych mogą zajmować wiele godzin, kilka sesji, co jest męczące dla chorego i czasochłonne dla specjalisty.

Kolejną cechą pomiaru z zastosowaniem IRT, niezwykle ważną w badaniu neuropsychologicznym, jest otwarcie na możliwość analizy jakościowej odpowiedzi, a szczególnie odpowiedzi błędnych. Odpowiedzi nieprawidłowe (perseweraacje, prawie błędy, ominięcia, pomyłki w ramach kategorii, tendencja do popełniania określonego typu błędów, wyolbrzymianie trudności) są dla klinicystów jednym z kluczowych źródeł informacji o stanie pacjenta, jego deficytach, ale też zachowanych możliwościach, strategiach radzenia sobie z zadaniem, wkładanym wysiłkiem.

Analiza jakościowa danych umożliwia także zrozumienie, że poszczególne pozycje testowe mogą w różny sposób „działać”, spełniać swoją rolę w badaniu określonej osoby w zależności od czynników demograficznych, społeczno-kulturowych i środowiskowych (tj. wieku, płci, rasy, pochodzenia etnicznego, statusu społeczno-ekonomicznego, czynników językowych, narodowości itd.). Dogłębne poznanie oraz kontrola wpływu powyższych zmiennych na wykonywanie testów psychologicznych i neuropsychologicznych stanowi od dawna wielkie wyzwanie dla specjalistów (por. Jodzio, 2011).

Przykładowe badanie z zastosowaniem IRT

W celu zobrazowania znaczenia wdrażania nowoczesnej psychometrii poniżej zamieszczono opis pilotażowego badania z zastosowaniem IRT, które przeprowadziły Łojek i Stańczak (2010b)¹ na danych z polskiej normalizacji ($N = 478$) *Kalifornijskiego testu uczenia się językowego* (CVLT; *California Verbal Learning Test*). Klasyczna teoria testu zakłada stopniowe istotne pogarszanie się procesów uczenia się i pamięci wraz z wiekiem, co rzeczywiście potwierdzają dane normalizacyjne CVLT (Łojek i Stańczak, 2010a). W badaniu z zastosowaniem IRT zastanawiano się, czy powyższa prawidłowość ma związek z poszczególnymi pozycjami CVLT, a dokładniej szesnastoma słowami z listy A, których uczy się na pamięć osoba badana, powtarzając pięciokrotnie te słowa bezpośrednio po ekspozycji listy słów, a także po krótkim i długim odroczeniu, swobodnie, z pomocą oraz rozpoznając słowa właściwe wśród dystrakcyjnych.

Okazało się, że w każdej grupie wiekowej pewne pozycje testowe (określone słowa) wykazywały istotny negatywny związek z cechą ukrytą (zdolnością do zapamiętywania słów), ale związki te były odmienne w poszczególnych grupach wiekowych. U osób najmłodszych (16–34 lata) odnotowano najwięcej pozycji (dziewięć słów), w których poziom cechy latentnej był ujemnie związany z prawdopodobieństwem udzielenia poprawnej odpowiedzi. Ujawniło się to jednak przede wszystkim podczas pierwszego odtwarzania listy słów (zadanie 1) i nie utrzymywało się podczas kolejnych odtworzeń. Natomiast w grupie najstarszej (65–79 lat) pozycji takich było mniej (cztery słowa), ale ich specyficzne „zachowanie” utrzymywało się w większej liczbie zadań (pierwszym i piątym odtwarzaniu bezpośrednim, odtwarzaniu swobodnym i z pomocą po długim odroczeniu oraz rozpoznawaniu). Dalsze analizy wykazały, że prawdopodobieństwo poprawnego odtworzenia pozycji w zadaniu 1 przez osoby młode nie było związane z poziomem ich cechy latentnej, lecz z innymi czynnikami wpływającymi na kodowanie w pamięci krótkotrwałej słów bodźcowych. Potrzebne były dodatkowe dane, aby próbować wnioskować o tym, co to były za czynniki (np. słaba koncentracja, niska

¹ Analizy IRT zostały przeprowadzone z pomocą prof. Michała Kosińskiego, wówczas zatrudnionego w Cambridge Psychometrics Centre i współpracującego z Pracownią Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.

motywacją). Czynniki te przestawały jednak mieć znaczenie wówczas, gdy bodziec zaczął być kodowany w magazynie pamięci długotrwałej, czyli już od drugiego powtórzenia słów z listy A, po bezpośredniej ekspozycji. Ponadto okazało się, że wykształcenie miało istotne znaczenie dla wyników uzyskanych w grupie najstarszej. U seniorów przeważały osoby z wykształceniem zawodowym (co dobrze odzwierciedlało strukturę wykształcenia w populacji dla tej grupy wiekowej). Analizy IRT wykazały, że niektóre pozycje (np. słowo „śrubokręt”) były prawdopodobnie z powodu osobistych doświadczeń zawodowych lepiej zapamiętywane przez osoby z wykształceniem zawodowym niż przez osoby z wykształceniem wyższym niż zasadnicze zawodowe. Można przypuszczać, że zapamiętywanie słów u osób w podeszłym wieku było w większym stopniu pochodną ich życiowych doświadczeń (związanych np. z wykonywanym zawodem) niż to miało miejsce w przypadku badania osób młodych. Powyższe badanie pilotażowe z zastosowaniem IRT wskazuje, że te same pozycje testowe mogą angażować i mierzyć nieco odmienne funkcje neuropsychologiczne ze względu na cechy demograficzne i doświadczenie życiowe osób badanych. Tego typu informacje mogłyby być dzięki modelowaniu IRT uwzględnione w procesie diagnozy neuropsychologicznej, wzmacniając trafność rozpoznania oraz zaleceń do interwencji.

IRT nie jest jedynym narzędziem psychometrycznym, którego wykorzystanie jest postulowane dla unowocześnienia metod pomiaru w neuropsychologii. Przykładem innego mogą być modele oparte na matematycznej teorii przestrzeni wiedzy (KST; *knowledge space theory*), wykorzystujące inteligentne maszyny uczące się (por. de Chiusole i in., 2024). Modele te ujmują wiedzę człowieka jako strukturę kombinatoryczną, zawierającą skończony zbiór pojęć i zdolności. W odróżnieniu od klasycznych testów psychologicznych, nie mają one na celu oceny wyników osoby badanej według liniowo uporządkowanych danych liczbowych (norm). Służą raczej do opisu, w jakim obszarze dana jednostka wykazuje największą wiedzę i najwyższe możliwości poznawcze, gdzie ma trudności, oraz czego jest zdolna się nauczyć. Zdania są w trakcie pomiaru dopasowywane do potrzeb i wymagań osoby badanej (badanie adaptacyjne). Ocena wiedzy i zdolności uczenia się techniką KST może stanowić swoiste połączenie badania diagnostycznego oraz ćwiczeń w ramach programu rehabilitacyjnego dla osób z uszkodzonym mózgiem (por. de Chiusole i in., 2024).

Należy podkreślić, że założenia nowoczesnych modeli psychometrycznych (IRT czy KST) spełniają wiele postulatów formułowanych przez Łurę (1967) w podejściu jakościowym do badania chorych z uszkodzonym mózgiem. Analiza sposobu dochodzenia do rozwiązania zadania, niepoprzestawanie na ostatecznym wyniku, opis mocnych i słabych stron funkcjonowania chorego to wyznaczniki podejścia jakościowego. Chociaż badania cech latentnych nie należy utożsamiać z poszukiwaniem deficytu podstawowego w rozumieniu Łurii, to jednak analizy IRT mogą wspomagać odkrywanie złożonych uwarunkowań zachowania czy sposobu rozwiązywania zadań przez chorego. Dla zrozumienia mózgowych mechanizmów zaburzeń funkcji psychicznych konieczna jest wspomniana wyżej integracja tak pogłębionej wiedzy psychologicznej z informacjami o stanie i pracy mózgu oraz danych biomedycznych, co może nastąpić przy dalszym wykorzystaniu inteligentnych maszyn uczących się.

Nowoczesna psychometria nie jest wolna od wielu ograniczeń (por. Germeine i in., 2019; Schmand, 2019). Algorytmy IRT opierają się na wielkich danych, które w neuropsychologii klinicznej zaczynają być dopiero gromadzone. Ich stosowanie wymaga zaawansowanej wiedzy psychometrycznej, matematycznej, odpowiedniego sprzętu oraz oprogramowania, które najczęściej nie są dostępne dla neuropsychologów praktyków. Stosowanie algorytmów matematycznych i maszyn uczących się musi być dogłębnie rozumiane i kontrolowane przez człowieka na każdym etapie – od wprowadzania odpowiednich danych poprzez analizę według ściśle określonych kryteriów do wyników końcowych. Efekty pracy algorytmów matematycznych muszą być ostatecznie interpretowane przez specjalistę, gdyż w praktyce klinicznej to człowiek czy zespół ludzi decyduje o życiu i zdrowiu jednostki.

Ponadto konstrukcja testów neuropsychologicznych nie zawsze odpowiada warunkom określonego modelowania matematycznego. Dla przykładu w próbach polegających na jak najszybszym powtarzaniu czynności psychomotorycznej (np. przyciskaniu dźwigni palcem, wkładaniu kołeczek do otworów czy wykreślanu liter) algorytm może pomóc w dopasowaniu trudności zadania, wychwyceniu zmęczenia u osoby badanej, tendencji do wyolbrzymiania swoich trudności. Jednak z założenia nie będzie miał możliwości zróżnicowania pozycji testowych pod względem cechy latentnej.

Podsumowując, nowoczesne modele psychometryczne mają potencjalnie bardzo duże zastosowanie do analizy danych neuropsychologicznych gromadzonych zarówno w badaniach naukowych, jak i praktyce klinicznej. Ich zaletą jest zindywidualizowane, jakościowe podejście do badania i pracy z chorymi

z uszkodzonym mózgiem, przy jednoczesnym możliwym utrzymaniu wysokich standardów trafności i rzetelności pomiaru. Ponadto mogą wspomagać proces modernizacji dotychczasowych oraz tworzenia nowych narzędzi pomiaru neuropsychologicznego. Ze względu na wymagania dotyczące choćby dużej ilości danych oraz konieczności korzystania z zaawansowanych statystyk, proces wdrażania nowoczesnej psychometrii nie jest jednak łatwy.

Monitorowanie sprawności w życiu codziennym

Stan osoby z dysfunkcjami mózgu może ulegać zmianom krótko- (minuty, godziny, dni) i długoterminowym (miesiące, lata). Wahanie te służą jako wskaźniki tendencji do poprawy zdrowia pacjenta, ale również zwiastują postęp choroby (Jodzio, 2011). Monitorowanie kondycji chorych, wczesne wykrywanie sygnałów o pogarszaniu się ich funkcjonowania, ale także systematyczna kontrola efektywności leczenia i rehabilitacji to poważne wyzwania dla systemów ochrony zdrowia. Problem ten dotyczy również badań neuropsychologicznych w praktyce klinicznej. Chory zazwyczaj odbywa regularne wizyty kontrolne (np. co pół roku, co rok). Jednak powtarzanie pomiarów z zastosowaniem tych samych testów czy prób neuropsychologicznych niesie ze sobą wiele ograniczeń z uwagi na wpływ uczenia się, znużenie pacjenta oraz wartość ekologiczną testów, których wyniki w pełni oddają sprawność pacjenta w życiu codziennym (Howieson, 2019). Monitorowanie stanu chorego z użyciem tradycyjnych neuropsychologicznych metod i procedur badania jest zatem z wielu powodów mocno ograniczone.

Nadzieje na poprawę wiąże się z obserwacją efektów rewolucji, jaka dokonuje się w medycynie, dzięki wdrażaniu systemów zdalnej diagnostyki i monitoringu pacjentów. Nowoczesne narzędzia bioinformatyczne zbierają i przekazują zdalnie dane o pracy organów wewnętrznych (np. serca i układu krążenia), wagi ciała, masy mięśniowej, jakości słuchu, ilości snu, poziomie glukozy, aktywności ruchowej chorego itd. (por. Kordiyak i Shakhovska, 2016). Tak zbierane, różnorodne dane służą specjalistom w opiece nad pacjentami, ale w przypadku osób zdrowych (np. w podeszłym wieku) mogą wspomagać zdrowy tryb życia i prewencję chorób. Funkcje neuropsychologiczne konieczne do radzenia sobie z wyzwaniami codzienności (m.in. komunikacja słowna, wykonywanie czynności celowych, uczenie się i pamięć, szybkość psychomotoryczna, orientacja wzrokowo-przestrzenna) również mogłyby być monitorowane w podobny

sposób. Służyłyby do tego połączone z internetem smartfony, komputery oraz sprzęty codziennego użytku, tzw. „internet rzeczy” (IoT; *internet of things*).

Wdrażanie powyższych wizji w życie nie jest jednak proste. Brakuje koncepcji, które definiowałyby zakres i sposób użycia tych narzędzi do badania, monitorowania i wspomagania funkcji neuropsychologicznych u pacjentów w życiu codziennym. Do wątpliwości, które opisano powyżej odnośnie do modernizacji metod diagnostycznych w neuropsychologii, dochodzą również zastrzeżenia natury etycznej dotyczące granicy prywatności, której nie należy przekraczać, wchodząc z wszechobecną technologią w życie chorego.

PODSUMOWANIE I WNIOSKI

Celem niniejszego rozdziału było przedstawienie dyskusji, która się toczy w środowisku neuropsychologów na temat metod diagnostycznych i procedur badania stosowanych w praktyce klinicznej. Głównym problemem wydaje się zbyt powolna i niewystarczająca modernizacja stosowanych od wielu dziesięcioleci narzędzi pomiaru diagnostycznego stosowanych w praktyce klinicznej oraz silny opór w odniesieniu do wdrażania nowych, oryginalnych metod, które charakteryzowałyby się nowoczesnymi postawami teoretycznymi, konstrukcją, procedurą pomiaru, sposobami analizy wyników. Zjawisko to jest charakterystyczne dla metod i procedur stosowanych w praktyce klinicznej, w odróżnieniu od intensywnego rozwoju i różnorodności narzędzi używanych w badaniach naukowych z obszaru neuropsychologii czy neurokognitywistyki. Starano się wykazać, że przyczyny tego oporu mogą być uwarunkowane wieloma czynnikami. Należą do nich przede wszystkim troska o zachowanie wypracowanych w toku rozwoju dziedziny wysokich standardów psychometrycznych przyjętych dla metod oraz procedur badania diagnostycznego. Standardy te kształtowały się przez lata w ramach komplementarnych podejść: jakościowego i ilościowego. Modernizacja i tworzenie nowych metod pomiaru, które będą miały zastosowanie w praktyce klinicznej mogą dokonywać się między innymi w efekcie integracji neuronauk i nauk o zdrowiu, wdrażania nowoczesnych algorytmów psychometrycznych i matematycznych, w tym sztucznej inteligencji oraz poprzez dane pozyskiwane przez urządzenia techniczne, z których korzysta pacjent w życiu codziennym.

Niezależnie od tego, w którą stronę rozwiną się metody badania neuropsychologicznego, należy przypuszczać, że przełom będzie możliwy wyłącznie

przy zaangażowaniu zespołu różnych specjalistów: neuropsychologów, matematyków, inżynierów i ekspertów w zakresie psychometrii. Nie bez znaczenia jest również zrozumienie ze strony wydawców testów psychologicznych oraz aktywne wspieranie przez nich procesu zmian i modernizacji metod pomiaru neuropsychologicznego.

Podsumowując, należy stwierdzić, że problem z koniecznością unowocześniania metod i procedur badania w praktyce neuropsychologicznej nie polega na pytaniu „czy”, ale „jak” to zrobić, aby przy wprowadzaniu zmian zachować najwyższe standardy psychometryczne i wymogi kliniczne. Celem musi być zawsze dobro chorych i nie można „eksperymentować” na nich, stosując nowoczesne technologie tylko dlatego, że są dostępne.

BIBLIOGRAFIA

- Ardila, A. (2013). A new neuropsychology for the XXI century. *Archives of Clinical Neuropsychology*, 28, 751–762. <https://doi.org/10.1093/arclin/act036>
- Bilder, R. (2011). Neuropsychology 3.0: Evidence-based science and practice. *Journal of the International Neuropsychological Society*, 17(1), 7–13. <https://doi.org/10.1017/S1355617710001396>
- Bilder, R. M., Postal, K. S., Barisa, M., Aase, D. M., Cullum, C. M., Gillaspay, S. R., Harder, L., Kanter, G., Lanca, M., Lechuga, D. M., Morgan, J. M., Most, R., Puente, A. E., Salinas, C. M., Woodhouse, J. (2020). Inter Organizational practice committee recommendations/guidance for teleneuropsychology in response to the COVID-19 pandemic. *Archives of Clinical Neuropsychology*, 35(6), 647–659. <https://doi.org/10.1093/arclin/acia046>
- Brzeziński, J., Gaul, M., Hornowska, E., Jaworowska, A., Machowski, A., Zakrzewska, M. (2004). *Skale inteligencji D. Wechslera dla dorosłych. Wersja zrewidowana – renormalizacja WAIS-R(PL). Podręcznik*. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G. (2012). *MMPI-2. Minnesocki wielowymiarowy inwentarz osobowości – 2. Podręcznik stosowania, oceny i interpretacji. Wersja zrewidowana (polska normalizacja: U. Brzezińska, M. Koć-Januchta i J. Stańczak)*. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Casaletto, K. B., Heaton, R. K. (2017). Neuropsychological assessment: Past and future. *Journal of the International Neuropsychological Society*, 23(9–10), 778–790. <https://doi.org/10.1017/S1355617717001060>

- Choynowski, M. (1959). *Skala pamięci. Podręcznik tymczasowy* [nieopublikowany maszynopis]. Pracownia Psychometryczna, Polska Akademia Nauk.
- Christensen, A. L., Goldberg, E., Bougakov, D. (red.). (2009). *Luria's legacy in the 21st century*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195176704.001.0001>
- Cysique, L. A., Łojek, E., Cheung, T. C.-K., Cullen, B., Egbert, A. R., Evans, J., Gawron, N., Gouse, H., Hansen, K., Holas, P., Hyniewska, S., Malinowska, E., Marcopulos, B. A., Merkle, T. L., Muñoz-Moreno, J. A., Ramsden, C., Salas, C., Sikkes, S. A. M., Silva, A. R., [...] NeuroCOVID International Neuropsychology Taskforce. (2022). Assessment of neurocognitive functions, olfaction, taste, mental, and psycho-social health in COVID-19 in adults: Recommendations for harmonization of research and implications for clinical practice. *The Journal of the International Neuropsychological Society*, 28(6), 642–660. <https://doi.org/10.1017/S1355617721000862>
- De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education*, 44(1), 109–117. <https://doi.org/10.1111/j.1365-2923.2009.03425.x>
- de Chiusole, D., Spinoso, M., Anselmi, P., Bacherini, A., Balboni, G., Mazzoni, N., Brancaccio, A., Epifania, O. M., Orsoni, M., Giovagnoli, S., Garofalo, S., Benassi, M., Robusto, E., Stefanutti, L., Pierluigi, I. (2024). PsycAssist: A web-based artificial intelligence system designed for adaptive neuropsychological assessment and training. *Brain Sciences*, 14(2), artykuł 122. <https://doi.org/10.3390/brainsci14020122>
- Feenstra, H. E., Vermeulen, I. E., Murre, J. M., Schagen, S. B. (2017). Online cognition: Factors facilitating reliable online neuropsychological test results. *The Clinical Neuropsychologist*, 31(1), 59–84. <https://doi.org/10.1080/13854046.2016.1190405>
- Fernandez, A. L. (2019). Modern neuropsychological tests for a diversity of cultural contexts. *The Clinical Neuropsychologist*, 33(2), 438–445. <https://doi.org/10.1080/13854046.2018.1560501>
- Geraldo, A., Azeredo, A., Pasion, R., Dores, A. R., Barbosa, F. (2019). Fostering advances to neuropsychological assessment based on the research domain criteria: The bridge between cognitive functioning and physiology. *The Clinical Neuropsychologist*, 33(2), 327–356. <https://doi.org/10.1080/13854046.2018.1523467>
- Germine, L., Reinecke, K., Chaytor, N. S. (2019). Digital neuropsychology: Challenges and opportunities at the intersection of science and software. *The Clinical Neuropsychologist*, 33(2), 271–286. <https://doi.org/10.1080/13854046.2018.1535662>
- Goldberg, E. (red.). (1990). *Contemporary neuropsychology and the legacy of Luria*. Lawrence Erlbaum Associates Publishers.
- Golden, C. J., Hammeke, T. A., Purisch, A. D. (1980). *Luria-Nebraska Neuropsychological Battery. Manual*. Western Psychological Services.

- Guedalia, J., Finkelstein, Y., Drukker, A., Frishberg, Y. (2000). The use of Luria's method for the neurobehavioral assessment of encephalopathy in an adolescent: Application in a rehabilitation setting. *Archives of Clinical Neuropsychology*, 15(2), 177–184. <https://doi.org/10.1093/arclin/15.2.177>
- Howieson, D. (2019). Current limitations of neuropsychological tests and assessment procedures. *The Clinical Neuropsychologist*, 33(2), 200–208. <https://doi.org/10.1080/13854046.2018.1552762>
- Janzen, L. A., Guger, S. (2016). Clinical neuropsychology practice and training in Canada. *The Clinical Neuropsychologist*, 30(8), 1193–1206. <https://doi.org/10.1080/13854046.2016.1175668>
- Jodzio, K. (2011). *Diagnostyka neuropsychologiczna w praktyce klinicznej*. Wydawnictwo Difin.
- Kądziaława, D. (1983). *Czynności rozumienia mowy. Analiza neuropsychologiczna*. Zakład Narodowy im. Ossolińskich. Wydawnictwo Polskiej Akademii Nauk.
- Kądziaława, D. (1998). Zaburzenia językowe po uszkodzeniach struktur podkorowych mózgu. W: A. Herzyk i D. Kądziaława (red.), *Związek mózg–zachowanie w ujęciu neuropsychologii klinicznej* (s. 111–155). Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
- Kądziaława, D. (2002). Neuropsychologia kliniczna: charakterystyka dyscypliny. W: J. Strelau (red.), *Psychologia. Podręcznik akademicki* (t. 3, s. 649–662). Gdańskie Wydawnictwo Psychologiczne.
- Kessels, R. P. C. (2019). Improving precision in neuropsychological assessment: Bridging the gap between classic paper-and-pencil tests and paradigms from cognitive neuroscience. *The Clinical Neuropsychologist*, 33(2), 357–368. <https://doi.org/10.1080/13854046.2018.1518489>
- Klimkowski, M. (1987). Poglądy na związek psychiki z mózgiem a diagnoza neuropsychologiczna. W: M. Klimkowski i A. Herzyk (red.), *Diagnoza neuropsychologiczna. Przegląd zagadnień* (s. 9–30). Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
- Kondratak, B., Pokropek, A. (2013). IRT i pomiar edukacyjny. *Edukacja*, 124(4), 42–66.
- Kordiyak, D., Shakhovska, N. (2016). Analytical review of medical diagnostic systems. *Econtechmod: An International Quarterly Journal on Economics of Technology and Modelling Processes*, 5(2), 11–16.
- Lezak, M. D. (1995). *Neuropsychological assessment* (wyd. 3). Oxford University Press.
- Lezak, M. D., Howieson, D. B., Bigler, E. D., Tranel, D. (2012). *Neuropsychological assessment* (wyd. 5). Oxford University Press.
- Łojek, E. (2009). Imaging communication in the brain. W: K. Bryan (red.), *Interdisciplinary Communication Studies: t. 1. Communication in the Healthcare* (s. 1–65). Peter Lang European Academic Publishers.

- Łojek, E. (21 maja 2014). *Recent developments in neuropsychology and their possible influence on testing and assessment* [wykład]. European Test Publisher Group, Istambuł, Turcja.
- Łojek, E. (11 kwietnia 2019). *Are modern neuropsychological methods really modern?* [wykład]. Spotkanie naukowe w ramach grantu ERASMUS PLUS Mobility Staff, Helsinki, Finlandia.
- Łojek, E., Stańczak, J. (2010a). *CVLT. Podręcznik do Kalifornijskiego testu uczenia się językowego Deana C. Delisa, Joela H. Kramera, Edith Kaplan i Beth A. Ober. Polska normalizacja*. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Łojek, E., Stańczak, J. (30 czerwca–3 lipca 2010b). *Developing connections between qualitative and quantitative approaches in neuropsychological assessment* [wykład]. International Neuropsychological Society 2010 Mid-Year Meeting, Kraków, Polska.
- Łojek, E., Stańczak, J. (2012). *CTT. Kolorowy test połączeń. Wersja dla dorosłych. Podręcznik. Polska normalizacja*. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Łucki, W. (1967). *Zestaw prób do badania zaburzeń wyższych czynności nerwowych u chorych z ogniskowymi uszkodzeniami mózgu*. Zespół Neurochirurgii Centrum Medycyny Doświadczalnej i Klinicznej Polskiej Akademii Nauk.
- Łucki, W. (1995). *Zestaw prób do badania procesów poznawczych u pacjentów z uszkodzeniami mózgu. Zeszyty A, B, C i D*. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Łuria, A. (1967). *Zaburzenia wyższych czynności korowych wskutek ogniskowych uszkodzeń mózgu. Wprowadzenie do neuropsychologii*. Państwowe Wydawnictwo Naukowe.
- Marcopulos, B., Łojek, E. (6–8 lipca 2016). *Are modern neuropsychological testing and assessment really “modern”?* [sympozjum]. Konferencja International Neuropsychological Society Meeting, Londyn, Wielka Brytania.
- Marcopulos, B. A., Łojek, E. (2019). Introduction to the special issue: Are modern neuropsychological assessment methods really “modern”? Reflections on the current neuropsychological test armamentarium. *The Clinical Neuropsychologist*, 33(2), 187–199. <https://doi.org/10.1080/13854046.2018.1560502>
- Maruszewski, M. (1969). Zadania psychologa w klinice ogniskowych uszkodzeń mózgu. W: A. Lewicki (red.), *Psychologia kliniczna* (s. 322–361). Państwowe Wydawnictwo Naukowe.
- Mikadze, Y. V., Ardila, A., Akhutina, T. V. (2019). A. R. Luria’s approach to neuropsychological assessment and rehabilitation. *Archives of Clinical Neuropsychology*, 34, 795–802. <https://doi.org/10.1093/arclin/acy095>

- Parsons, T. D., Duffield, T. (2019). National Institutes of Health initiatives for advancing scientific developments in clinical neuropsychology. *The Clinical Neuropsychologist*, 33(2), 246–270. <https://doi.org/10.1080/13854046.2018.1523465>
- Price, C. J. (2018). The evolution of cognitive models: From neuropsychology to neuroimaging and back. *Cortex*, 107, 37–49. <https://doi.org/10.1016/j.cortex.2017.12.020>
- Rabin, L. A., Barr, W. B., Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, 20, 33–65. <https://doi.org/10.1016/j.acn.2004.02.005>
- Rabin, L. A., Paolillo, E., Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology*, 3, 206–230. <https://doi.org/10.1093/arclin/acw007>
- Schmand, B. (2019). Why are neuropsychologists so reluctant to embrace modern assessment techniques? *The Clinical Neuropsychologist*, 33(2), 209–219. <https://doi.org/10.1080/13854046.2018.1523468>
- Siuta, J. (2006). *NEO-PI-R. Inwentarz Osobowości Paula T. Costy Jr. i Roberta R. McCrae. Adaptacja polska. Podręcznik*. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Stańczak, J. Matczak, A., Jaworowska, A., Bac, I. (2020). *WISC-V. Skale inteligencji Wechslera dla dzieci* (wyd. 5). Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Strauss, E., Sherman, E. M. S., Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary*. Oxford University Press.
- Strelau, J. (2008). *Temperament as a regulator of behavior: After fifty years of research*. Eliot Werner Publications.
- Sweet, J. J., Peck, E. A., III, Abramowitz, C., Etzweiler, S. (2002). National Academy of Neuropsychology/Division 40 of the American Psychological Association Practice Survey of Clinical Neuropsychology in the United States, Part I: Practitioner and practice characteristics, professional activities, and time requirements. *The Clinical Neuropsychologist*, 16(2), 109–127. <https://doi.org/10.1076/clin.16.2.109.13237>
- Toeplitz, Z. (1982). *Krytyczna analiza Kwestionariusza osobowości H. J. Eysencka (Eysenck Personality Questionnaire, EPQ) pod kątem próby adaptacji tego testu do warunków polskich*. Zakład Narodowy im. Ossolińskich.
- Zawadzki, B., Strelau, J. (2018). Formal characteristics of behavior: Temperament Inventory. W: V. Zeigler-Hill i T. K. Shackelford (red.), *Encyclopedia of personality and individual differences*. Springer. https://doi.org/10.1007/978-3-319-28099-8_2285-1

AGNIESZKA E. ŁYŚ  <https://orcid.org/0000-0001-8346-1862>

Wydział Psychologii i Kognitywistyki, Uniwersytet im. Adama Mickiewicza
w Poznaniu

KAMILA BARGIEL-MATUSIEWICZ  <https://orcid.org/0000-0002-9467-2267>

Wydział Psychologii, Uniwersytet Warszawski

Diagnoza psychologiczna związana z postępowaniem dowodowym w sytuacjach przemocy seksualnej – wyzwanie dla współczesnej psychologii sądowej

Psychological assessment related to taking evidence in situations of sexual violence – a challenge for modern forensic psychology

Autor korespondencyjny: Agnieszka E. Łyś

e-mail: agnieszka.lys@amu.edu.pl

Publikacja finansowana z dotacji MNiSW dla Uniwersytetu im. Adama Mickiewicza
w Poznaniu (506000, 4305000000, BN002024, 511).

Sposób cytowania: Łyś, A. E., Bargiel-Matusiewicz, K. (2024). Diagnoza psychologiczna związana z postępowaniem dowodowym w sytuacjach przemocy seksualnej – wyzwanie dla współczesnej psychologii sądowej. W: B. Zawadzki, K. Fronczyk i K. Bargiel-Matusiewicz (red.), *Wybrane problemy współczesnej diagnostyki klinicznej w psychologii* (s. 47–64). Wydawnictwo Liberi Libri. <https://doi.org/10.47943/lib.9788363487676.rozdzial02>

Streszczenie

Problematyka podjęta w niniejszym rozdziale dotyczy zagadnień diagnozy psychologicznej w kontekście perspektywy prawnej oraz społecznej zjawiska przemocy seksualnej. Analizie zostają poddane możliwości i ograniczenia metod analizy treści zeznań, takich jak *Statement Validity Assessment*, *Multivariable Adults' Statement Assessment Method*, *Psychologically Based Credibility Assessment Tool* czy *Global Evaluation System*. W rozdziale opisane są także narzędzia do diagnozowania wystąpienia przemocy seksualnej, jak *Child Sexual Behavior Inventory-2* – narzędzie służące do oceny prawdopodobieństwa wystąpienia doświadczenia wykorzystania seksualnego u dzieci i *Pict-o-Stat* – narzędzie służące do oceny prawdopodobieństwa wystąpienia doświadczenia wykorzystania seksualnego u osób z niepełnosprawnościami. Została omówiona także rola przekonań społecznych na temat problemu przestępstw seksualnych.

Słowa kluczowe: przemoc seksualna, przestępstwa seksualne, opiniowanie sądowo-psychologiczne, opiniowanie sądowo-seksuologiczne

Abstract

The problems raised in this chapter concern the issues of psychological assessment in the context of the legal and social perspective of the phenomenon of sexual violence. The possibilities and limitations of the methods of analysis of the content of witness statements, such as *Statement Validity Assessment*, *Multivariable Adults' Statement Assessment Method*, *Psychologically Based Credibility Assessment Tool*, or *Global Evaluation System*, are analyzed. The chapter also describes tools for assessing the occurrence of sexual violence, such as the *Child Sexual Behavior Inventory-2* – a tool for assessing the likelihood of experiencing sexual abuse in children, and *Pict-o-Stat* – a tool for assessing the likelihood of experiencing sexual abuse in people with disabilities. The role of social beliefs concerning the problem of sexual crimes was also discussed.

Keywords: sexual violence, sexual crimes, forensic psychological assessment, forensic sexual assessment

Problematyka podjęta w niniejszym rozdziale dotyczy zagadnień diagnozy psychologicznej w kontekście perspektywy prawnej oraz społecznej zjawiska przemocy seksualnej. Według Europejskiego Instytutu ds. Równości Płci (European Institute for Gender Equality, 2017) pojęcie przemocy seksualnej oznacza jakiegokolwiek zachowanie seksualne dokonane bez zgody ofiary. Warto mieć na uwadze, że pod to pojęcie podlegają zarówno zachowania seksualne związane z kontaktem fizycznym, jak np. wymuszona penetracja czy niechciany dotyk, jak i niezwiązane z kontaktem fizycznym, jak np. zmuszanie kogoś do oglądania treści o charakterze seksualnym (Smith i in., 2017). Jak wynika z literatury przedmiotu oraz dostępnych statystyk bardzo niewielka część przestępstw seksualnych ostatecznie podlega karze. Na podstawie dostępnych oszacowań można stwierdzić, że prawdopodobnie zaledwie ok. 15% przypadków przemocy seksualnej wobec kobiet w Unii Europejskiej trafia do wiadomości organów ścigania (Fundamental Rights Agency, 2014). Ponadto w Polsce bardzo duży odsetek spraw dotyczących zgwałceń, zgłaszanych na policję, kończy się umorzeniem bądź odmową wszczęcia postępowania, przy czym większość z tych decyzji opiera się na braku wystarczających danych uzasadniających stwierdzenie popełnienia przestępstwa (Dudka, 2012). Jak bowiem podkreśla Ferenz (2014), niejednokrotnie zeznania ofiary zgwałcenia są jedynym źródłem dowodowym w sprawie. Stąd też w sprawach dotyczących przestępstw seksualnych niezwykle ważna jest rola psychologa jako biegłego sądowego.

Z uwagi na wykazywany w literaturze przedmiotu związek doświadczenia przemocy seksualnej z zaburzeniami, takimi jak np. depresja czy PTSD (np. Chen i in., 2010), badanie osoby pokrzywdzonej przestępstwem seksualnym wymaga ogromnej wrażliwości i uważności ze strony specjalisty przeprowadzającego badanie. Niezwykle ważna jest rola psychologa obecnego przy przesłuchaniu pokrzywdzonej osoby małoletniej – ma on za zadanie ją wspierać i dbać o to, by stres związany z przesłuchaniem był jak najmniej dotkliwy (Toeplitz-Winiewska, 2017). Zarazem jednak biegły psycholog powinien mieć na uwadze, że jego rola istotnie różni się od roli psychoterapeuty – głównym

celem jego działania nie jest niesienie pomocy, lecz odpowiedź na pytania organu procesowego, stąd nieodzowne w pracy biegłego psychologa sądowego są neutralność i obiektywność (Łyś i in., 2019; Rode i in., 2020). Konieczność pogodzenia tych wymagań to niemałe wyzwanie. Dlatego niezmiernie ważne jest stosowanie trafnych i rzetelnych narzędzi diagnostycznych, pozwalających uzyskać odpowiedzi na pytania organów procesowych, pojawiające się w toku spraw dotyczących przestępstw seksualnych.

OCENA PSYCHOLOGICZNEJ WIARYGODNOŚCI ZEZNAŃ – MOŻLIWOŚCI I OGRANICZENIA

Jednym z pytań zadawanych biegłym psychologom przez organy procesowe jest pytanie o ocenę psychologicznej wiarygodności zeznań. Rzecz jasna, ocena wiarygodności zeznań należy do organu procesowego, nie zaś do biegłego psychologa, niemniej jednak w praktyce biegłych psychologów sądowych używane są metody analizy treści, mające na celu określenie prawdopodobieństwa tego, że zeznania świadka są szczere. Dlatego niezmiernie ważne jest rozróżnienie między pojęciem wiarygodności zeznań – ocenianej przez sąd na podstawie tego, czy treść zeznań jest spójna z pozostałym materiałem dowodowym, a pojęciem wiarygodności psychologicznej – ocenianej przez biegłych psychologów sądowych na podstawie tego, czy treść zeznań spełnia psychologiczne kryteria wiarygodności (Pastwa-Wojciechowska i in., 2023).

Psychologiczne kryteria wiarygodności opierają się na hipotezie Undeutscha, według którego zeznania szczere różnią się treścią od nieszczerých (Undeutsch, 1967, za: Steller, 1989). Do metod opartych na tychże kryteriach należą np. *Statement Validity Assessment* (SVA; Steller, 1989), *Multivariable Adults' Statement Assessment Method* (MASAM; Wojciechowski i in., 2018) czy *Psychologically Based Credibility Assessment Tool* (PBCAT; Evans i in., 2013).

SVA (Steller, 1989) jest metodą stosowaną nie tylko w Polsce, lecz także w USA oraz w kilku krajach Europy Zachodniej, w tym Holandii, Szwecji, Hiszpanii i Niemczech (Vrij, 2014). SVA zawiera *Criteria-Based Content Analysis* (CBCA) – zbiór 19 kryteriów, mających świadczyć o szczerości osoby składającej zeznania (jak np. logiczna struktura czy elementy odwołujące się do własnego stanu psychicznego w momencie zdarzenia) oraz *Listę kontrolną prawdziwości*, uwzględniającą takie czynniki, jak np. podatność na sugestie czy ewentualną motywację do składania fałszywych zeznań (Rassin, 2000). Rzetelność i trafność tej metody stoją jednak pod znakiem zapytania. Krytycy

zwracają uwagę na brak ustalonej liczby kryteriów, jakie muszą być spełnione, by zeznanie można było uznać za szczere, brak standaryzacji, prowadzący do wybiórczego stosowania kryteriów oraz znacznego wpływu subiektywnej oceny osoby przeprowadzającej badanie na wnioski formułowane w oparciu o użycie tego narzędzia (Arce, 2017; Burchard, 2009; Rassin, 2000; Ruby i Brigham, 1997; Vrij i Akehurst, 1998/2003). Pod znakiem zapytania stoi też kwestia uniwersalności kulturowej tego narzędzia – Ruby i Brigham (1998), porównując wypowiedzi osób o różnym pochodzeniu etnicznym, wykazali, że niektóre kryteria CBCA są silniejszym predyktorem prawdy u jednej z grup etnicznych, a słabszym u innej. Warto mieć na uwadze, że choć obecnie narzędzie to stosuje się także do analizy zeznań świadków dorosłych, pierwotnie miało ono służyć do analizy zeznań małoletnich świadków, a kryteria CBCA zostały wyłonione na podstawie badań eksperymentalnych prowadzonych wśród dzieci i młodzieży (Steller, 1989). Ponadto kryteria CBCA wskazują na szczerść zeznania, jednak brak spełnienia odpowiedniej liczby kryteriów nie dowodzi, że zeznanie jest nieszczerze (Steller, 1989). Należy także podkreślić, że nawet fałszywe zeznania zawierają niektóre elementy prawdziwe, co może prowadzić do zawyżenia liczby kryteriów prawdziwości (Rassin, 2000; Steller, 1989). Według Rassin (2000) wykrywalność zeznań prawdziwych za pomocą CBCA wynosi 80%, jednak wykrywalność zeznań fałszywych zaledwie ok. 8–35%. Amado i in. (2016) na podstawie przeprowadzonej przez siebie metaanalizy zwracają uwagę, że wielkości efektów mierzące stopień, w jakim zastosowanie CBCA pozwala na odróżnienie zeznań prawdziwych od fałszywych znacznie się różnią pomiędzy poszczególnymi badaniami, np. w badaniu Arce i in. (2000, za: Amado i in., 2016) okazało się, że zaledwie 51% zeznań prawdziwych zawierało więcej spełnionych kryteriów szczerości niż zeznania fałszywe, a co więcej – na podstawie kryteriów CBCA aż 44% prawdziwych zeznań branych pod uwagę w kolejnym badaniu Arce i in. (2010) zostało błędnie zidentyfikowanych jako fałszywe. Jak podkreśla Arce (2017), tak duży błąd rozpoznania sprawia, że kryteria CBCA nie spełniają kryteriów Dauberta, według których, aby badanie daną metodą można było wykorzystywać jako dowód w postępowaniu sądowym, metoda musi mieć potwierdzoną trafność i rzetelność (Fournier, 2016). Oznacza to, że analiza oparta na kryteriach CBCA nie powinna być wykorzystywana jako dowód w postępowaniu sądowym (Arce, 2017). Ponadto nie wszystkie kryteria CBCA w równym stopniu pozwalają odróżnić zeznania prawdziwe od fałszywych. Według metaanalizy Amado i in. (2016) niektóre kryteria szczerości, np. duża

liczba szczegółów, charakteryzują się wysoką trafnością, inne zaś, jak np. próby usprawiedliwienia sprawcy, mają znikomą trafność.

Wskutek ograniczeń SVA, pokazanych przez wymienione wyżej badania, badacze zajmujący się problematyką psychologii zeznań świadków zaczęli poszukiwać nowych, doskonalszych metod odróżniania zeznań szczerych od nieszczerych. Jedną z owych metod jest GES (*Global Evaluation System*), stworzona przez Arce i Fariña (2005). Kryteria szczerości według GES opierają się na zmodyfikowanej wersji CBCA. Ponadto GES bierze pod uwagę też czynniki nieuwzględniane w SVA, jak objawy PTSD (*post-traumatic stress disorder*) będącego skutkiem przestępstwa i konieczność wykluczenia ewentualnej symulacji tych objawów, a także sposób przeprowadzenia przesłuchania i dostosowanie go do wieku i sprawności świadka. Wstępne analizy wskazują na to, że GES może być przydatną metodą analizy zeznań (Arce, 2017), jednak brak odpowiednio dużej liczby badań uniemożliwia przeprowadzenie metaanalizy i bardziej wnikliwą ocenę wartości diagnostycznej narzędzia. Wymaga ono zatem dalszych badań.

Kolejnym narzędziem, mającym za zadanie pomóc w odróżnieniu zeznań szczerych od nieszczerych, jest PBCAT (Evans i in., 2013). Składa się ono z zestawu tych kryteriów CBCA, które okazały się najbardziej rzetelne i trafne w dotychczasowych badaniach, a także z dodatkowych kryteriów, jak np. związanych z przeciążeniem poznawczym, opartych na założeniu, że kłamstwo wymaga więcej wysiłku niż mówienie prawdy.

Badanie przeprowadzone przez Evans i in. (2013) wskazuje na trafność i rzetelność narzędzia. Jednak, podobnie jak w przypadku GES, brak odpowiednio dużej liczby badań uniemożliwia przeprowadzenie metaanalizy i bardziej wnikliwą ocenę wartości diagnostycznej narzędzia. Ponadto, aby mieć pełen obraz trafności i rzetelności narzędzia, a co za tym idzie, możliwości zastosowania go w postępowaniu dowodowym, należałoby porównać je z SVA.

Innym narzędziem, mającym pomóc w odróżnieniu zeznań szczerych od nieszczerych, jest MASAM (Wojciechowski i in., 2018). Metoda ta zawiera dodatkowe kryteria, biorące pod uwagę fakt, że konkretne zeznania mogą zawierać elementy zarówno fałszywe, jak i prawdziwe. Wyniki badań przeprowadzonych przez Wojciechowskiego i in. (2018) sugerują, że MASAM jest trafnym i rzetelnym narzędziem, jednak, podobnie jak w przypadku GES i PBCAT, brak odpowiednio dużej liczby badań uniemożliwia przeprowadzenie metaanalizy i bardziej wnikliwą ocenę wartości diagnostycznej narzędzia.

Problemem dotyczącym obecnej wiedzy na temat trafności i rzetelności analizy treści zeznań jest także to, że w dużej mierze opiera się ona na bada-

niach eksperymentalnych (Amado i in., 2016; Arce, 2017; Steller, 1989). Ocena szczerości wypowiedzi generowanych w warunkach laboratoryjnych ma niską trafność ekologiczną – trudno ekstrapolować prawidłowości dotyczące wypowiedzi generowanych w takich warunkach na warunki procesu sądowego, zwłaszcza biorąc pod uwagę, że sytuacja składania zeznań dotyczących przestępstwa seksualnego wiąże się z silnymi emocjami, których nie da się w sposób etyczny stworzyć w laboratorium. W przyszłości zatem warto byłoby przeprowadzić badania dotyczące rzeczywistych spraw sądowych.

NARZĘDZIA DO DIAGNOZOWANIA PRZEMOCY SEKSUALNEJ

Dużym problemem w psychologii sądowej jest też to, że objawy doświadczenia przemocy seksualnej są bardzo niespecyficzne. Dobrze widać to na przykładzie wykorzystywania seksualnego dzieci. Przegląd systematyczny Kendall-Tackett i in. (1993) wskazuje, że symptomy wielu zaburzeń, jak np. zaburzenia lękowe czy depresyjne, częściej pojawiają się u dzieci, które doświadczyły wykorzystania seksualnego niż u dzieci, które nie mają takich doświadczeń. Podobne problemy dotyczą dorosłych ofiar przestępstw seksualnych (Chen i in., 2010). Nie ma jednak żadnego specyficznego zestawu symptomów, które składałyby się na syndrom charakterystyczny dla osób pokrzywdzonych przestępstwami seksualnymi (Kendall-Tackett i in., 1993), co również utrudnia diagnozę przemocy seksualnej.

W przypadku diagnozowania wykorzystywania seksualnego u dzieci przydatne mogą być kryteria normalności aktywności seksualnej dziecka lub nastolatka, opracowane przez Beisert (2006, 2008b). Według Beisert aktywność seksualna dziecka lub nastolatka mieści się w normie rozwojowej, jeśli:

- nie utrudnia realizacji zadań rozwojowych przewidzianych dla okresu, w którym aktualnie znajduje się dziecko/nastolatek,
- mieści się w repertuarze zachowań seksualnych charakterystycznych dla danego wieku,
- dokonuje się między osobami będącymi w zbliżonym wieku,
- oparta jest na zasadzie dobrowolności,
- prowadzi do realizacji celów seksualnych (tu należy odróżnić np. masturbację mającą na celu rozładowanie napięcia seksualnego od masturbacji mającej na celu zredukowanie poziomu stresu),
- nie narusza zdrowia,
- nie narusza rażąco przyjętego porządku społecznego.

Kryteria te są w dużej mierze zbieżne z kryteriami normy partnerskiej w seksuologii, stworzonymi w odniesieniu do osób dorosłych, kładącymi nacisk na dobrowolność i zgodę (Lew-Starowicz i Stusiński, 2015b). Zachowania seksualne dziecka niezgodne z normą rozwojową, choć nie stanowią bezpośredniego dowodu na to, że dziecko doświadczyło wykorzystania seksualnego, mogą wskazywać na takie doświadczenie.

Friedrich i in. (1992, 2012) stworzyli narzędzie *Child Sexual Behavior Inventory-2* (CSBI-2), mające na celu pomoc w ocenie prawdopodobieństwa, że dziecko doświadczyło wykorzystywania seksualnego. Narzędzie zostało przygotowane w oparciu o porównanie częstotliwości występowania 36 zachowań seksualnych u dzieci w wieku 2–12 lat, które doświadczyły wykorzystania seksualnego i tych, które go nie doświadczyły. Zachowania te są podzielone na kilka kategorii: przestrzeganie granic osobistych, ekshibicjonizm, zachowania związane z rolami płciowymi, autostymulacja, niepokój seksualny, zainteresowanie seksualnością, natarczywość seksualna, wiedza seksualna oraz zachowania voyeurystyczne. Zachowania te były mierzone na podstawie kwestionariusza wypełnionego przez matki lub opiekunki dzieci. Rzetelność narzędzia okazała się wysoka ($\alpha = 0,93$ u dzieci, które doświadczyły wykorzystania seksualnego i $\alpha = 0,82$ u tych, które go nie doświadczyły), jak i stabilność bezwzględna ($r = 0,85$). Okazało się także, że w przypadku aż 27 zachowań wystąpiły istotne różnice między dziećmi, które doświadczyły wykorzystywania seksualnego i tymi, które go nie doświadczyły, co – zdaniem autorów – wskazuje na wysoką trafność narzędzia. Co więcej, Friedrich i in. (1997) wykazali, że dzieci, które doświadczyły wykorzystania seksualnego mają wyższy wskaźnik zachowań wymienionych w CSBI-2 nie tylko w porównaniu z dziećmi z grup nieklinicznych, ale także z dziećmi mającymi zaburzenia psychiczne, a niemającymi za sobą doświadczenia wykorzystania seksualnego. Oznacza to, że wyniki dzieci, które doświadczyły wykorzystania seksualnego różnią się od wyników dzieci, które zmagają się z zaburzeniami psychicznymi mającymi inną etiologię. Kolejna wersja narzędzia, *Child Sexual Behavior Inventory-3* (CSBI-3; Friedrich i in., 2001), w której część pozycji z CSBI-2 została przeformułowana pod kątem klarowności, również ma satysfakcjonującą zgodność wewnętrzną ($\alpha = 0,93$ u dzieci, które doświadczyły wykorzystania seksualnego, $\alpha = 0,83$ u tych, które go nie doświadczyły, ale leczą się psychiatrycznie z innych powodów i $\alpha = 0,72$ u dzieci z grupy nieklinicznej, co autorzy wyjaśniają mniejszą wariancją wyników występującą w tej grupie) i stabilność bezwzględną ($r = 0,91$). Okazało się jednak, że różnice między dziećmi, które doświadczyły wykorzystania

seksualnego, a dziećmi z grupy mającej zaburzenia psychiczne spowodowane innymi czynnikami są mniejsze niż różnice między dziećmi, które doświadczyły wykorzystania seksualnego, a dziećmi z grupy niemającej stwierdzonych zaburzeń psychicznych, co oznacza, że trafność różnicowa tego narzędzia wymaga dalszych badań. Ponadto z przeglądu systematycznego dotyczącego właściwości CSBI-2 i CSBI-3, przeprowadzonego przez Vrolijk-Bosschaart i in. (2018) wynika, że, choć wyniki u dzieci, które doświadczyły wykorzystania seksualnego i tych, które nie mają takich doświadczeń, różnią się istotnie, to w niektórych grupach narzędzie charakteryzuje się niską czułością, co może przyczynić się do niewykrycia wykorzystania seksualnego, a w innych niską specyficznością, co może przyczynić się do błędnej diagnozy wykorzystania seksualnego. Poza tym, jak podkreślają Vrolijk-Bosschaart i in. (2018), nie wszystkie dzieci wykorzystywane seksualnie przejawiają nienormatywne zachowania seksualne. Brak takich zachowań nie może zatem stanowić dowodu, że do wykorzystania seksualnego nie doszło. Vrolijk-Bosschaart i in. (2018) wskazują także, że CSBI jako metoda oparta na danych deklarowanych przez rodziców, nie jest odporna na sytuację, w której to sam sprawca bierze udział w badaniu. Sytuacja taka jest całkiem prawdopodobna, biorąc pod uwagę, że ok. 20% sprawców wykorzystywania seksualnego dzieci to dorośli członkowie rodziny (da Costa i in., 2022; Ferragut i in., 2021). Ponadto, jak zwraca uwagę Beisert (2008a), niekiedy zdarza się, że partnerki sprawców wykorzystania seksualnego dzieci nie chcą, aby ich czyny ujrzały światło dzienne, co także może prowadzić do udzielania nieszczerých odpowiedzi. Ponadto CSBI nie jest narzędziem neutralnym kulturowo, np. autorzy koreańskiej wersji CSBI (Jin i in., 2013) wskazują, że występowanie niektórych zachowań seksualnych było negowane przez niemal wszystkich rodziców biorących udział w opisywanych przez nich badaniach, co – jak przyznają sami autorzy – może być spowodowane konserwatywnymi wartościami, inspirowanymi konfucjanizmem. Stworzenie polskiej wersji CSBI wymagałoby zatem gruntownych badań w polskiej próbie.

OSOBY Z NIEPEŁNOSPRAWNOŚCIAMI JAKO OFIARY PRZEMOCY SEKSUALNEJ

Dotychczasowe badania pokazują, że, w porównaniu z populacją ogólną, osoby z niepełnosprawnościami są bardziej narażone na doświadczenie przemocy seksualnej (Basile i in., 2016; Breiding i Armour, 2015; Martin i in., 2006; Mitra i in., 2016). W przypadku osób z niektórymi rodzajami

niepełnosprawności rodzi to dodatkowe trudności w doborze narzędzi diagnostycznych. W przypadku osób głuchych problemem może być przykładowo znalezienie biegłego psychologa sądowego, który włada Polskim Językiem Migowym. Ponadto przeciętne narzędzia diagnostyczne nie są dostosowane do osób z niepełnosprawnością intelektualną, co jest o tyle ważne, że osoby z niepełnosprawnością intelektualną, w porównaniu do osób z innymi typami niepełnosprawności, częściej doznają przemocy seksualnej (Lin i in., 2009). Obiecującym sposobem na ograniczenie tego problemu wydają się narzędzia oparte na metodzie *Audio Computer-Assisted Self-Interview* (ACASI), np. *Pic-O-Stat* (Kristensen i Möller, 2017). Pytania w tym narzędziu są sformułowane w prostszy sposób niż w standardowych narzędziach do badania doświadczeń przemocy seksualnej. Oprócz wersji pisemnej mają też wersję audio, a ponadto są opatrzone rysunkami. Badanie pilotażowe przeprowadzone wśród dziewięciu osób z niepełnosprawnością intelektualną mieszkających w Szwecji wstępnie wykazało, że pytania są zrozumiałe dla osób z niepełnosprawnością intelektualną w stopniu lekkim. Badania ilościowe (Punjabi i in., 2021; van der Elst i in., 2009) wskazują, że odsetek doświadczeń przemocy okazuje się wyższy, gdy osoby badane odpowiadają przy użyciu narzędzi opartych na ACASI niż wtedy, gdy odpowiadają przy użyciu tradycyjnych metod, co może wskazywać na wyższą wykrywalność przemocy, gdy jest ona badana przy użyciu narzędzi opartych na ACASI. Aby jednak tego typu narzędzie dało się zastosować do diagnozy indywidualnej, niezbędne są badania nad jego czułością i specyficznością. Falb i in. (2016) na podstawie badań przeprowadzonych we wschodniej części Demokratycznej Republiki Konga oraz w obozach uchodźczych na terenie Etiopii wskazują, że narzędzia oparte na ACASI mogą być przydatne także do diagnozowania doświadczenia przemocy seksualnej wśród osób, dla których język, w jakim odbywa się badanie, nie jest językiem ojczystym, jak np. w środowiskach uchodźczych.

INNE TRUDNOŚCI ZWIĄZANE Z DIAGNOZĄ PRZEMOCY SEKSUALNEJ

Trudności dowodowe nie są jedyną prawdopodobną przyczyną niskiego odsetka wyroków skazujących za przestępstwa seksualne. Inną przyczyną mogą być np. mity na temat zgwałceń, ujmowane w literaturze jako podszyte uprzedzeniami, stereotypowe lub fałszywe przekonania dotyczące zgwałceń, ofiar zgwałceń i gwałcicieli (Burt, 1980). Według Eurobarometru (European

Commission, 2016) 27% mieszkańców Unii Europejskiej twierdzi, że istnieją okoliczności, które usprawiedliwiają aktywność seksualną bez zgody drugiej strony (np. upojenie alkoholowe czy wcześniejszy flirt ze sprawcą). Od mitów na temat zgwałceń nie są wolni również przedstawiciele organów ścigania. Badania wśród amerykańskich policjantów pokazują, że co prawda odsetek policjantów akceptujących mity na temat zgwałceń jest niski, ale jednak wielu policjantów przeszacowuje odsetek fałszywych oskarżeń wśród wszystkich zgłoszeń dotyczących zgwałceń (Mennicke i in., 2014; Venema, 2018). Wskazuje na to także badanie Constantinou i Butorac (2023), którzy wykazali, że aż 60,9% z 220 cypryjskich policjantów, którzy brali udział w badaniu, wierzy, że kobiety często fałszywie oskarżają mężczyzn, aby się na nich zemścić lub otrzymać odszkodowanie. Przekonania te nie mają odzwierciedlenia w statystykach. Badania nie wskazują bowiem, by odsetek fałszywych oskarżeń o przestępstwa seksualne był wyższy niż odsetek fałszywych oskarżeń o inne przestępstwa (Lisak i in., 2010). Ponadto wiedza na temat zgwałceń niekoniecznie musi przekładać się na postawy wobec osób zgłaszających zgwałcenie. Page (2010), która przeprowadziła badanie wśród 891 amerykańskich policjantów, wskazała, że 93% z nich zgodziło się z twierdzeniem, że każda kobieta może zostać zgwałcona, jednak 18% z nich miałooby problem z uwierzeniem zamężnej kobiecie zgłaszającej zgwałcenie przez męża, zaś aż 44% z uwierzeniem osobie świadczącej usługi seksualne, zgłaszającej zgwałcenie. Krahé i in. (2008) wskazują, że nawet studenci prawa oraz uczestnicy aplikacji prawniczych nie są wolni od stereotypowego postrzegania zgwałceń – są bowiem skłonni uważać ofiarę za bardziej winną, jeśli przed zdarzeniem spożywała alkohol oraz jeśli знаła sprawcę. To drugie zjawisko jest o tyle groźne, że, jak pokazują badania (np. Fundamental Rights Agency, 2014), większość zgwałceń popełniana jest przez osoby znane ofierze. Co więcej, Bitsch i Klemetsen (2017) na podstawie badań rzeczywistych spraw sądowych w Norwegii wskazują, że sprawcy zgwałceń, którzy znali wcześniej ofiarę, otrzymują o 18% niższe wyroki, niż ci, którzy w momencie zdarzenia byli obcy dla ofiary, co pokazuje, że stereotypowa wizja „prawdziwego gwałtu”, popełnianego przez nieznanego sprawcę w ciemnej uliczce, może mieć realny wpływ na działanie wymiaru sprawiedliwości. Dlatego, oprócz doskonalenia narzędzi diagnostycznych, ważnym wyzwaniem dla psychologów jest także rzetelna edukacja społeczeństwa, dotycząca przyczyn i mechanizmów przestępstw seksualnych.

Nie wiemy, jaka jest skala akceptacji mitów na temat przestępstw seksualnych przez biegłych psychologów i seksuologów sądowych. W literaturze dotyczącej

najczęstszych błędów popełnianych przez biegłych seksuologów (Depko, 2023; Lew-Starowicz i Stusiński, 2015a) pojawiają się jednak m.in. takie błędy, jak moralizowanie, etykietowanie, używanie określeń nacechowanych emocjonalnie, kierowanie się osobistymi opiniami, niemającymi potwierdzenia w wynikach badań naukowych czy kierowanie się stereotypami. Obserwacje te sugerują, że nawet część biegłych psychologów i seksuologów może mieć skłonność do ulegania mitom i stereotypom na temat przestępstw seksualnych. Kwestia ta wymaga zatem dalszych badań.

PODSUMOWANIE

Warto pamiętać, że żadne narzędzie, jakim posługują się biegli psychologowie i seksuologowie, nie daje stuprocentowo pewnej odpowiedzi na pytania organów procesowych (Leśniak i Leśniak, 2018). Opinia psychologiczna bądź seksuologiczna powinna być zatem traktowana jak opinia prawdopodobna, nie zaś kategoryczna. Warto też pamiętać, że nie istnieje jedna, modelowa reakcja na pokrzywdzenie przestępstwem seksualnym, np. Randall (2010) wskazuje na oczekiwania społeczne dotyczące tego, że ofiara zgwałcenia będzie stawiać opór. Tymczasem, jak podkreślają Coxell i King (2010), wiele ofiar zgwałceń doświadcza bezruchu tonicznego w momencie zdarzenia. W literaturze pojawiają się też wzmianki o reakcji typu *fawn*, polegającej na tym, że ofiara zachowuje się wobec sprawcy w sposób uległy, licząc, że jego działania staną się mniej brutalne (Walker, 2013, za: Hicks, 2019). Długofalowe skutki pokrzywdzenia przestępstwem seksualnym również są bardzo różne. Lemieux i Byers (2008) wykazały, że doświadczenie wykorzystania seksualnego w dzieciństwie jest predyktorem ryzykownych zachowań seksualnych, ale także unikania aktywności seksualnej.

Ten brak jednej, modelowej reakcji sprawia, że diagnoza przemocy seksualnej zawsze będzie obciążona błędem, który należy brać pod uwagę. Warto jednak podejmować wszelkie starania, aby ten błąd był jak najmniejszy. Stąd ważne jest, by tworzyć i adaptować rzetelne i trafne narzędzia do diagnozy przemocy seksualnej. Niezmiernie ważna jest także edukacja osób pracujących w organach ścigania i wymiarze sprawiedliwości. McKee i in. (2020) wykazali bowiem, że szkolenie policjantów wpływa na ich postawy wobec ofiar zgwałceń. Lathan i in. (2022) wskazują ponadto na związek wypalenia zawodowego z akceptacją mitów na temat zgwałceń. Ważne jest zatem, by dbać

o edukację zwiększającą świadomość prawidłowości dotyczących przestępstw seksualnych, niwelując błędne przekonania w tym zakresie, ale także – by dbać o dobrostan i przeciwdziałać wypaleniu zawodowemu osób pracujących w organach ścigania i wymiarze sprawiedliwości.

BIBLIOGRAFIA

- Amado, B. G., Arce, R., Fariña, F., Vilariño, M. (2016). Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, 16(2), 201–210. <https://doi.org/10.1016/j.ijchp.2016.01.002>
- Arce, R. (2017). Análisis de contenido de las declaraciones de testigos: Evaluación de la validez científica y judicial de la hipótesis y la prueba forense. *Acción Psicológica*, 14(2), 171–190. <https://doi.org/10.5944/ap.14.2.21347>
- Arce, R., Fariña, F. (2005). Psychological evidence in court on statement credibility, psychological injury, and malingering: The Global Evaluation System (GES). *Papeles Del Psicólogo*, 26, 59–77.
- Arce, R., Fariña, F., Vilariño, M. (2010). Contraste de la efectividad del CBCA en la evaluación de la credibilidad en casos de violencia de género. *Psychosocial Intervention*, 19(2), 109–119. <https://doi.org/10.5093/in2010v19n2a2>
- Basile, K. C., Breiding, M. J., Smith, S. G. (2016). Disability and risk of recent sexual violence in the United States. *American Journal of Public Health*, 106(5), 928–933. <https://doi.org/10.2105/AJPH.2015.303004>
- Beisert, M. (2006). Rozwojowa norma seksuologiczna jako kryterium oceny zachowań seksualnych dzieci i młodzieży. *Dziecko Krzywdzone. Teoria, badania, praktyka*, 5(3), 43–57.
- Beisert, M. (2008a). *Kazirodztwo. Rodzice w roli sprawców* (wyd. 2). Wydawnictwo Naukowe Scholar.
- Beisert, M. (2008b). Kryteria diagnostyczne przydatne dla oceny zachowań seksualnych małego dziecka. *Przegląd Seksuologiczny*, 14, 11–19.
- Bitsch, A., Klemetsen, M. E. (2017). The legal grading of sexual citizenship: Sentencing practices in Norwegian rape cases. *Gender, Place & Culture*, 24(2), 174–188. <https://doi.org/10.1080/0966369X.2017.1298572>
- Breiding, M. J., Armour, B. S. (2015). The association between disability and intimate partner violence in the United States. *Annals of Epidemiology*, 25(6), 455–457. <https://doi.org/10.1016/j.annepidem.2015.03.017>

- Burchard, M. (2009). Ocena wiarygodności dowodu z przesłuchania poznawczego w procesie karnym. *Przegląd Prawa i Administracji*, 81, 175–180.
- Burt, M. R. (1980). Cultural myths and supports for rape. *Journal of Personality and Social Psychology*, 38(2), 217–230. <https://doi.org/10.1037//0022-3514.38.2.217>
- Chen, L. P., Murad, M. H., Paras, M. L., Colbenson, K. M., Sattler, A. L., Goranson, E. N., Elamin, M. B., Seime, R. J., Shinozaki, G., Prokop, L. J., Zirakzadeh, A. (2010). Sexual abuse and lifetime diagnosis of psychiatric disorders: Systematic review and meta-analysis. *Mayo Clinic Proceedings*, 85(7), 618–629. <https://doi.org/10.4065/mcp.2009.0583>
- Constantinou, A. G., Butorac, K. (2023). Rape myth acceptance among Croatian and Cypriot police officers. *Criminology & Criminal Justice*. <https://doi.org/10.1177/17488958231160251>
- Coxell, A. W., King, M. B. (2010). Adult male rape and sexual assault: Prevalence, revictimisation and the tonic immobility response. *Sexual and Relationship Therapy*, 25(4), 372–379. <https://doi.org/10.1080/14681991003747430>
- da Costa, E. L., Faúndes, A., Nunes, R. (2022). The association between victim-offender relationship and the age of children and adolescents who suffer sexual violence: A cross-sectional study. *Jornal de Pediatria*, 98(3), 310–315. <https://doi.org/10.1016/j.jped.2021.07.001>
- Depko, A. (2023). Błędy w opiniowaniu sądowo-seksuologicznym. Rodzaje i przyczyny. W: A. Depko i A. Krasowska (red.), *Seksuologia sądowa III. Opiniowanie seksuologiczne w praktyce* (s. 107–133). Wydawnictwo Lekarskie PZWL.
- Dudka, K. (2012). *Podstawy umorzenia i odmowy wszczęcia postępowania karnego w sprawach o przestępstwa zgwałcenia*. Instytut Wymiaru Sprawiedliwości. <https://iws.gov.pl/wp-content/uploads/2018/08/Dudka-K-Zgwałcenia-podstawy-umorzenia-i-odmowy-wszczęcia-2012.pdf>
- European Commission (2016). *Special Eurobarometer 449: Gender-based violence*. https://publications.europa.eu/resource/cellar/f60437fd-e9db-11e6-ad7c-01aa75e-d71a1.0001.01/DOC_1
- European Institute for Gender Equality (2017). *Glossary of definitions of rape, femicide and intimate partner violence*. <https://eige.europa.eu/publications-resources/publications/glossary-definitions-rape-femicide-and-intimate-partner-violence>
- Evans, J. R., Michael, S. W., Meissner, C. A., Brandon, S. E. (2013). Validating a new assessment method for deception detection: Introducing a Psychologically Based Credibility Assessment Tool. *Journal of Applied Research in Memory and Cognition*, 2(1), 33–41. <https://doi.org/10.1016/j.jarmac.2013.02.002>
- Falb, K., Tanner, S., Asghar, K., Souidi, S., Mierzwa, S., Assazenew, A., Bakomere, T., Mallinga, P., Robinette, K., Tibebu, W., Stark, L. (2016). Implementation of

- Audio-Computer Assisted Self-Interview (ACASI) among adolescent girls in humanitarian settings: Feasibility, acceptability, and lessons learned. *Conflict and Health*, 10(1), artykuł 32. <https://doi.org/10.1186/s13031-016-0098-1>
- Ferenz, J. M. (2014). Prawidłowość trybu ścigania przestępstwa zgwałcenia w kontekście gwarancji konstytucyjnych. *Forum Prawnicze*, 25(5), 28–41.
- Ferragut, M., Ortiz-Tallo, M., Blanca, M. J. (2021). Victims and perpetrators of child sexual abuse: Abusive contact and penetration experiences. *International Journal of Environmental Research and Public Health*, 18(18), artykuł 9593. <https://doi.org/10.3390/ijerph18189593>
- Fournier, L. R. (2016). The Daubert guidelines: Usefulness, utilization, and suggestions for improving quality control. *Journal of Applied Research in Memory and Cognition*, 5, 308–313. <https://doi.org/10.1016/j.jarmac.2016.06.012>
- Friedrich, W. N., Fisher, J., Broughton, D., Houston, M., Shafraan, C. R. (2012). Normatywne zachowania seksualne u dzieci – Nowe ustalenia badawcze. *Dziecko Krzywdzone. Teoria, Badania, Praktyka*, 11(4), 51–69.
- Friedrich, W. N., Fisher, J. L., Dittner, C. A., Acton, R., Berliner, L., Butler, J., Damon, L., Davies, W. H., Gray, A., Wright, J. (2001). Child Sexual Behavior Inventory: Normative, psychiatric, and sexual abuse comparisons. *Child Maltreatment*, 6(1), 37–49. <https://doi.org/10.1177/1077559501006001004>
- Friedrich, W. N., Grambsch, P., Damon, L., Hewitt, S. K., Koverola, C., Lang, R. A., Wolfe, V., Broughton, D. (1992). Child Sexual Behavior Inventory: Normative and clinical comparisons. *Psychological Assessment*, 4, 303–311. <https://doi.org/10.1037/1040-3590.4.3.303>
- Friedrich, W. N., Jaworski, T. M., Huxsahl, J. E., Bengtson, B. S. (1997). Dissociative and sexual behaviors in children and adolescents with sexual abuse and psychiatric histories. *Journal of Interpersonal Violence*, 12(2), 155–171. <https://doi.org/10.1177/088626097012002001>
- Fundamental Rights Agency (2014). *Violence against women: An EU-wide survey*. <https://fra.europa.eu/en/publication/2014/violence-against-women-eu-wide-survey-main-results-report>
- Hicks, J. M. (2019). *Elucidating unconscious drivers in clandestine sexual practices: Means and methods for ego-syntonic dynamic sexual consent* [praca magisterska, Pacifica Graduate Institute]. Pacifica Graduate Institute ProQuest Dissertations & Theses.
- Jin, Y., Chung, U.-S., Jeong, S.-H., Lee, W. K. (2013). The reliability and validity of the Korean version of the Child Sexual Behavior Inventory. *Psychiatry Investigation*, 10(4), 336–345. <https://doi.org/10.4306/pi.2013.10.4.336>

- Kendall-Tackett, K. A., Williams, L. M., Finkelhor, D. (1993). Impact of sexual abuse on children: A review and synthesis of recent empirical studies. *Psychological Bulletin*, 113(1), 164–180. <https://doi.org/10.1037/0033-2909.113.1.164>
- Krahé, B., Temkin, J., Bieneck, S., Berger, A. (2008). Prospective lawyers' rape stereotypes and schematic decision making about rape cases. *Psychology, Crime & Law*, 14(5), 461–479. <https://doi.org/10.1080/10683160801932380>
- Kristensen, K., Möller, A. (2017). Development of Pict-O-Stat: Violence – focus on experiences of violence and a survey that can be completed by the individuals themselves. *The Journal of Adult Protection*, 19(3), 146–157. <https://doi.org/10.1108/JAP-08-2016-0018>
- Lathan, E. C., Langhinrichsen-Rohling, J., Stefurak, J. T., Duncan, J., Selwyn, C. N. (2022). Trauma-informed sexual assault investigations training: Lessons learned attempting to enhance sexual assault-related understanding and reduce rape myth beliefs and burnout among police officers. *Policing: A Journal of Policy and Practice*, 16(4), 539–559. <https://doi.org/10.1093/police/paab076>
- Lemieux, S. R., Byers, E. S. (2008). The sexual well-being of women who have experienced child sexual abuse. *Psychology of Women Quarterly*, 32(2), 126–144. <https://doi.org/10.1111/j.1471-6402.2008.00418.x>
- Leśniak, B., Leśniak, M. (2018). Subiektywność w opiniach biegłych sądowych – prawne, etyczne i psychologiczne aspekty. *Zeszyty Naukowe Politechniki Śląskiej. Organizacja i Zarządzanie*, 131, 287–297. <https://doi.org/10.29119/1641-3466.2018.131.22>
- Lew-Starowicz, Z., Stusiński, J. (2015a). Etyka w diagnostyce seksuologicznej i orzecznictwie sądowo-seksuologicznym. W: Z. Lew-Starowicz, A. Przyłuska-Fiszer i J. Stusiński (red.), *Normy i kontrowersje etyczne w seksuologii* (s. 237–250). Gdańskie Wydawnictwo Psychologiczne.
- Lew-Starowicz, Z., Stusiński, J. (2015b). Kryteria normy w seksuologii. W: Z. Lew-Starowicz, A. Przyłuska-Fiszer i J. Stusiński (red.), *Normy i kontrowersje etyczne w seksuologii* (s. 13–68). Gdańskie Wydawnictwo Psychologiczne.
- Lin, L.-P., Yen, C.-F., Kuo, F.-Y., Wu, J.-L., Lin, J.-D. (2009). Sexual assault of people with disabilities: Results of a 2002–2007 national report in Taiwan. *Research in Developmental Disabilities*, 30(5), 969–975. <https://doi.org/10.1016/j.ridd.2009.02.001>
- Lisak, D., Gardinier, L., Nicksa, S. C., Cote, A. M. (2010). False allegations of sexual assault: An analysis of ten years of reported cases. *Violence Against Women*, 16(12), 1318–1334. <https://doi.org/10.1177/1077801210387747>
- Łyś, A., Bargiel-Matusiewicz, K., Kuligowski, K., Studzińska, A. (2019). Psychologiczne, społeczne i prawne przesłanki do działań w zakresie profilaktyki przemocy seksualnej wobec dzieci i młodzieży. *Dziecko Krzywdzone. Teoria, Badania, Praktyka*, 18(2), 21–37.

- Martin, S. L., Ray, N., Sotres-Alvarez, D., Kupper, L. L., Moracco, K. E., Dickens, P. A., Scandlin, D., Gizlice, Z. (2006). Physical and sexual assault of women with disabilities. *Violence Against Women*, 12(9), 823–837. <https://doi.org/10.1177/1077801206292672>
- McKee, Z., Mueller-Johnson, K., Strang, H. (2020). Impact of a training programme on police attitudes towards victims of rape: A randomised controlled trial. *Cambridge Journal of Evidence-Based Policing*, 4(1–2), 39–55. <https://doi.org/10.1007/s41887-020-00044-1>
- Mennicke, A., Anderson, D., Oehme, K., Kennedy, S. (2014). Law enforcement officers' perception of rape and rape victims: A multimethod study. *Violence and Victims*, 29(5), 814–827. <https://doi.org/10.1891/0886-6708.VV-D-13-00017>
- Mitra, M., Mouradian, V. E., Fox, M. H., Pratt, C. (2016). Prevalence and characteristics of sexual violence against men with disabilities. *American Journal of Preventive Medicine*, 50(3), 311–317. <https://doi.org/10.1016/j.amepre.2015.07.030>
- Page, A. D. (2010). True colors: Police officers and rape myth acceptance. *Feminist Criminology*, 5(4), 315–334. <https://doi.org/10.1177/1557085110384108>
- Pastwa-Wojciechowska, B., Grzegorzewska, I., Markiewicz, I. (2023). Diagnostyka psychologiczna w opiniowaniu seksuologicznym. W: A. Depko i A. Krasowska (red.), *Seksuologia sądowa II. Opiniowanie w sprawach karnych* (s. 217–293). Wydawnictwo Lekarskie PZWL.
- Punjabi, M., Norman, J., Edwards, L., Muyingo, P. (2021). *Using ACASI to measure gender-based violence in Ugandan primary schools*. RTI Press Publication. <https://doi.org/10.3768/rtipress.2021.rb.0025.2104>
- Randall, M. (2010). Sexual assault law, credibility, and “ideal victims”: Consent, resistance, and victim blaming. *Canadian Journal of Women and the Law*, 22(2), 397–433. <https://doi.org/10.3138/cjwl.22.2.397>
- Rassin, E. (2000). Criteria Based Content Analysis: The less scientific road to truth. *Expert Evidence*, 7(4), 265–278. <https://doi.org/10.1023/A:1016627527082>
- Rode, D., Dukała, K., Kabzińska, J., Zalewska-Łunkiewicz, K. (2020). *Kliniczna psychologia sądowa*. Wydawnictwo Naukowe PWN.
- Ruby, C. L., Brigham, J. C. (1997). The usefulness of the Criteria-Based Content Analysis technique in distinguishing between truthful and fabricated allegations: A critical review. *Psychology, Public Policy, and Law*, 3(4), 705–737. <https://doi.org/10.1037/1076-8971.3.4.705>
- Ruby, C. L., Brigham, J. C. (1998). Can Criteria-Based Content Analysis distinguish between true and false statements of African-American speakers? *Law and Human Behavior*, 22(4), 369–388. <https://doi.org/10.1023/A:1025766825429>
- Smith, S. G., Chen, J., Basile, K. C., Gilbert, L. K., Merrick, M. T., Patel, N., Walling, M., Jain, A. (2017). *The National Intimate Partner and Sexual Vi-*

- olence Survey (NISVS): 2010–2012 state report*. National Center for Injury Prevention and Control, Centers for Disease Control and Prevention. https://stacks.cdc.gov/view/cdc/46305/cdc_46305_DS1.pdf
- Steller, M. (1989). Recent developments in statement analysis. W: J. C. Yuille (red.), *Credibility Assessment* (s. 135–154). Springer. https://doi.org/10.1007/978-94-015-7856-1_8
- Toeplitz-Winiewska, M. (2017). Problemy i dylematy etyczne w pracy psychologa sądowego. W: J. Brzeziński, B. Chyrowicz, Z. Toeplitz i M. Toeplitz-Winiewska (red.), *Etyka zawodu psychologa* (s. 389–404). Wydawnictwo Naukowe PWN.
- van der Elst, E. M., Okuku, H. S., Nakamya, P., Muhaari, A., Davies, A., McClelland, R. S., Price, M. A., Smith, A. D., Graham, S. M., Sanders, E. J. (2009). Is Audio Computer-Assisted Self-Interview (ACASI) useful in risk behaviour assessment of female and male sex workers, Mombasa, Kenya? *PLoS ONE*, 4(5), artykuł e5340. <https://doi.org/10.1371/journal.pone.0005340>
- Venema, R. M. (2018). Police officers' rape myth acceptance: Examining the role of officer characteristics, estimates of false reporting, and social desirability bias. *Violence and Victims*, 33(1), 176–200. <https://doi.org/10.1891/0886-6708.33.1.176>
- Vrij, A. (2014). Verbal lie detection tools: Statement validity analysis, reality monitoring and scientific content analysis. W: P. A. Granhag, A. Vrij i B. Verschuere (red.), *Detecting Deception. Current challenges and cognitive approaches* (s. 3–35). John Wiley & Sons. <https://doi.org/10.1002/9781118510001.ch1>
- Vrij, A., Akehurst, L. (1998/2003). Komunikacja werbalna a wiarygodność: Ocenianie prawdziwości zeznań. W: A. Memon, A. Vrij i R. Bull (red.), *Prawo i psychologia. Wiarygodność zeznań i materiału dowodowego* (s. 15–47). Gdańskie Wydawnictwo Psychologiczne.
- Vrolijk-Bosschaart, T. F., Verlinden, E., Langendam, M. W., De Smet, V., Teeuw, A. H., Brilleslijper-Kater, S. N., Benninga, Marc. A., Lindauer, R. J. L. (2018). The diagnostic utility of the Child Sexual Behavior Inventory for sexual abuse: A systematic review. *Journal of Child Sexual Abuse*, 27(7), 729–751. <https://doi.org/10.1080/10538712.2018.1477215>
- Wojciechowski, B., Gräns, M., Lidén, M. (2018). A true denial or a false confession? Assessing veracity of suspects' statements using MASAM and SVA. *PLoS ONE*, 13(6), artykuł e0198211. <https://doi.org/10.1371/journal.pone.0198211>

ANNA JĘDRASZKIEWICZ  <https://orcid.org/0000-0002-4588-0262>
Wydział Psychologii, Uniwersytet Warszawski

Zastosowanie modeli psychiatrii obliczeniowej do badań nad PTSD

The application of computational psychiatry models in PTSD research

Autor korespondencyjny: Anna Jędraszekiewicz
e-mail: a.jedraszkiewicz@uw.edu.pl

Sposób cytowania: Jędraszekiewicz, A. (2024). Zastosowanie modeli psychiatrii obliczeniowej do badań nad PTSD. W: B. Zawadzki, K. Fronczyk i K. Bargiel-Matusiewicz (red.), *Wybrane problemy współczesnej diagnostyki klinicznej w psychologii* (s. 65–99). Wydawnictwo Liberi Libri. <https://doi.org/10.47943/lib.9788363487676.rozdzial03>

Streszczenie

Rozdział dotyczy psychiatrii obliczeniowej – nowego interdyscyplinarnego podejścia do badania zaburzeń psychicznych opartego na tworzeniu modeli obliczeniowych. Tekst wprowadza podstawowe założenia teoretyczne podejścia, które mimo swojej niejednorodności skupia się na wspólnym celu wyjaśniania konkretnych dysfunkcji za pomocą precyzyjnych metod matematycznych i symulacji komputerowych. Krytykuje jednocześnie nadmierne koncentrowanie się na opracowywaniu klasyfikacji w systemach psychiatrycznych, podkreślając, że lepsze zrozumienie fundamentalnych zjawisk z perspektywy przetwarzania informacji ułatwi formułowanie trafniejszych diagnoz. W rozdziale omówiono dwie ramy teoretyczne wewnątrz psychiatrii obliczeniowej: uczenie ze wzmocnieniem oraz wnioskowanie bayesowskie. Obie stanowią ogólne zbiory założeń teoretycznych określających relacje między terminami matematycznymi i ich odpowiednikami w postaci konstruktywów psychologicznych. Konkretnie zastosowania w badaniach opisano na przykładzie pourazowego zaburzenia stresowego. Przedstawiono korzyści z modeli obliczeniowych i krytyczne uwagi dotyczące ich wykorzystywania. Zaprezentowano stanowiska badaczy, którzy zachęcają m.in. do przywiązywania większej wagi do trafności teoretycznej oraz tworzenia stabilnych między badaniami parametrów modeli. Rozdział kończy się zarysem dalszych kierunków badania pourazowego zaburzenia stresowego w ujęciach obliczeniowych.

Słowa kluczowe: psychiatria obliczeniowa, stresowe zaburzenie pourazowe, bayesowskie modele poznania, uczenie ze wzmocnieniem, zaburzenia psychiatryczne, kryzys w teorii

Abstract

The chapter explores computational psychiatry – a novel interdisciplinary approach to investigating mental disorders through the development of computational models. It introduces the foundational theoretical

assumptions of this approach, which, despite its lack of uniformity, shares a common goal: explaining specific dysfunctions using precise mathematical methods and computer simulations. Computational psychiatry critiques the excessive emphasis on constructing psychiatric classifications while highlighting that a deeper understanding of fundamental phenomena from an information processing perspective can lead to more accurate diagnoses. Two research frameworks within computational psychiatry are presented: Reinforcement Learning and Bayesian Inference. These frameworks serve as general theoretical approaches that specify relationships between mathematical concepts and their psychological equivalents. Specific empirical applications are discussed in the context of Post-Traumatic Stress Disorder. The chapter highlights both the advantages and critical points of computational modeling in psychiatry, including researchers' perspectives on the importance of construct validity and reliable parameters across studies. Additionally, it proposes possible further directions for computationally-oriented Post-Traumatic Stress Disorder research.

Keywords: computational psychiatry, posttraumatic stress disorder, bayesian models of cognition, reinforcement learning, psychiatric disorders, theory crisis

W rozdziale tym opisuję wybrane aspekty psychiatrii obliczeniowej – jednego z nurtów w badaniach nad zaburzeniami psychicznymi. Rozdział nie dotyczy bezpośrednio zagadnień diagnostycznych, jak inne części tego tomu. Skupia się natomiast na ramach teoretycznych i wynikach badań podstawowych, które mogą pomóc lepiej zrozumieć, jak dokładnie wyglądają dysfunkcje w zaburzeniach psychicznych. Lepsze rozumienie konkretnych dysfunkcji może przyczynić się do rozwoju trafniejszych narzędzi diagnostycznych i podejmowania optymalnych decyzji przez specjalistów.

W języku polskim nie ma wiele tekstów wprowadzających w problematykę psychiatrii obliczeniowej, zatem treści tu przedstawione mogą umożliwić zaznajomienie się z podstawowymi założeniami tego stosunkowo nowego

podejścia. Konkretnie modele omówione zostają na przykładzie stresowego zaburzenia pourazowego (*post-traumatic stress disorder* [PTSD]). Odwołuję się do kryteriów diagnostycznych PTSD obecnych w DSM-5, jednak nie opisuję ich szczegółowo ani nie skupiam się na przeglądzie teorii wyjaśniających poszczególne objawy.

Moim celem jest przede wszystkim ukazanie na poziomie dość ogólnym, jakie możliwości daje perspektywa obliczeniowa w zakresie analizy procesów poznawczych (głównie podejmowania decyzji i uczenia się) w kontekście zaburzeń psychicznych (z naciskiem na kontekst objawów PTSD). Prezentuję również argumenty pojawiające się we współczesnych dyskusjach na temat konsekwencji metodologicznych stosowania modeli obliczeniowych.

PSYCHIATRIA OBLICZENIOWA – WSTĘPNA CHARAKTERYSTYKA

Zarys podejścia

Psychiatrię obliczeniową można określić jako szeroki i niejednorodny zbiór podejść wykorzystujących modelowanie obliczeniowe w celu wyjaśniania i badania związków między zachowaniami, procesami poznawczymi i mechanizmami¹ neurobiologicznymi, które identyfikuje się u osób z objawami zaburzeń psychicznych (Adams i in., 2016; Huys i in., 2021; Stephan i Mathys, 2014). Podejścia te opierają się na wynikach badań i elementach teoretycznych z wielu dyscyplin, m.in. psychiatrii, psychologii klinicznej, psychologii poznawczej, informatyki i neuronauki. Przy czym warto zaznaczyć, że termin *psychiatria obliczeniowa* jest mylący, gdyż sugeruje, że dotyczy wyłącznie psychiatrii, podczas gdy w rzeczywistości obejmuje przede wszystkim psychologię kliniczną, którą w interdyscyplinarnych badaniach integruje się z innymi dziedzinami. Zaletą tego terminu jest jednak jego zwięzłość w porównaniu do *obliczeniowej psychologii klinicznej*, co prawdopodobnie przyczynia się do jego szerokiego zastosowania w anglojęzycznej literaturze. Na najogólniejszym poziomie

¹ W wielu tekstach twierdzi się, że psychiatria obliczeniowa skupia się na wyjaśnianiu mechanizmów (nie tylko neurobiologicznych); chciałabym jednak zachować ostrożność i nie posługiwać się tym złożonym pojęciem, tak żeby uniknąć trudności związanych z jego wieloznacznością oraz niepoprawnymi użyciami, na co zwraca się uwagę w tekstach z filozofii umysłu. Informacje wprowadzające do podejść mechanistycznych zainteresowany Czytelnik znajdzie tutaj: Hommel, 2020; Miłkowski, 2016; Povich i Craver, 2017.

podejścia psychiatrii obliczeniowej łączy ujmowanie umysłu jako systemu przetwarzającego informacje i wykonującego obliczenia w interakcji ze środowiskiem. Niekiedy określenie *obliczeniowy* badacze stosują w znaczeniu metod analizy danych eksperymentalnych lub korelacyjnych, nie określając przy tym explicite założeń teoretycznych – o problemach z tego wynikających traktuje część *Komentarze krytyczne* w tym rozdziale. W różnym stopniu wykorzystują również miary neuronaukowe (np. neurofizjologiczne czy neuroanatomiczne). Na przykład w niektórych pracach pomija się aspekt neurobiologiczny; w takich przypadkach badania integrują konkretny matematyczny model, który ma ukazywać przebieg procesów poznawczych i jest zaimplementowany w postaci kodu komputerowego, jedynie z behawioralnym zadaniem eksperymentalnym. Niejednoznaczny jest także stosunek psychiatrii obliczeniowej do tradycyjnych klasyfikacji psychiatrycznych – w części prac badacze odnoszą się wprost do narzędzi klinicznych. Wtedy klasyfikacje zaburzeń nadają pewne granice i pozwalają na porównywanie nieprawidłowości w funkcjonowaniu między grupą osób z diagnozą oraz grupą bez diagnozy czy też porównywanie między osobami o różnym nasileniu objawów. Natomiast jeśli abstrahuje się od klasyfikacji psychiatrycznych, hipotezy formułowane są na podstawie zjawisk wykrytych w badaniach empirycznych, łączonych potencjalnie z cierpieniem psychicznym (takim zjawiskiem może być np. hamowanie poznawcze).

W literaturze można spotkać się z terminem *precision psychiatry*, który ma podkreślać, że celem badań w psychiatrii obliczeniowej jest ułatwienie formułowania precyzyjnych predykcji co do przebiegu leczenia zaburzeń i tworzenia zindywidualizowanych oddziaływań terapeutycznych (Friston i in., 2017). Objawy w kategoriach diagnostycznych traktowane są często jako przydatne w praktyce klinicznej obserwowalne jednostki klasyfikacji (do czego nawiązuję później w podrozdziale *Wartość dodana*), ale jednocześnie sądzi się, że większą uwagę należy skierować na próby identyfikacji procesów i struktur leżących u podstaw tych objawów (Wiecki i in., 2015). Stąd w części prac duże znaczenie przypisuje się analizom danych statystycznych na poziomie indywidualnym. Aby przybliżyć szerszy kontekst takiego podejścia, warto zauważyć, że psychiatria obliczeniowa bazuje na metodach analizy wypracowanych w ogólniejszym obszarze modelowania obliczeniowego procesów poznawczych (*computational cognitive science*; Sun, 2023a). Model obliczeniowy² definiuje się w uproszczeniu

² Warto podkreślić, że tworzenie modeli obliczeniowych nie wyklucza korzystania z teorii i modeli werbalnych, tj. takich, które nie są matematycznie sformalizowane i które opierają

jako „a model that is implemented as a computer program” („model, który jest implementowany w postaci programu komputerowego”³; Fum i in., 2007, s. 135). Zatem w porównaniu z modelami matematycznymi i statystycznymi modele obliczeniowe wykorzystuje się nie tylko do opisywania zachowań, ale przede wszystkim do ich symulacji (odtworzenia). Modele obliczeniowe ukazują, jakie kroki trzeba wykonać lub jakie informacje są przetwarzane i dokładnie w jaki sposób, żeby w trakcie zadania eksperymentalnego uzyskać efekt w formie np. obserwowalnego zachowania. Ten rodzaj zainteresowania modelami czy teoriami opartymi na procesach (*process models*) można skonstruować z podejściami skupionymi na wytworach (*product models*), które znane są również jako teorie czarnej skrzynki (*black box theories*) lub modele wejścia-wyjścia (*input-output theories*; Sun, 2023a); podział ten koresponduje z rozróżnieniem Fodora (1968) na mocną (obejmującą proces) i słabą (obejmującą tylko wejście i wyjście) równoważność symulacji. I faktycznie porównanie umysłu z czarną skrzynką zdaje się trafnie ilustrować teoretyczne oraz metodologiczne założenia w wielu obszarach nauk o poznaniu. Aby nakreślić tę różnicę, można przywołać przykład modelu ABC stosowanego w psychoterapii poznawczo-behawioralnej do wyjaśniania przyczyn utrzymywania się u pacjentów dezadaptacyjnych schematów myślenia. Zgodnie z tym modelem ludzie doświadczają określonych emocji (C – *consequences*) w wyniku poznawczej interpretacji wydarzeń (B – *beliefs*), a nie ze względu na sam fakt, że wystąpiła np. trudna sytuacja (A – *activating events*). Zatem na etapie projektowania badania formułowane są hipotezy na temat przyczynowych i funkcjonalnych związków między zmiennymi (interesuje nas wpływ poznawczej interpretacji wydarzenia na emocje), weryfikuje się je także empirycznie. Jednak nawet w schematach eksperymentalnych, pozwalających na wnioskowanie o relacjach przyczynowo-skutkowych, testujemy jedynie *input-output* model. Nie badamy więc dokładnie, jakie procesy zachodzą pomiędzy danymi wejściowymi (*input*), czyli interpretacjami wydarzeń, a danymi wyjściowymi (*output*), czyli nasileniem emocji. Należy dodać, że tworzenie modeli obliczeniowych stanowi jeden z wielu etapów badawczych; kluczowe jest bowiem wykazanie zgodności (np. za pomocą narzędzi statystycznych) między procesami symulowanymi

się na założeniach wyrażanych w języku naturalnym (np. teorie osobowości w ujęciu cech).

³ Kwestie niesłusznych uproszczeń istoty obliczeniowych podejść do umysłu oraz wprowadzającego w błąd określenia tzw. metafory komputerowej (napotykanego w części tekstów z zakresu psychologii poznawczej) zostają poruszone m.in. w artykule Rescorli (2020).

komputerowo a realnym funkcjonowaniem umysłu. W tym celu zestawia się właściwości modeli obliczeniowych z wynikami różnych metod: testów poznawczych, miar psychofizjologicznych, neuronalnych, behawioralnych, kwestionariuszy do pomiaru różnic indywidualnych itp. Ponadto wyniki badań empirycznych dostarczają wiedzy o tym, jakie konkretnie zjawiska chcemy modelować i które parametry powinniśmy włączyć do modeli. Warto jednak zaznaczyć, że psychiatria obliczeniowa nie ogranicza się do analiz na poziomie indywidualnym, np. na danych grupowych opierają się sposoby walidacji modeli obliczeniowych (Wiecki i in., 2015). Analizy na poziomie grupowym odgrywają także ważną rolę w nurtach psychiatrii obliczeniowej zogniskowanych wokół uczenia maszynowego i *Big Data*, które stosuje się jako metody klasyfikacji zaburzeń psychicznych na podstawie tzw. neuroobliczeniowych fenotypów (Huys i in., 2016); natomiast w tym rozdziale te zagadnienia nie są omawiane.

W tekście skupię się na dwóch grupach założeń teoretycznych (czy inaczej ramach badawczych; *frameworks*) wykorzystywanych do tworzenia modeli matematycznych: uczeniu ze wzmocnieniem (*reinforcement learning* [RL]) i wnioskowaniu bayesowskim (*Bayesian inference* [BI]), które przedstawiane są jako jedne z częściej stosowanych (Stephan i Mathys, 2014). Opiszę je skrótowo (dokładne charakterystyki, uwzględniające wzory i równania matematyczne, można znaleźć m.in. w: Sun, 2023b), aby nakreślić kontekst dla dalszych części. Skoncentruję się na sposobie ich rozumienia w tekstach z zakresu psychiatrii obliczeniowej, jednak warto pamiętać, że mają one także swoje określone znaczenia w matematyce, informatyce i neuronauce.

Uczenie ze wzmacnianiem

Modele RL mają obrazować, jak ludzie przypisują wartości działaniom podejmowanym w reakcji na zmiany w środowisku oraz jak dokonują wyborów, które pozwalają im uzyskać maksymalną nagrodę na koniec zadania (Adams i in., 2016)⁴. Można wyodrębnić trzy elementy wspólne dla różnych modeli RL: stany, działania i nagrody (czy wzmocnienia), wyrażane za pomocą wzorów i równań matematycznych (Seriès, 2020). Stany opisują wszystko to, co dana

⁴ Kierując się literaturą (zob. Seriès, 2020; Sutton i Barto, 2018), odwołuję się do opisu RL opartego na formalizmie procesu decyzyjnego Markowa.

osoba⁵ może wykorzystać do podejmowania działań, czyli np. bodźce w środowisku lub pobudzenie psychofizjologiczne; stany obejmują nie tylko informacje percepcyjne, lecz także te bardziej abstrakcyjne, jak np. symbole oznaczające obiekty w otoczeniu (Sutton i Barto, 2018). Modele RL konstruowane są do wyjaśniania procesów zachodzących m.in. podczas warunkowania klasycznego, warunkowania instrumentalnego czy podejmowania decyzji. Zadania eksperymentalne składają się z sekwencji kroków czasowych (Rangel i in., 2008). W każdym z tych kroków osoba reaguje i oddziałuje na otoczenie – obserwuje, że znajduje się w określonym stanie (jednym ze skończonej liczby możliwych stanów) oraz zależnie od tego stanu podejmuje działanie (jedno z dostępnego zbioru możliwych w tym stanie działań). Stany się zmieniają, a każde działanie wiąże się z różnej wielkości nagrodą. Należy więc w taki sposób przewidywać wartości swoich działań, żeby kumulująca się w całej sekwencji nagroda była jak największa. Na podstawie obserwacji efektów swoich wyborów (metodą prób i błędów lub opierając się na planowaniu) osoba uczy się zależności między stanami a działaniami. Niektóre wybory przynoszą więcej korzyści długoterminowo i to właśnie je należy konsekwentnie powtarzać mimo niepewności w otoczeniu. Wartość danego stanu oblicza się jako sumę natychmiastowej nagrody i uśrednionych wartości przyszłych stanów (Seriès, 2020). Zatem zadaniem osoby jest szacowanie wartości kolejnego stanu – może być tak, że stan nie wiąże się z natychmiastową nagrodą, jednak w dłuższej perspektywie przybliża do maksymalizacji gromadzonej całościowo nagrody.

W psychologii badacze szczególnie interesują się dwoma rodzajami RL, tj. opartym na modelu środowiska (*model-based*) i tym bez modelu (*model-free*), które odpowiednio przyrównuje się do uczenia wymagającego planowania oraz uczenia się nawyków (Huys i in., 2021; Lee i in., 2012). Istnieje bardzo dużo możliwości tworzenia algorytmów opisujących wiele zależności, np. to, jak obliczana jest nagroda, jakie elementy środowiska są brane pod uwagę, jak środowisko jest reprezentowane, jak przebiega proces uczenia. Postać konkretnego modelu zależy od wyborów badacza i rodzaju badanego zjawiska.

Modele RL wykorzystuje się w badaniach nad uzależnieniami oraz innymi zaburzeniami z impulsywnymi i kompulsywnymi zachowaniami w obrazie klinicznym (Huys i in., 2021). Dostarczają wówczas formalnych ram dla ana-

⁵ Używam zamiennie określeń *podmiot (agent)* i *osoba*. Problemy uczenia maszynowego odnoszą się właśnie do podmiotów, niekoniecznie będących systemami biologicznymi, natomiast w przypadku psychiatrii obliczeniowej domyślnie piszę o osobach.

lizowania procesów wskazujących na dysfunkcje układu dopaminergicznego, takich jak przewidywanie nagrody i kontrola impulsów. Niejednokrotnie modelowanie z użyciem RL okazuje się również przydatne w analizach zaburzeń lękowych czy nastroju, w których istotna może być kwestia nieadaptacyjnych sposobów nadawania wartości bodźcom i działaniom (Adams i in., 2016).

Bayesowskie podejścia do poznania

Kolejna grupa założeń teoretycznych bazuje na BI. Twierdzenie Bayesa wyraża matematyczne relacje między prawdopodobieństwem a priori⁶, obserwacjami oraz prawdopodobieństwem a posteriori. W odróżnieniu od częstościowej interpretacji prawdopodobieństwa, ujęcie bayesowskie uwzględnia istnienie wcześniejszej wiedzy czy hipotez na temat zestawu danych, które wykorzystujemy do wnioskowania. Rozpoczynamy więc od ustalenia prawdopodobieństwa hipotez (H) dotyczących danych, czyli prawdopodobieństwa a priori, zapisywanego jako $P(H)$. Następnie obserwujemy faktyczne dane (D) – ich prawdopodobieństwo ma symbol $P(D)$. Możemy wówczas zapytać, jak prawdopodobne są zaobserwowane dane, kiedy porównujemy je z naszymi wcześniejszymi założeniami; interesuje nas zatem prawdopodobieństwo danych pod warunkiem określonego prawdopodobieństwa hipotez – $P(D|H)$. Celem wnioskowania jest dowiedzenie się, do jakiego stopnia należy zaktualizować prawdopodobieństwo a priori w momencie napotkania nowych obserwacji, tak aby uzyskać prawdopodobieństwo a posteriori, czyli prawdopodobieństwo hipotez z uwzględnieniem danych $P(H|D)$. Oblicza się je ze wzoru

$$P(H|D) = \frac{P(H)P(D|H)}{P(D)}.$$

Twierdzenie to stanowi dla części badaczy fundament rozumienia, jak mózg przetwarza informacje (Stephan i Mathys, 2014). Zważywszy na nieodłączność niepewności świata, możemy orientować się poznawczo w środowisku jedynie na podstawie naszych w różnym stopniu prawdopodobnych hipotez i domysłów. Nieustannie interpretujemy rzeczywistość i tworzymy różne predykcje – korzystamy więc z wewnętrznego modelu świata, który ukierunkowuje odbiór bodźców zmysłowych (Piekarski, 2020; Seriès, 2020). Kiedy naszą wcześniejszą wiedzę

⁶ W literaturze stosowane są również terminy *prawdopodobieństwo uprzednie* oraz *prawdopodobieństwo końcowe*.

konfrontujemy z nowymi informacjami (poznawczymi czy percepcyjnymi), to nowe dane powinny zostać odpowiednio uwzględnione w dotychczasowych interpretacjach. Wówczas zaktualizowana wiedza może dać początek kolejnym hipotezom i proces się powtarza. Aktualizacje mają zachodzić zgodnie z regułą Bayesa. Przy czym rozumowanie dokładnie odwzorowujące BI trudno uznać za biologicznie realistyczne, jako że wymagałoby to precyzyjnego śledzenia gęstości prawdopodobieństw różnych możliwych zdarzeń. Według niektórych autorów bardziej wiarygodne może być to, że umysł dokonuje przybliżonych szacowań, aproksymacji reguły Bayesa (Rescorla, 2019).

Podjęcie bayesowskie do procesów poznawczych nie jest jednolite, w jego ramach próbuje się konstruować koncepcje teoretyczne, które umożliwiłyby wyjaśnienie, jak obliczenia związane z BI mogą być realizowane w strukturach neuronalnych (Adams i in., 2016; Seriès, 2020). Nurtem łączącym do pewnego stopnia część tego rodzaju koncepcji jest bayesowskie przetwarzanie predykcyjne w wariancie hierarchicznym (Clark, 2013; Williams, 2023). Jego zwolennicy twierdzą, że umysł działa jak hierarchiczny i wielopoziomowy model generatywny (*generative model*), który na każdym z poziomów hierarchii stara się zrozumieć i przewidzieć, w jaki sposób ukryte mechanizmy środowiska odpowiadają za obserwowane dane oraz po zaobserwowaniu danych sprawdza, czy mogły one faktycznie zostać wygenerowane na podstawie początkowych predykcji (Kiefer i Hohwy, 2018; Williams, 2018, 2020). Kontakt modelu ze środowiskiem zachodzi dwukierunkowo i w odniesieniu do zasady minimalizacji błędu predykcji: model ogólnie konstruuje przewidywania, a oddolnie przetwarza informacje pochodzące od sygnałów sensorycznych; kiedy pojawia się niezgodność między predykcją (wywiedzioną z wcześniejszej wiedzy) a sygnałem sensorycznym, połączeniem „dół-góra” przesyłana jest informacja o błędzie i, zależnie od jego wiarygodności, model modyfikuje predykcje na odpowiednim poziomie hierarchii lub utrzymuje je niezmiennie (Corlett i in., 2020). Wielopoziomowość modelu zakłada się, aby móc wyróżniać poziomy o różnym stopniu abstrakcji, a hierarchiczna struktura modelu generatywnego ma korespondować z hierarchiczną organizacją kory nowej mózgu (Williams, 2023).

W psychiatrii obliczeniowej źródeł zaburzeń psychicznych upatruje się w nieprawidłowościach w zakresie modelu generatywnego, które polegają m.in. na niewłaściwych operacjach wykonywanych na modelu (Adams i in., 2016; Seriès, 2020). Z objawami zaburzeń psychicznych wiązać się mogą przykładowo z jednej strony zbyt sztywne i niepoddające się łatwej modyfikacji predykcje, czyli takie,

które są niewystarczająco wrażliwe na sygnały sensoryczne, a z drugiej strony predykcje niestabilne i chwiejne, które nieadekwatnie do kontekstu zmieniają się pod wpływem bodźców płynących ze środowiska (Friston, 2023; Huys i in., 2021). Przekłada się to na dysfunkcje w obszarze uczenia się w reakcji na nowe doświadczenia. Pojęcia predykcji czy wcześniejszej wiedzy (więc w przybliżeniu rozkładów prawdopodobieństwa a priori) używa się często zamiennie z przekonaniami (*beliefs*), także uzupełnionymi o przymiotnik „aprioryczne”. Badacze próbują zatem niejako połączyć konceptualnie przekonania w ujęciu bayesowskim z psychologicznym rozumieniem nieadaptacyjnych przekonań np. w ujęciu psychoterapii poznawczo-behawioralnej (*cognitive-behavioral psychotherapy* [CBT]), mierzonymi np. za pomocą samoopisowych kwestionariuszy (Moutoussis i in., 2018). Należy jednak zauważyć, że część autorów predykcje (*predictions*) i hipotezy (*hypotheses*) odnosi do niższych poziomów hierarchii, czyli tych bliższych danym sensorycznym, natomiast przekonania czy przekonania aprioryczne (*beliefs/priors/prior beliefs*) rezerwuje dla poziomów wyższych – wtedy rozumie się je jako bardziej uogólnione założenia na temat świata (Litwin i Miłkowski, 2020), choć wciąż rozpatrywane są one w kategoriach rozkładów prawdopodobieństwa. W rozdziale będę stosować terminy obecne w poszczególnych oryginalnych tekstach, więc za autorami napiszę o predykcjach w przypadku *predictions* lub o przekonaniach w przypadku *beliefs*; przy czym warto pamiętać, że wszystkie te pojęcia łączy zazwyczaj to, że pełnią funkcję predykcji formułowanych w modelu z góry na dół, czyli od wyższych do niższych poziomów hierarchii (Sims, 2017).

Wartość dodana

Można zapytać, w jaki sposób podejścia w ramach psychiatrii obliczeniowej przybliżają do lepszego rozumienia natury zaburzeń psychicznych. W poszukiwaniu odpowiedzi warto wyjść od przedstawienia krytyki kierowanej w stronę psychiatrii tradycyjnej, tj. opartej na klasyfikacjach zaburzeń psychicznych DSM (*Diagnostic and Statistical Manual of Mental Disorders*) i ICD (*International Classification of Diseases*; Cuthbert, 2022). Z jednej strony trudno podważać znaczenie teorii klasyfikujących, które pozwalają na systematyczne porządkowanie danych i wyznaczanie granic między odrębnymi zjawiskami (Miłkowski, 2023). Dążą one do zwiększania rzetelności diagnoz, czyli dużej zgodności w rozpoznawaniu konkretnych kategorii zaburzeń przez klinicy-

stów (Cuthbert, 2022). Cel ten nie zawsze jest jednak realizowany – zwraca się uwagę m.in. na to, że zestawom bardzo różnych objawów nadaje się tę samą kategorię diagnostyczną (Stephan i in., 2016). Problem ten uwidacznia się już w samym tytule artykułu z 2013 r. (Galatzer-Levy i Bryant): *636,120 ways to have posttraumatic stress disorder*; faktycznie taka mnogość kombinacji objawów może utrudniać zrozumienie, co stanowi istotę zaburzenia. Również za niepożądane uznaje się sytuacje, w których mimo doświadczania tych samych objawów pacjenci inaczej reagują na leczenie psychofarmakologiczne czy psychoterapeutyczne (Stephan i in., 2016). Tym samym proces decyzyjny co do rodzaju planowanej farmakoterapii staje się złożony i kłopotliwy, a dopasowywanie skutecznych leków redukujących objawy często przebiega metodą prób i błędów (Stephan i in., 2016).

Zwolennicy podejścia obliczeniowego do psychiatrii sugerują, że problemy z klasyfikacją częściowo można by wyeliminować, gdyby większą wagę przywiązywano do jednej z kluczowych funkcji teorii naukowych, jaką jest wyjaśnianie. Podobne wnioski wysuwają badacze zaburzeń psychicznych w ramach przedsięwzięcia *Research Domain Criteria* (RDoC). Choć głównie w swoich metodach opierają się oni na opracowywaniu domenowych kryteriów badawczych (Cuthbert, 2022), ich celem jest także systematyczna identyfikacja wyraźnie zarysowanych zjawisk psychicznych oraz odkrywanie procesów i struktur w sensie przyczynowym odpowiedzialnych za powstawanie zjawisk. Psychiatria obliczeniowa i podejście RDoC zazębiają się pod względem założeń (Kato i in., 2020), natomiast według części autorów (Adams i in., 2016) RDoC kładzie zbyt duży nacisk na wyjaśnienia neurobiologiczne (tzw. biomarkery). W RDoC wśród fundamentalnych jednostek analizy zaburzeń wymienia się geny, molekuły, komórki, obwody neuronalne, fizjologię, a na końcu zachowania, co tworzy duży przeskok conceptualny między fizjologią a zachowaniami. Tymczasem psychiatria obliczeniowa bierze pod uwagę poziom przetwarzania informacji przez wykorzystywanie formalnych modeli symulujących przebieg procesów poznawczych, dzięki czemu zbliża się do podejścia psychologii poznawczej. Ma to znaczenie w kontekście prób integrowania wiedzy na temat zaburzeń z wielu dyscyplin.

Rozwiązania psychiatrii obliczeniowej, którą współtworzą również psychologowie, korespondują z zaleceniami formułowanymi w reakcji na kryzys w teorii opisywany w literaturze psychologicznej z ostatnich lat (Guest i Martin, 2021; Haslbeck i in., 2021; Robinaugh i in., 2021). Podstawową zaletą formalnych modeli ukazujących działanie procesów psychicznych jest ich

precyzja (Browning i in., 2023). Tworzenie symulacji komputerowych może pociągać za sobą konieczność wyrażenia *explicite* różnych nieuświadomionych przez badaczy założeń związanych z teorią i metodami oraz przedstawienia ich za pomocą operacji matematycznych⁷. Dzięki temu osoba opracowująca zagadnienia teoretyczne, leżące u podstaw jej hipotez, może zastanowić się głębiej nad ograniczeniami, które powinny dotyczyć każdej propozycji teoretycznej (Scheel i in., 2021). To częściowo odpowiada na jeden z przejawów wspomnianego kryzysu, czyli problem nadmiaru teorii psychologicznych, których autorzy nie określają przestrzeni przypadków, do których ich teorie czy predykcje się nie odnoszą. Zdarza się więc, że hipotezy w badaniach są na tyle niejasne, że niemal dowolne wyniki z analiz statystycznych zależności okazują się dopasowane, przez co teoria nie wymaga większych modyfikacji.

WSTĘPNE WYNIKI BADAŃ I PROPOZYCJE TEORETYCZNE NA PRZYKŁADZIE PTSD

W ramach psychiatrii obliczeniowej prowadzi się badania dotyczące wielu zaburzeń, głównie są to: spektrum schizofrenii i inne zaburzenia psychotyczne, zaburzenia depresyjne, zaburzenia afektywne dwubiegunowe, zaburzenia neurorozwojowe: spektrum autyzmu i ADHD, zaburzenia pourazowe, zaburzenia lękowe, zaburzenia obsesyjno-kompulsywne i zaburzenie osobowości z pogranicza (dokładne przeglądy można znaleźć tutaj: Heinz, 2017; Moustafa, 2017; Seriès, 2020). Przegląd literatury wskazuje, że PTSD nie należy do kategorii zaburzeń najczęściej analizowanych w omawianym kontekście. Tymczasem analiza PTSD z perspektywy obliczeniowej może okazać się szczególnie ciekawa i przydatna, gdy uwzględni się takie charakterystyki zaburzenia, jak różnorodność objawów, niewielka replikowalność pojedynczych eksperymentów oraz zmiany w klasyfikacjach oparte głównie na analizach czynnikowych, które

⁷ Jest to rodzaj idealizacji – w praktyce badawczej zdarza się, że specyfikacja modelu oraz jego implementacja nie są wyraźnie rozgraniczone (Cooper i Guest, 2014). Specyfikacja modelu powinna dotyczyć konceptualnych rozwiązań i bardziej abstrakcyjnych założeń teoretycznych, i w tym sensie specyfikacja może wiązać się z wieloma implementacjami, o ile są one zgodne z teorią. Natomiast implementacja odnosi się do konkretnego zastosowania i obejmuje m.in. kod komputerowy, zastosowane algorytmy i te techniczne szczegóły, które umożliwiają odтворzenie poszczególnych kroków w modelu w niezależnych badaniach. Niewystarczający opis decyzji, które zostały podjęte w związku zarówno ze specyfikacją, jak i implementacją, może prowadzić do trudności w komunikacji podstawowych idei istotnych dla teorii.

w małym stopniu obejmują rozważania teoretyczne. W tym podrozdziale szczegółowo odnoś się do dwóch tekstów o PTSD: w jednym wykorzystywane jest RL (Homan i in., 2019), a w drugim, bardziej teoretycznym, przeważają bayesowskie podejścia do poznania (Kube i in., 2020).

W klasyfikacji DSM-5 (American Psychiatric Association, 2013) PTSD rozpoznaje się przy uwzględnieniu ośmiu kryteriów, spośród których cztery dotyczą objawów. Aby wydarzenie mogło zostać określone jako traumatyczne, a więc potencjalnie prowadzące do rozwoju PTSD, powinno wiązać się w określonych warunkach z zagrożeniem życia, poważnymi urazami lub przemocą seksualną (kryterium A). Zgodnie z kolejnymi kryteriami B–E pacjent:

- doświadcza intruzji, czyli niechcianych i niekontrolowanych myśli czy wspomnień w innej formie, które powodują, że trauma przeżywana jest w subiektywnym doświadczeniu na nowo,
- próbuje unikać bodźców kojarzących się z wydarzeniem traumatycznym (przy czym unikanie polega nie tylko na fizycznym dystansowaniu się od niektórych miejsc czy ludzi, ale dotyczy również niedopuszczania do siebie określonych myśli czy emocji),
- przejawia odmienne w porównaniu z zachowaniem przed traumą wzorce reagowania emocjonalnego i poznawczego (dotyczy to m.in. zmian w postrzeganiu siebie i świata),
- wykazuje wzmożone pobudzenie i zachowania świadczące o nadmiernej (w porównaniu ze stanami sprzed traumy) reaktywności układu nerwowego.

Kryteria F, G i H odnoszą się do czasu trwania (objawy nie ustępują od co najmniej miesiąca od wydarzenia), znacznie utrudnionego funkcjonowania oraz wykluczenia innego podłoża objawów. Szczegółowo objawy w PTSD według DSM-5 prezentują Zawadzki i in., w tym tomie.

Różnice neuroobliczeniowe w uczeniu

Cel i wstępne informacje o badaniu

Autorzy pierwszej omawianej pracy (Homan i in., 2019) koncentrują się na asocjacyjnym uczeniu się zagrożenia (*associative learning of threat*) u osób z objawami PTSD. Choć sami o tym nie wspominają, to na podstawie m.in. literatury można przypuszczać, że pojęcie *learning of threat* jest pokrewne pojęciu warunkowania strachu (*fear conditioning*). Niektórzy obydwą po-

jęcia stosują zamiennie (jak np. Ojala i Bach, 2020), natomiast Homan i in. tłumaczą jedynie, że w przypadku *learning of threat* warunkowanie dotyczy reakcji obronnych na bodziec awersyjny (*conditioned defensive responses*) mierzonych przez aktywność elektrodermalną (*skin conductance response* [SCR]). Można to zinterpretować jako uogólnioną reakcję wywołaną przetwarzaniem informacji o zagrożeniu (w badaniu zagrażający ma być szok elektryczny); reakcja ta może, ale nie musi, być strachem (jak w sytuacji warunkowania strachu).

Naukowcy zwracają uwagę na to, że choć dysponujemy wynikami ukazującymi trudności związane z uczeniem się u osób z PTSD, takie jak nadmierna generalizacja negatywnych bodźców czy problemy z utrzymaniem efektów wygaszania reakcji, to trudno określić, w jaki sposób poszczególne składowe procesy uczenia się przyczyniają się do rozwoju dysfunkcji. Odpowiedzi należy szukać w modelach obliczeniowych, które pozwalają przyjrzeć się, jak dokładnie dochodzi do powstawania dysfunkcji w zakresie przetwarzania informacji o zagrożeniu. Jako przykład wskaźnika obliczeniowego potencjalnie pomocnego w odkrywaniu nieprawidłowości autorzy podają tempo uczenia się (*learning rate*). Wskaźniki (czy parametry) w psychiatrii obliczeniowej nie są bezpośrednio obserwowalne, więc badacze wnioskują na ich temat na podstawie danych behawioralnych. Homan i in. podkreślają, że połączenie metod obliczeniowych z miarami neuronaukowymi może dostarczyć pełniejszego obrazu konsekwencji psychicznych traumy u osób doświadczających objawów PTSD. Formułują hipotezę, że analiza samych danych behawioralnych nie ujawnia różnic w zachowaniu podczas warunkowania z udziałem zagrażającego bodźca – zarówno weterani wojenni niezdiagnozowani z PTSD, jak i ci z diagnozą zaburzenia nie mają trudności z wykonaniem obydwu etapów zadania: nabywania skojarzenia bodźca z reakcją, a następnie odwracania wyuczonego skojarzenia (*reversal learning*). Tymczasem skonstruowanie modelu obliczeniowego, przy założeniu, że obliczeniom (czy procesom przetwarzania informacji) odpowiada aktywność neuronalna i psychofizjologiczna, pozwoli ukazać odmienne wzorce uczenia się osób z PTSD. W efekcie możliwe stanie się – po pierwsze – przeanalizowanie, jak badani w reakcji na bodźce i szoki elektryczne tworzą przewidywania co do zmieniających się z próby na próbę zależności między bodźcem wizualnym a obecnością lub brakiem szoku oraz – po drugie – jak dostosowują swoje zachowania zgodnie z wielkościami różnic między tym, czego się spodziewają, a tym, jakie zależności między bodźcami i obecnością lub brakiem szoku występują faktycznie w zadaniu.

Poza SCR w badaniu wykorzystywane są techniki fMRI (*functional magnetic resonance imaging*) i MRI (*magnetic resonance imaging*); najwięcej uwagi autorzy poświęcają funkcjom i strukturze ciała migdałowatego, gdyż odgrywa ono rolę w uczeniu się asocjacyjnym oraz kodowaniu wartości (*value encoding*). Znany jest związek podwyższonej aktywności ciała migdałowatego w odpowiedzi na bodźce o negatywnej walencji emocjonalnej z objawami PTSD. Autorzy zakładają, że obliczenia związane z uczeniem się podczas wykonywania zadania odwracania wyuczonych skojarzeń są realizowane w ciele migdałowatym; chcą sprawdzić, jak implementacje obliczeń w mózgu łączą się z objawami PTSD.

W badaniu wzięło udział 54 weteranów wojennych, w tym 24 z diagnozą PTSD. Zadanie eksperymentalne składało się z dwóch etapów: nabywania skojarzenia (*acquisition*) oraz jego odwracania (*reversal*). Na etapie nabywania prezentowano dwa bodźce (A i B) – twarze wyrażające złość. Bodziec A (jedna twarz) skojarzony został z szokiem elektrycznym, natomiast bodziec B z szokiem nie został skojarzony. Na tym etapie pojawiły się 3 rodzaje prób, które badani musieli nauczyć się różnicować: (1) bodziec B w żadnym wypadku nie pociągał za sobą szoku elektrycznego, (2) bodziec A w większości wypadków nie pociągał za sobą szoku, (3) bodziec A pociągał za sobą szok w 1/3 wypadków. W drugiej fazie, która nastąpiła bez zapowiedzi, skojarzenia bodźców zostały zamienione. Po bodźcu A nie występował już w żadnym wypadku szok elektryczny, natomiast bodźcowi B szok towarzyszył teraz w 1/3 wypadków. Przed rozpoczęciem badania uczestnicy zostali poinformowani, że będą oglądać obrazki na ekranie komputera i otrzymywać szoki elektryczne o odpowiednio dostosowanej intensywności.

Specyfika modelowania

Dla zrozumienia całego procesu modelowania istotna jest interpretacja parametrów obliczeniowych, dlatego zrekonstruuję wnioskowanie przeprowadzone w omawianym tekście właśnie w odniesieniu do wybranych przez badaczy parametrów. Autorzy stosują model hybrydowy, który łączy klasyczne modele uczenia się Pearce’a-Halla (P-H) oraz Rescorli-Wagnera (R-W). Jak twierdzą, oba modele wpisują się w obliczeniowe teorie RL. Zgodnie z modelami P-H i R-W uczenie się polega na aktualizowaniu w każdej próbie wartości (*value*) przypisywanej wskazówkom, które informują o tym, czy nastąpi przewidywany rezultat, czy też nie (wskazówkami w badaniu są bodźce wizualne w postaci twarzy). Aktualizowanie zachodzi na podstawie błędu predykcji (*prediction*

error), czyli rozbieżności między przewidywanym a otrzymywanym rezultatem wystąpienia danej wskazówki (w tym badaniu rezultatem jest szok elektryczny).

Jednym z kluczowych parametrów modeli RL jest tempo uczenia się, które przekłada się na to, do jakiego stopnia błędy predykcji zostaną wykorzystane do zaktualizowania wartości kolejnego przewidywanego bodźca (Atlas i in., 2022). W podstawowym modelu R-W parametr tempa uczenia się jest stały, co oznacza, że wpływ błędu predykcji na uczenie się w pełni zależy od samej wielkości tego błędu, a nie np. stopnia pewności co do tego, czy przewidywania uznaje się za wiarygodne (Harris i in., 2022). Autorzy decydują się jednak na model hybrydowy, oparty nie tylko na modelu R-W, ale wprowadzający także elementy modelu P-H. Dokonują takiego wyboru, gdyż P-H umożliwia modelowanie zmiennego tempa uczenia się, które jest bardziej realistyczne psychologicznie ze względu na takie cechy środowiska, jak zmienność i niepewność. Posługują się zatem parametrem *associability*, tj. wielkością opisującą stopień, do którego dana wskazówka doprowadziła w poprzednich próbach do spodziewanego rezultatu, czyli w jakiej mierze była wiarygodna. Proces uczenia się o wartościach wskazówek przyspiesza wtedy, gdy napotykamy wskazówkę niewiarygodną, tzn. niezgodną z naszymi przewidywaniami o braku lub obecności szoku, a więc zaskakującą. Natomiast uczenie się spowalnia, gdy wskazówki są wiarygodne, czyli nie dostarczają nowych informacji o środowisku (Li i in., 2011). Autorzy piszą o zmiennym tempie uczenia się w kategoriach uwagowych i wyjaśniają, że parametr *associability* odpowiada ilości uwagi, jaką osoba ucząca się poświęca wskazówkom. Im bardziej wskazówka powiązana jest z zaskoczeniem, tym więcej uwagi otrzymuje (później w tekście *associability* tłumaczą jako stopień powiązania wskazówki z zaskoczeniem, poza tym w nawiasie zapisują nazwę oryginalną).

Kolejny parametr modelu, waga dla błędu predykcji (*prediction error weight*), jest wielkością szacowaną dla każdej osoby na podstawie SCR i informuje o tym, jak dużo wagi należy nadać błędnym przewidywaniam podczas aktualizowania stopnia powiązań wskazówek z zaskoczeniem (*associability*); inaczej można to rozumieć jako tempo uczenia się dostosowane do stopnia powiązań wskazówek z zaskoczeniem. Ponadto zdaniem autorów wagi błędów predykcji informują o wartościach wskazówek kodowanych w sensie obliczeniowym (zgodnie z modelem hybrydowym) przez ciało migdałowe podczas wykonywania zadania odwracania wyuczonych skojarzeń.

Podsumowując część o parametrach, można zdefiniować asocjacyjne uczenie się o zagrożeniu jako proces tworzenia skojarzeń między wskazówkami

(bodźcami wizualnymi) a rezultatami tych wskazówek (obecnością zagrożenia lub jego brakiem), w którym wyróżnia się trzy oddzielne (choć pozostające w matematycznych relacjach) aspekty: ważenie błędów predykcji, aktualizowanie stopnia powiązania wskazówek z zaskoczeniem czy określania ich wiarygodności oraz aktualizowanie wartości wskazówek.

W celu przedstawienia sposobu modelowania w tym badaniu, opisane zostaną ogólniejsze zasady tworzenia modeli obliczeniowych. Choć techniki i założenia mogą różnić się znacznie między badaniami, pewne zasady pozostają podobne (por. Nassar i Frank, 2016; Wilson i Collins, 2019). Aby zobrazować zależności występujące między danymi w tej pracy (ale również w innych pracach w psychiatrii obliczeniowej), można wyodrębnić dwa częściowo niezależne modele łączone na pewnym etapie w podejściu obliczeniowym (Delgado i in., 2011), mianowicie model uczenia się (*learning model*) i model obserwacji (*observation model*). Model uczenia się określa dynamikę wewnętrznych zmiennych (*internal variables*) modelu, czyli w omawianym badaniu pozwala uchwycić, jak kształtują się procesy uczenia się o wartościach bodźców zapowiadających szoki elektryczne. Model obserwacji ukazuje zaś, jak zmienne wewnętrzne modelu matematycznego przekładają się na dane behawioralne, a więc w przypadku tego badania, jak dynamika różnic w wielkości wag przypisywanych błędów predykcji generuje aktywność psychofizjologiczną.

Model uczenia się według autorów ma charakter deterministyczny, co oznacza, że jeśli określimy początkowy zestaw parametrów i dane wejściowe w modelu, to dane wyjściowe będą zawsze takie same; w opozycji mamy probabilistyczny model obserwacji, który łączy (lub inaczej: odwzorowuje) zmienne wewnętrzne (te w modelu uczenia się) z obserwacjami w taki sposób, że przypisuje możliwym obserwacjom różne stopnie prawdopodobieństwa. Tłumacząc to prościej: badaczy interesuje, jak nieobserwowalne zmienne wewnętrzne (czy parametry) produkują obserwowalne dane z eksperymentu, przy czym do jednego modelu uczenia się można przypisać więcej niż jeden model obserwacji z tego powodu, że poziom realnych obserwacji zniekształcony jest przez szum. Charakter relacji między tymi dwoma modelami może zostać określony jako generatywny, co odnosi się do sposobu, w jaki model uczenia się jest w stanie wygenerować model obserwacji – chodzi zatem o znalezienie jak najlepszego dopasowania między danymi eksperymentalnymi a modelami matematycznymi. Nie można przy tym utożsamić tak ujmowanego modelu generatywnego z modelem generatywnym rozumianym jako poznawczy model świata konstruowany przez umysł m.in. w podejściach bayesowskich (do

drugiego użycia pojęcia będę nawiązywać w części podrozdziału *PTSD z perspektywy przetwarzania predykcyjnego*⁸. Autorzy artykułu jedynie dwa razy wprost stosują określenie *generative*⁹, stąd konieczne wydaje się doprecyzowanie terminów. W nieco zawołowany sposób piszą także o zmiennych wewnętrznych modelu, prawdopodobnie utożsamiając je ze zmiennymi latentnymi.

W przybliżeniu idei stojących za modelowaniem obliczeniowym może pomóc prezentacja w punktach etapów takiego modelowania. Celem nie jest dokładne odtworzenie matematycznych operacji, lecz przyjrzenie się wnioskowaniu badaczy konstruujących modele w psychiatrii obliczeniowej:

1. Identyfikacja poznawczych procesów do modelowania – autorzy omawianego artykułu skupili się na uczeniu się w reakcji na zagrożenie.
2. Wybór ram teoretycznych, które pozwolą opisać zjawisko – określenie, jak przetwarzane są informacje w sensie rodzaju wykonywanych obliczeń. Ten etap wiąże się z ustaleniem początkowych nieobserwowalnych parametrów. W omawianym badaniu są to parametry wykorzystywane w modelach P-H i R-W.
3. Wybór zadania eksperymentalnego – dobór takiego zadania, które pozwoli zbadać konkretne poznawcze zjawisko. W artykule wykorzystano zadanie odwracania wyuczonych skojarzeń.

⁸ Bardzo dobrze o tej różnicy (pomijanej w większości tekstów) piszą Haines i in. (2023); przytaczam fragment w oryginale:

Generative Model — a statistical model of the joint probability distribution $p(y, \theta)$ on observed data (y) and unobserved parameters (θ). Conceptually, generative models can be thought of as probabilistic models that can simulate forward all aspects of both the observed and unobserved data/variables of interest. For example, a trial-level reinforcement learning model with specified parameters can generate/simulate data that resembles the structure of a participant's actual choices in a decision-making task, with latent parameters such as learning rate, representing unobserved variables that influence behavior. Note: this definition is distinct from the neuroscience construct of “generative model” used in the context of “predictive coding”, “active inference”, and “free energy”. (s. 16)

⁹ Raz w badaniu (w części metodologicznej) pojawiają się określenia deterministycznego modelu uczenia się i probabilistycznego modelu obserwacji. Najprawdopodobniej jest to bezpośrednie odwołanie do pojęć, które znajdujemy u Delgado i in. (2012) – tam również *generative* występuje jednokrotnie: *The observation model provides a generative account of how a particular sort of measurement may be impacted by the underlying learning* (s. 18). Mówimy jednak o tym samym procesie generowania danych.

4. Tworzenie symulacji komputerowych – wygenerowanie sztucznych danych na podstawie modeli matematycznych zgodnych z zakładaną teorią opisującą zachowanie. Jest to etap, który poprzedza analizę danych zebranych w badaniu eksperymentalnym. Przeprowadza się go po to, by móc dokładnie prześledzić, jak zmieniają się różne wartości parametrów, a wraz z nimi zachowania systemu (na razie sztucznego, później realnego). Na tym etapie porównuje się również symulacje różnych wariantów modeli obliczeniowych (w tym badaniu są to trzy warianty modelu hybrydowego i klasyczny model R-W).
5. Estymacja – na podstawie symulowanych wariantów modeli dokonuje się estymacji parametrów, których wartości nie są znane i które odgrywają kluczową rolę w wyjaśnianiu różnic w zachowaniu. Estymuje się je np. za pomocą *hierarchical Bayesian estimation* – stosowane są więc techniki statystyczne, które określają wzajemne relacje między parametrami w modelu (w przypadku techniki *hierarchical Bayesian estimation*, wykorzystanej w badaniu Homana i in., prawdopodobieństwo jednego parametru zależy od wartości innego parametru; Kruschke i Vanpaemel, 2015).
6. Rekonstrukcja (*recovery of parameters*) – porównywanie estymowanych wartości parametrów z prawdziwymi wartościami parametrów z symulacji. Ocenia się, do jakiego stopnia estymacje trafnie rekonstruują (czy odzyskują) prawdziwe parametry w symulacjach. Jeśli trafność jest wysoka, modele z wyestymowanymi parametrami można dopasować do danych zebranych w eksperymencie.
7. Walidacja modelu – sprawdzanie, czy model został dobrze dopasowany do realnych danych z eksperymentu. Następnie porównuje się go z innymi modelami, aby wykazać, że wybrany model najlepiej opisuje dane (w badaniu jest to model hybrydowy łączący R-W i P-H). Najczęściej jest to proces cykliczny, w którym testuje się różne wersje modeli i wielokrotnie modyfikuje parametry.

Wyniki badania

W badaniu nie odnotowano różnic w poziomie SCR, które świadczyłyby o trudnościach w odwracaniu wyuczonych skojarzeń u osób z PTSD. Zgodnie z hipotezami większe nasilenie objawów PTSD badane za pomocą wywiadu klinicznego przekłada się na większe przywiązywanie wagi do błędów predykcji. Autorzy, odwołując się do wcześniej zarysowanej konceptualnej korespondencji

między wagami błędów predykcji i wartościami wskazówek zmieniającymi się wraz ze zmianami w aktywności ciała migdałowatego, wykazują, że osoby z diagnozą PTSD przejawiają osłabioną aktywność neuronalną w ciele migdałowatym i mniejszą objętość tej struktury. Nie występują jednak bezpośrednie związki wag błędów predykcji z objętością ciała migdałowatego. Sprawdzono ponadto, czy ciało migdałowe koduje inne parametry uczenia się, takie jak błędy predykcji oraz stopień powiązań wskazówek z zaskoczeniem. Choć w niektórych badaniach te parametry korelują z aktywnością ciała migdałowatego, w tym badaniu taka zależność się nie ujawniła. Wszystkie powyższe wyniki pokazują, że osoby z objawami PTSD bardziej ważą błędy predykcji, a także słabiej kodują wartości w ciele migdałowatym i mają mniejszą objętość tej struktury.

Autorzy proponują interpretację, zgodnie z którą większa wrażliwość na błędne predykcje odpowiada większej ilości uwagi, jaką weterani z PTSD kierują na niewiarygodne wskazówki. Osoby z wyższym nasileniem objawów w porównaniu z osobami o niższym nasileniu w większym stopniu dostosowują swoje przewidywania do niewiarygodnych wskazówek zmieniających się z próby na próbę. Zbieżnie z hipotezą o występowaniu różnic na poziomie obliczeniowym pokazano, że u osób z większym nasileniem objawów PTSD można wyróżnić w procesie uczenia się o zagrażających bodźcach odmienne obliczeniowe wzorce, którym odpowiadają różnice w aktywności ciała migdałowatego.

Badacze odnoszą się również do pewnej niezgodności w wynikach: wążenie błędów predykcji decyduje o kształtowaniu się wartości wskazówek, przy czym objawy PTSD korelują dodatnio z wagami błędów predykcji, zaś ujemnie z kodowaniem wartości w ciele migdałowatym. Przeprowadzono analizy mediacji i sprawdzono związki parametrów uczenia się z aktywnością w innych obszarach mózgu niż ciało migdałowe. Najważniejsze wyniki dotyczą przetwarzania informacji w prążkowiu. Osłabione kodowanie przez prążkowie stopnia powiązania wartości wskazówek z zaskoczeniem okazało się częściowym mediatorem dodatniego związku objawów PTSD i wag błędów predykcji. Według autorów zwiększone wagi i zmniejszone śledzenie wartości mogą działać na zasadzie kompensacji, lecz nie jest jasne, jak należy rozumieć te rozbieżności.

Mając na względzie ograniczenia związane z wnioskowaniem na temat przyczynowości relacji, badacze konkludują, że objawy PTSD mogą rozwijać się wtedy, gdy pojawiają się nieprawidłowości w zakresie neuroobliczeniowego przetwarzania wartości wskazówek w różnych strukturach mózgowych podczas

asocjacyjnego uczenia się o zagrożeniu. Sądzą również, że nieodnotowanie różnic między grupami na poziomie danych behawioralnych oraz ujawnienie różnic w zakresie zmiennych latentnych w modelach obliczeniowych świadczą o przydatności tych drugich; dostarczają one więcej informacji o rzeczywistym funkcjonowaniu poznawczym. Pomiar fizjologiczny zdaniem autorów można uznać za zniekształconą szumem realizację modelu latentnego, co częściowo tłumaczy trudności w replikacji bardziej klasycznych badań nad warunkowaniem w reakcji na zagrożenie w PTSD.

PTSD z perspektywy przetwarzania predykcyjnego

Objawy PTSD próbuje się wyjaśniać również zgodnie z różnymi założeniami teoretycznymi opartymi na BI (Kube i in., 2020; Linson i Friston, 2019; Lyndon i Corlett, 2020; Seriès i in., 2024). O ile inne zaburzenia w tych nurtach bada się empirycznie, to w przypadku PTSD powstały dotychczas głównie szkice i trudno weryfikowalne symulacje komputerowe. Niemniej jest to kierunek rozwijany, a propozycje interpretacji w myśl BI znaleźć można m.in. w tekście autorstwa Kubego i in. (2020). Potrzebę tworzenia nowych wyjaśnień autorzy uzasadniają tym, że wciąż brakuje nam dostatecznej wiedzy, dlaczego tylko część osób z doświadczeniem traumy rozwija objawy spełniające kryteria PTSD w klasyfikacjach diagnostycznych – mechanizmy powstawania zaburzenia są nadal niejasne.

Podstawy teoretyczne

Badacze w swoim tekście wskazują, że rozumienie trudności pacjentów w nurcie hierarchicznego bayesowskiego przetwarzania predykcyjnego zbiega się z rozumieniem przyjmowanym w terapii CBT. Skutki doświadczenia traumatycznego oraz wynikające z niego zaburzenia uwidaczniają się w sposobie formułowania predykcji na temat rzeczywistości, przy czym w założeniach Kubego i in. szczególnie dużą rolę odgrywa przetwarzanie sygnałów percepcyjnych. Badacze zaznaczają, że odwołują się do wnioskowania aktywnego (*active inference*), czyli do podejścia, zgodnie z którym ludzie w celu zminimalizowania niezgodności z dotychczasowymi predykcjami nie tylko poszukują informacji, ale również podejmują (aktywnie) odpowiednie działania.

Wydarzenie traumatyczne zakłóca funkcjonowanie systemu percepcyjnego. Trauma jest doświadczeniem bardzo trudnym i intensywnym, naruszającym

silnie poczucie bezpieczeństwa. Oznacza to, że predykcjom sformułowanym na podstawie sygnałów percepcyjnych odebranych podczas zdarzenia traumatycznego przypisana zostaje bardzo duża precyzja. Precyzję autorzy rozumieją jako pewność co do wiarygodności predykcji lub danych sensorycznych; umysł na drodze ważenia precyzji nadaje wagę błędom predykcji, tj. ustala wpływ błędów predykcji (na różnych poziomach hierarchii) na aktualizowane przekonania (aposterioryczne). W sensie statystycznym waga błędów predykcji dotyczy stosunku precyzji (czyli wariancji odwrotnej) rozkładu prawdopodobieństwa obserwowanych danych względem precyzji rozkładu prawdopodobieństwa przekonań apriorycznych (Litwin i Miłkowski, 2020; Seriès, 2020). Duża precyzja jest równoznaczna z małą wariancją, więc małą niepewnością, natomiast mała precyzja wiąże się z dużą niepewnością. Wynika z tego, że precyzyjniejsze predykcje są bardziej odporne na zmiany pod wpływem nowych danych sensorycznych. W takim przypadku model generatywny nie musi uwzględniać błędów predykcji, a proces uczenia się zachodzi wtedy w mniejszym stopniu (Corlett i in., 2020; Williams, 2020).

Dysfunkcjonalność przetwarzania informacji w PTSD

Kube i in. (2020) opisują, że w subiektywnym doświadczeniu osoby doświadczającej traumy najlepszym modelem wyjaśniającym rzeczywistość w danej chwili może wydawać się przykładowo taki, zgodnie z którym osoba jest teraz w niebezpieczeństwie. Ma to konsekwencje dla dalszego percepcyjnego wnioskowania w sytuacjach, które intensywnością nie przypominają już traumy. Predykcje z wyższych poziomów (czyli te o wysokiej precyzji przypisanej podczas wydarzenia traumatycznego) ze względu na swoją mało zmienną nadmiernie dużą precyzję zaczynają dominować w systemie percepcyjnym nad mało precyzyjnymi predykcjami z niższych poziomów hierarchii modelu (czyli tymi bardziej zależnymi od kontekstu obecnie przetwarzanych sygnałów sensorycznych). Sprawia to, że odbiór trafnych danych sensorycznych ze środowiska jest utrudniony, gdyż system nie uznaje ich za wystarczająco wiarygodne. Autorzy uważają, że właśnie te nadmiernie precyzyjne przekonania aprioryczne, nieaktualizowane podczas kontaktu z osłabionymi sygnałami sensorycznymi, mogą odpowiadać za objawy intruzji, czyli jedną z czterech kategorii objawów w PTSD według klasyfikacji DSM-5 (z wyłączeniem kategorii pierwszej dotyczącej samego wydarzenia traumatycznego). Zatem pewność co do wiarygodności przekonania o ciągłym zagrożeniu (lub inaczej: stosowanie nadmiernie precyzyjnej percepcyjnej hipotezy zakodowanej w momencie

doświadczania traumy) przekłada się na powtórne przeżywanie stresora traumatycznego, mimo że nie jest on obecny w środowisku.

W rozwoju i utrzymywaniu intruzji rolę odgrywa również interpretowanie sygnałów interoceptywnych, a więc płynących z ciała. Kiedy osoba z zaburzeniem znajdzie się w sytuacji postrzeganej nawet nieświadomie jako podobnej do wydarzenia traumatycznego, może doświadczyć silnego pobudzenia psychofizjologicznego, np. pod postacią lęku. Opierając się na wrażeniach sensorycznych, umysł próbuje wnioskować o ich przyczynie, a zgodnie z wcześniejszym doświadczeniem najbardziej prawdopodobna jest hipoteza percepcyjna o obecności poważnego zagrożenia. I w efekcie zgodność z hipotezą dostarcza kolejnego dowodu na słuszność przekonań apriorycznych. Ponadto w wyniku uczenia się osoby przypisują zbyt dużą precyzję przewidywaniom, że wydarzy się coś nieprzewidywalnego, co wiąże się z objawem nieustającej nadmiernej czujności. Sztywne przekonania aprioryczne wpływają na selekcję działań bazujących na unikaniu – trudno wobec tego mieć styczność z błędami predykcji, a co za tym idzie, zmodyfikować predykcje wyższego rzędu.

Kontekst dotychczasowych badań

Badacze przekonują, że ramy hierarchicznego przetwarzania predykcyjnego i aktywnego wnioskowania są zbieżne z dotychczasowymi teoriami wyjaśniającymi PTSD. Cytują m.in. prace o poznawczym procesie nadmiernej akomodacji. Akomodacja jest uznawana za adaptacyjną, gdyż umożliwia integrację nowych informacji z wcześniejszą wiedzą (np. przedtraumatyczne przekonanie o świecie jako miejscu zawsze bezpiecznym można zmienić na potraumatyczne przekonanie o świecie jako miejscu w większości przypadków bezpiecznym). Jednak w nadmiernej akomodacji dochodzi do zbyt dużej generalizacji przyjmowanych nowych treści, czyli w wyniku jednego zagrażającego zdarzenia formułuje się przekonanie o ogólnie groźnym i niebezpiecznym świecie. Powstałe na drodze nadmiernej akomodacji przekonania, tak samo, jak zbyt precyzyjne predykcje z wyższych poziomów hierarchii, oddziałują na przyszłe interpretacje rzeczywistości i wybór podejmowanych działań. Kube i in. (2020) przytaczają wyniki badań empirycznych spójne z ich propozycjami teoretycznymi, zauważając m.in., że nieadekwatnie wysoki lęk osób z objawami PTSD w sytuacji bezpieczeństwa oraz nieadekwatnie niski lęk w sytuacji zagrożenia potwierdzają osłabiony u pacjentów wpływ danych sensorycznych na aprioryczne przekonania.

Autorzy tłumaczą, że przydatność stosowania hierarchicznego przetwarzania predykcyjnego w kontekście PTSD sprowadza się do kilku punktów. Nowością ma być koncentracja na odmiennych właściwościach przekonań dysfunkcyjnych: ważne jest uzupełnianie badań nad treścią przekonań o badania nad relacjami między wielkością precyzji predykcji z wyższych poziomów hierarchii a stopniem aktualizowania przekonań. Kolejne punkty dotyczą jednego z podstawowych celów badaczy w obszarze psychiatrii obliczeniowej, jakim jest konsolidowanie analiz funkcjonowania umysłu z wynikami neurobiologicznymi. Autorzy opisują m.in. wstępne doniesienia nad skutecznością leków dopaminergicznych i zauważają potencjał badań nad rolą dopaminy w procesie przypisywania precyzji oraz wyborze działań. Wykorzystując wyniki neuroobrazowania, proponują również, że u podłoża trudności w aktualizowaniu przekonań opartych na błędach predykcji może leżeć brak odpowiedniego współdziałania wyższych (obszarów przedczołowych) i niższych struktur mózgu (np. prążkowiec brzuszny).

KOMENTARZE KRYTYCZNE

Badania empiryczne w ramach podejść psychiatrii obliczeniowej projektowane są od niedawna, szczególnie jeśli chodzi o problematykę PTSD. Natomiast już na tym etapie można zidentyfikować trudności związane z modelami.

Funkcje modeli obliczeniowych

Choć modele obliczeniowe mają na celu tworzenie precyzyjniejszych teorii, w niektórych wypadkach nie udaje się tego osiągnąć. Przyczyną może być tendencja badaczy do przypisywania twierdzeniom matematycznym nadmiernej wartości w niewłaściwych kontekstach. Dochodzi do sytuacji, w których konstruowanie modeli formalnych, mimo że w części prac występujących jedynie w roli metody symulacji komputerowej, zaczyna zastępować proces opracowywania teorii (Miłkowski i Litwin, 2022).

Rozdzielenia wymagają pytania o modelowanie rzeczywistego przebiegu procesów poznawczych (wtedy zakładamy, że umysł faktycznie wykonuje obliczenia i w konkretny sposób implementuje algorytm) i o heurystyki pomocne w opisie procesów niekoniecznie realistycznych psychologicznie, lecz dobrze

dopasowanych do danych (zatem tak jakby umysł przetwarzał informacje zgodnie z wyidealizowanym modelem, por. teorie *as-if* w: Gigerenzer, 2020). Pisząc wprost o takim rozdzieleniu, badacze są w stanie zredukować ryzyko ekwiwokacji i wysuwania nieprawidłowych wniosków.

Natomiast kiedy modele obliczeniowe sprowadzone zostają do funkcji narzędzi, może pojawić się zarzut, że w dużym stopniu upodabniają się do teorii behawiorystycznych (Jones i Love, 2011), czyli dotyczą jedynie adaptacji do środowiska zewnętrznego, a nie struktury wewnętrznej (poznawczej) badanych osób. Modele generatywne, o których szczególnie często wspominają zwolennicy szeroko rozumianego przetwarzania predykcyjnego, same w sobie nie odzwierciedlają sposobu funkcjonowania umysłu. Z pozoru może wydawać się inaczej, jako że w pracach z zakresu psychologii i psychiatrii niejednokrotnie pojęcie zmiennych latentnych (kluczowe w generowaniu danych w omawianych modelach) utożsamia się błędnie z tzw. *underlying processes*, czyli pewnymi podstawowymi procesami psychicznymi, które przyczyniają się do powstawania bardziej obserwowalnych zachowań (Maraun i Gabriel, 2013). I jeśli uznamy, że modele formalne stanowią tylko narzędzia niedające wglądu do przebiegu procesów poznawczych, to okaże się, że zgodnie z perspektywą behawiorystyczną ograniczamy się do analizy tego, jakie zachowania ludzie przejawiają w reakcji na właściwości środowiska. W przypadku badań omówionych bardziej szczegółowo w tym rozdziale możemy zauważyć, że modele generatywne stosowane są w dwóch znaczeniach: właśnie w znaczeniu narzędzi (w tekście Homana i in., 2019), jak i modeli świata tworzonych przez umysł (w tekście Kubego i in., 2020).

Nieścisłość terminologiczna

RL i BI są pewnymi ogólnymi i szerokimi zbiorami założeń teoretycznych. Oznacza to, że porównywanie między badaniami konkretnych modeli odwołujących się do tych zbiorów i pomijanie specyficznych parametrów może prowadzić do niepoprawnych wniosków. Przykładowo z licznych prac na temat przetwarzania predykcyjnego wyłaniają się trzy sposoby rozumienia, jaka jest jego istota, przy czym autorzy nie zawsze wprost odnotowują różnice w ujęciach. W niektórych warunkach za zgodne z przetwarzaniem predykcyjnym można uznać RL, choć zazwyczaj traktuje się je jako oddzielne grupy. Również w tekście omówionym w podrozdziale *PTSD z perspektywy przetwarzania predykcyjnego* autorzy odnoszą się do aktywnego wnioskowania, natomiast nie

wyjaśniają zależności między tym podejściem i przetwarzaniem predykcyjnym, wymieniają je obok siebie, a niejednokrotnie stosują zamiennie. Tymczasem aktywne wnioskowanie jest algorytmem wykorzystywanym w badaniach, które wpisują się w szeroko ujmowane przetwarzanie predykcyjne, z tym, że algorytm może sprawdzać się w modelach o różnych założeniach i w formie odmiennych wyrażen matematycznych (Miłkowski i Litwin, 2022).

Problemy z synonimicznym używaniem terminów psychologicznych w odniesieniu do zbliżonych zjawisk oraz z brakiem określania relacji znaczeniowych między pojęciami matematycznymi i psychologicznymi można zobrazować, ponownie sięgając do przykładów z tekstu Kubego i in. (2020). W artykule czytamy więc o przekonaniach, oczekiwaniach, predykcjach czy hipotezach. Prawdopodobnie wszystkie oznaczają w przybliżeniu wcześniejszą wiedzę czy reprezentacje umysłowe osoby, która musi swoje bardziej lub mniej utrwalone struktury poznawcze skonfrontować z sygnałami z zewnętrznego środowiska. Wydaje się jednak, że autorzy oczekują *implicite*, że przekonania w niektórych fragmentach tekstu będą interpretowane jako rozkłady prawdopodobieństwa zgodnie z BI, a w innych – jako schematy poznawcze nadające sens rzeczywistości. Można się zastanowić, czy w pewnych przypadkach nie brakuje uzasadnienia dla użycia pojęć matematycznych w znaczeniu konstruktów psychologicznych. W RL złożoność tego rodzaju widoczna jest np. w pojęciu śledzenia (lub kodowania) wartości, które ma być matematycznym odpowiednikiem psychologicznych procesów uwagowych. Poza tym definicje terminów mogą ściśle wiązać się z metodami pomiaru; inaczej bada się przewidywania w kontekście tzw. *expectations violation* w terapii ekspozycji, a jeszcze inaczej predykcje we wnioskowaniu typowo percepcyjnym; mimo to część badaczy nie przypisuje dużego znaczenia tym różnicom i traktuje oba sposoby rozumienia jednakowo dla potwierdzenia spójności proponowanych przez siebie założeń (Kube i Rozenkrantz, 2021).

Warto się zastanowić, do jakiego stopnia nieścisłość terminologiczna wpływa z interdyscyplinarności psychiatrii obliczeniowej, która integruje wiedzę z wielu dyscyplin. Próby spojrzenia na zagadnienia zaburzeń z uwzględnieniem dorobku różnych obszarów mogą przyczyniać się do lepszego ich rozumienia, jednak wskazana byłaby ostrożność w operowaniu terminami wieloznacznymi. Nie precyzując, z której dyscypliny czerpiemy definicję pojęcia, nieintencjonalnie utrudniamy interpretację wyników badawczych (Mäki, 2016). Widać to, gdy weźmiemy pod uwagę trzy znaczenia RL. W uczeniu maszynowym jest to grupa algorytmów, w psychologii – rodzaj uczenia się, zaś neuronauka

odwołuje się do specyficznych struktur mózgowych biorących udział w uczeniu się o motywacyjnej wartości bodźców (Eckstein i in., 2021). Te sposoby rozumienia często pokrywają się ze sobą, tj. algorytmy uczenia maszynowego są wykorzystywane do opisu uczenia się na podstawie informacji zwrotnych o nagrodzie lub karze. Lecz występują również rozbieżności, np. trudno mówić o roli dopaminy (rozumienie neuronaukowe) w kontroli motorycznej (rozumienie psychologiczne). W badaniu Homana i in. (2019) omówionym w podrozdziale *Różnice neuroobliczeniowe w uczeniu* można odnieść wrażenie, że te trzy znaczenia się przenikają, a sens niektórych określeń (jak np. *value tracking* lub *value updating*) nie jest oczywisty do wywnioskowania z kontekstu.

WNIOSKI KOŃCOWE

W tym rozdziale starałam się zarysować główne założenia modeli RL i podejść bayesowskich w psychiatrii obliczeniowej. Przedstawiłam przykłady, jak rozpatrywać objawy PTSD z perspektywy tych stosunkowo nowych kierunków badawczych w nauce o zaburzeniach psychicznych. Przytoczyłam również krytyczne uwagi dotyczące niedoskonałości modeli obliczeniowych.

Kluczowa dla rozwoju psychiatrii opartej na modelach formalnych wydaje się koncentracja na teoriach, np. poprzez systematyczne opracowywanie trafności teoretycznej pojęć. Dwie cechy parametrów w modelach obliczeniowych wymagają szczególnej uwagi (Eckstein i in., 2021, 2022; Karvelis i in., 2023): interpretowalność (czyli zdolność parametrów do uchwycenia odrębnych rzeczywistych procesów poznawczych i neuronalnych) oraz możliwość uogólniania (czyli stopień, w jakim parametry opisują względnie stałe między zadaniami właściwości danej osoby, a nie jej środowiska).

W kontekście PTSD warto zastanowić się nad relacją między modelami obliczeniowymi a metodami kwestionariuszowymi i wywiadami klinicznymi. Zarówno objawy, które próbuje się wyjaśniać za pomocą modeli formalnych łączonych z miarami neuronaukowymi, jak i pozycje w kwestionariuszach badających nasilenie objawów, odnoszą się do treści poznawczych. Można zatem zapytać, czy najwyraźniejszą różnicą nie byłby stopień utrwalenia treści poznawczych (analogicznie do rozróżnienia na cechy i stany). Problem zdaje się złożony w sensie metodologicznym i teoretycznym, podobnie jak inne kwestie rozważane w podejściu obliczeniowym do zaburzeń.

W przyszłych badaniach, wciąż podążając tropem interdyscyplinarności, można oprzeć się na współczesnych tekstach z filozofii umysłu (Smortchkova i in., 2020) i zrezygnować z używania pojęcia *przekonania* w rozumieniu nastawienia wobec sądu (zdania w tzw. języku myśli). Lepszym terminem jest *reprezentacja umysłowa*, która poza tym, że obejmuje formaty quasi-językowe i formaty rozkładów prawdopodobieństwa, zawiera w sobie również reprezentacje ikoniczne. Odnosząc się do formatów reprezentacji, unikamy trudności z utożsamianiem przekonań–rozkładów prawdopodobieństwa i przekonań–zdań pojawiających się w kwestionariuszach. Jaśniejsze stają się wtedy dwa cele: ustalenie, w jaki sposób psychologicznie możliwe jest reprezentowanie rozkładów prawdopodobieństwa oraz rozważenie znaczenia różnych formatów reprezentacji dla trudności w przetwarzaniu poznawczym; ze względu na wizualny charakter intruzji w PTSD szczególnie ciekawe mogłyby okazać się badania nad rolą reprezentacji ikonicznych w uczeniu się o zagrożeniu.

Skupianie się na obliczeniowym podejściu do zaburzeń nie musi oznaczać braku zainteresowania perspektywą subiektywną, zwłaszcza że wśród powodów sięgania po pomoc psychiatryczną czy psychoterapeutyczną wymienia się subiektywnie odczuwane cierpienie. Próby włączenia punktu widzenia osób badanych w kontekście psychiatrii obliczeniowej można znaleźć m.in. w: Rouault i in. (2019) oraz Hoven i in. (2023). Autorzy badają metapoznanie, które rozumieją jako umiejętność oceny własnych działań i podejmowania decyzji podczas zadania poznawczego. Badacze posługują się metodami obliczeniowymi oraz krótkimi skalami ujmującymi stopień subiektywnego przekonania o słuszności decyzji, tak aby określić relacje między wieloma pojedynczymi ocenami decyzji podejmowanych z próby na próbę a zgeneralizowanym stopniem pewności we własne wybory w całym zadaniu. Kolejne prace mogłyby koncentrować się wokół podobnie operacjonalizowanego metapoznania i jego związków z poczuciem własnej skuteczności, jedną ze zmiennych mierzonych samoopisowo w ramach szerszego zbioru *self-beliefs* w PTSD.

Warto również zadawać pytania badawcze o różnice indywidualne, np. cechy temperamentu, które mogą stanowić czynniki ryzyka w rozwoju PTSD (Cyniak-Cieciura i Zawadzki, 2021). Różnice we wzorcach reagowania emocjonalnego, choć często traktowane jako nieodłączna część rozumienia zaburzeń psychicznych, w psychiatrii obliczeniowej zdają się schodzić na drugi plan, przynajmniej w ujęciu bliskim psychologii potocznej. Wyjaśnieniem dla takiego stanu rzeczy może być „chłodna” natura symulacji komputerowych – niemniej

wysiłki w kierunku tworzenia modeli obliczeniowych z uwzględnieniem aspektów emocjonalnych są zauważalne w najnowszej¹⁰ literaturze (Kao i in., 2023).

PODZIĘKOWANIA

Za wartościowe i pomocne komentarze do rozdziału chciałabym podziękować Panu Profesorowi Bogdanowi Zawadzkiemu oraz Panu Profesorowi Marcinowi Miłkowskiemu. Dziękuję również za uwagi i recenzje Panu Doktorowi Krzysztofowi Fronczykowi, Pani Profesor Kamilli Bargiel-Matusiewicz, Panu Profesorowi Włodzimierzowi Strusowi oraz Panu Profesorowi Jerzemu Brzezińskiemu.

BIBLIOGRAFIA

- Adams, R. A., Huys, Q. J., Roiser, J. P. (2016). Computational psychiatry: Towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery, and Psychiatry*, 87(1), 53–63. <https://doi.org/10.1136/jnnp-2015-310737>
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (wyd. 5). <https://doi.org/10.1176/appi.books.9780890425596>
- Atlas, L. Y., Sandman, C. F., Phelps, E. A. (2022). Rating expectations can slow aversive reversal learning. *Psychophysiology*, 59(3), artykuł e13979. <https://doi.org/10.1111/psyp.13979>
- Browning, M., Paulus, M., Huys, Q. J. M. (2023). What is computational psychiatry good for? *Biological Psychiatry*, 93(8), 658–660. <https://doi.org/10.1016/j.biopsych.2022.08.030>
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181–204. <https://doi.org/10.1017/S0140525X12000477>
- Cooper, R. P., Guest, O. (2014). Implementations are not specifications: Specification, replication and experimentation in computational cognitive modeling. *Cognitive Systems Research*, 27, 42–49. <https://doi.org/10.1016/j.cogsys.2013.05.001>

¹⁰ Warto zauważyć, że choć w modelach bayesowskich faktycznie zaczyna się dopiero w ostatnich latach uwzględniać stany emocjonalne, to o podejściu obliczeniowym do emocji i motywacji pisał w latach 80. m.in. Stoman (1987).

- Corlett, P. R., Mohanty, A., MacDonald, A. W. (2020). What we think about when we think about predictive processing. *Journal of Abnormal Psychology*, 129(6), 529–533. <https://doi.org/10.1037/abn0000632>
- Cuthbert, B. N. (2022). Research Domain Criteria (RDoC): Progress and potential. *Current Directions in Psychological Science*, 31(2), 107–114. <https://doi.org/10.1177/09637214211051363>
- Cyniak-Cieciura, M., Zawadzki, B. (2021). The relationship between temperament traits and post-traumatic stress disorder symptoms and its moderators: Meta-analysis and meta-regression. *Trauma, Violence, & Abuse*, 22(4), 702–716. <https://doi.org/10.1177/1524838019876702>
- Delgado, M. R., Phelps, E. A., Robbins, T. W. (2011). *Decision making, affect, and learning*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199600434.001.0001>
- Eckstein, M. K., Master, S. L., Xia, L., Dahl, R. E., Wilbrecht, L., Collins, A. G. (2022). The interpretation of computational model parameters depends on the context. *eLife*, 11, artykuł e75474. <https://doi.org/10.7554/eLife.75474>
- Eckstein, M. K., Wilbrecht, L., Collins, A. G. (2021). What do reinforcement learning models measure? Interpreting model parameters in cognition and neuroscience. *Current Opinion in Behavioral Sciences*, 41, 128–137. <https://doi.org/10.1016/j.cobeha.2021.06.004>
- Fodor, J. A. (1968). *Psychological explanation: An introduction to the philosophy of psychology*. Random House.
- Friston, K. (2023). Computational psychiatry: From synapses to sentience. *Molecular Psychiatry*, 28(1), 256–268. <https://doi.org/10.1038/s41380-022-01743-z>
- Friston, K. J., Redish, A. D., Gordon, J. A. (2017). Computational nosology and precision psychiatry. *Computational Psychiatry*, 1, 2–23. https://doi.org/10.1162/cpsy_a_00001
- Fum, D., Missier, F. D., Stocco, A. (2007). The cognitive modeling of human behavior: Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research*, 8(3), 135–142. <https://doi.org/10.1016/j.cogsys.2007.07.001>
- Galatzer-Levy, I. R., Bryant, R. A. (2013). 636,120 ways to have posttraumatic stress disorder. *Perspectives on Psychological Science*, 8(6), 651–662. <https://doi.org/10.1177/1745691613504115>
- Gigerenzer, G. (2020). How to explain behavior? *Topics in Cognitive Science*, 12(4), 1363–1381. <https://doi.org/10.1111/tops.12480>
- Guest, O., Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, 16(4), 789–802. <https://doi.org/10.1177/1745691620970585>
- Haines, N., Sullivan-Toole, H., Olino, T. (2023). From classical methods to generative models: Tackling the unreliability of neuroscientific measures in mental health

- research. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 8(8), 822–831. <https://doi.org/10.1016/j.bpsc.2023.01.001>
- Harris, D. J., Arthur, T., Vine, S. J., Liu, J., Abd Rahman, H. R., Han, F., Wilson, M. R. (2022). Task-evoked pupillary responses track precision-weighted prediction errors and learning rate during interceptive visuomotor actions. *Scientific Reports*, 12(1), artykuł 22098. <https://doi.org/10.1038/s41598-022-26544-w>
- Haslbeck, J. M. B., Ryan, O., Robinaugh, D. J., Waldorp, L. J., Borsboom, D. (2021). Modeling psychopathology: From data models to formal theories. *Psychological Methods*, 27(6), 930–957. <https://doi.org/10.1037/met0000303>
- Heinz, A. (2017). *A new understanding of mental disorders: Computational models for dimensional psychiatry*. MIT Press.
- Homan, P., Levy, I., Feltham, E., Gordon, C., Hu, J., Li, J., Pietrzak, R. H., Southwick, S., Krystal, J. H., Harpaz-Rotem, I., Schiller, D. (2019). Neural computations of threat in the aftermath of combat trauma. *Nature Neuroscience*, 22(3), 470–476. <https://doi.org/10.1038/s41593-018-0315-x>
- Hommel, B. (2020). Pseudo-mechanistic explanations in psychology and cognitive neuroscience. *Topics in Cognitive Science*, 12(4), 1294–1305. <https://doi.org/10.1111/tops.12448>
- Hoven, M., Rouault, M., Van Holst, R., Luigjes, J. (2023). Differences in metacognitive functioning between obsessive–compulsive disorder patients and highly compulsive individuals from the general population. *Psychological Medicine*, 53(16), 7933–7942. <https://doi.org/10.1017/S003329172300209X>
- Huys, Q. J. M., Browning, M., Paulus, M. P., Frank, M. J. (2021). Advances in the computational understanding of mental illness. *Neuropsychopharmacology*, 46(1), 3–19. <https://doi.org/10.1038/s41386-020-0746-4>
- Huys, Q. J. M., Maia, T. V., Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, 19(3), 404–413. <https://doi.org/10.1038/nn.4238>
- Jones, M., Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(4), 169–188. <https://doi.org/10.1017/S0140525X10003134>
- Kao, C.-H., Feng, G. W., Hur, J. K., Jarvis, H., Rutledge, R. B. (2023). Computational models of subjective feelings in psychiatry. *Neuroscience & Biobehavioral Reviews*, 145, artykuł 105008. <https://doi.org/10.1016/j.neubiorev.2022.105008>
- Karvelis, P., Paulus, M. P., Diaconescu, A. O. (2023). Individual differences in computational psychiatry: A review of current challenges. *Neuroscience & Biobehavioral Reviews*, 148, artykuł 105137. <https://doi.org/10.1016/j.neubiorev.2023.105137>

- Kato, A., Kunisato, Y., Katahira, K., Okimura, T., Yamashita, Y. (2020). Computational Psychiatry Research Map (CPSYMAP): A new database for visualizing research papers. *Frontiers in Psychiatry*, 11, artykuł 578706. <https://doi.org/10.3389/fpsyt.2020.578706>
- Kiefer, A., Hohwy, J. (2018). Content and misrepresentation in hierarchical generative models. *Synthese*, 195(6), 2387–2415. <https://doi.org/10.1007/s11229-017-1435-7>
- Kruschke, J. K., Vanpaemel, W. (2015). Bayesian estimation in hierarchical models. W: J. R. Busemeyer, Z. Wang, J. T. Townsend i A. Eidels (red.), *Oxford handbook of computational and mathematical psychology* (s. 279–299). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199957996.013.13>
- Kube, T., Berg, M., Kleim, B., Herzog, P. (2020). Rethinking post-traumatic stress disorder – A predictive processing perspective. *Neuroscience & Biobehavioral Reviews*, 113, 448–460. <https://doi.org/10.1016/j.neubiorev.2020.04.014>
- Kube, T., Rozenkrantz, L. (2021). When beliefs face reality: An integrative review of belief updating in mental health and illness. *Perspectives on Psychological Science*, 16(2), 247–274. <https://doi.org/10.1177/1745691620931496>
- Lee, D., Seo, H., Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience*, 35(1), 287–308. <https://doi.org/10.1146/annurev-neuro-062111-150512>
- Li, J., Schiller, D., Schoenbaum, G., Phelps, E. A., Daw, N. D. (2011). Differential roles of human striatum and amygdala in associative learning. *Nature Neuroscience*, 14(10), 1250–1252. <https://doi.org/10.1038/nn.2904>
- Linson, A., Friston, K. (2019). Reframing PTSD for computational psychiatry with the active inference framework. *Cognitive Neuropsychiatry*, 24(5), 347–368. <https://doi.org/10.1080/13546805.2019.1665994>
- Litwin, P., Miłkowski, M. (2020). Unification by Fiat: Arrested development of predictive processing. *Cognitive Science*, 44(7), artykuł e12867. <https://doi.org/10.1111/cogs.12867>
- Lyndon, S., Corlett, P. R. (2020). Hallucinations in posttraumatic stress disorder: Insights from predictive coding. *Journal of Abnormal Psychology*, 129(6), 534–543. <https://doi.org/10.1037/abn0000531>
- Mäki, U. (2016). Philosophy of interdisciplinarity. What? Why? How? *European Journal for Philosophy of Science*, 6(3), 327–342. <https://doi.org/10.1007/s13194-016-0162-0>
- Maraun, M. D., Gabriel, S. M. (2013). Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling. *New Ideas in Psychology*, 31(1), 32–42. <https://doi.org/10.1016/j.newideapsych.2011.02.006>
- Miłkowski, M. (2016). A mechanistic account of computational explanation in cognitive science and computational neuroscience. W: V. C. Müller (red.), *Computing*

- and philosophy* (t. 375, s. 191–205). Springer International Publishing. https://doi.org/10.1007/978-3-319-23291-1_13
- Miłkowski, M. (2023). *Cognitive metascience: A new approach to the study of theories*. PsyArXiv. <https://doi.org/10.31234/osf.io/xvyz3>
- Miłkowski, M., Litwin, P. (2022). Testable or bust: Theoretical lessons for predictive processing. *Synthese*, 200(6), artykuł 462. <https://doi.org/10.1007/s11229-022-03891-9>
- Moustafa, A. (red.). (2017). *Computational models of brain and behavior*. Wiley-Blackwell.
- Moutoussis, M., Shahar, N., Hauser, T. U., Dolan, R. J. (2018). Computation in psychotherapy, or how computational psychiatry can aid learning-based psychological therapies. *Computational Psychiatry*, 2, 50–73. https://doi.org/10.1162/CPSY_a_00014
- Nassar, M. R., Frank, M. J. (2016). Taming the beast: Extracting generalizable knowledge from computational models of cognition. *Current Opinion in Behavioral Sciences*, 11, 49–54. <https://doi.org/10.1016/j.cobeha.2016.04.003>
- Ojala, K. E., Bach, D. R. (2020). Measuring learning in human classical threat conditioning: Translational, cognitive and methodological considerations. *Neuroscience & Biobehavioral Reviews*, 114, 96–112. <https://doi.org/10.1016/j.neubiorev.2020.04.019>
- Piekarski, M. (2020). *Mechanizmy predykcyjne i ich normatywność*. Wydawnictwo Liberi Libri. <https://doi.org/10.47943/lib.9788363487447>
- Povich, M., Craver, C. F. (2017). Mechanistic levels, reduction, and emergence. W: S. Glennan i P. M. Illari (red.), *The Routledge handbook of mechanisms and mechanical philosophy* (s. 185–197). Routledge.
- Rangel, A., Camerer, C., Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7), 545–556. <https://doi.org/10.1038/nrn2357>
- Rescorla, M. (2019). A realist perspective on Bayesian cognitive science. W: A. Nes i T. Chan (red.), *Inference and consciousness* (s. 40–73). Routledge.
- Rescorla, M. (2020). *The computational theory of mind*. The Stanford encyclopedia of philosophy. <https://plato.stanford.edu/archives/fall2020/entries/computational-mind>
- Robinaugh, D. J., Haslbeck, J. M. B., Ryan, O., Fried, E. I., Waldorp, L. J. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, 16(4), 725–743. <https://doi.org/10.1177/1745691620974697>
- Rouault, M., Dayan, P., Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, 10, artykuł 1141. <https://doi.org/10.1038/s41467-019-09075-3>
- Scheel, A. M., Tiokhin, L., Isager, P. M., Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*, 16(4), 744–755. <https://doi.org/10.1177/1745691620966795>

- Seriès, P. (red.). (2020). *Computational psychiatry: A primer*. The MIT Press.
- Seriès, P., Veerapa, E., Jardri, R. (2024). Can computational models help elucidate the link between complex trauma and hallucinations? *Schizophrenia Research*, 265, 66–73. <https://doi.org/10.1016/j.schres.2023.05.003>
- Sims, A. (2017). The problems with prediction. W: T. K. Metzinger i W. Wiese (red.), *Philosophy and predictive processing*. MIND Group. <https://doi.org/10.15502/9783958573246>
- Smortchkova, J., Dołęga, K., Schlicht, T. (red.). (2020). *What are mental representations?* Oxford University Press.
- Stephan, K. E., Bach, D. R., Fletcher, P. C., Flint, J., Frank, M. J., Friston, K. J., Heinz, A., Huys, Q. J. M., Owen, M. J., Binder, E. B., Dayan, P., Johnstone, E. C., Meyer-Lindenberg, A., Montague, P. R., Schnyder, U., Wang, X.-J., Breakspear, M. (2016). Charting the landscape of priority problems in psychiatry, part 1: Classification and diagnosis. *The Lancet Psychiatry*, 3(1), 77–83. [https://doi.org/10.1016/S2215-0366\(15\)00361-2](https://doi.org/10.1016/S2215-0366(15)00361-2)
- Stephan, K. E., Mathys, C. (2014). Computational approaches to psychiatry. *Current Opinion in Neurobiology*, 25, 85–92. <https://doi.org/10.1016/j.conb.2013.12.007>
- Stoman, A. (1987). Motives, mechanisms, and emotions. *Cognition and Emotion*, 1(3), 217–233. <https://doi.org/10.1080/02699938708408049>
- Sun, R. (2023a). An overview of computational cognitive sciences. W: R. Sun (red.), *The Cambridge handbook of computational cognitive sciences* (s. 3–26). Cambridge University Press. <https://doi.org/10.1017/9781108755610.003>
- Sun, R. (red.). (2023b). *The Cambridge handbook of computational cognitive sciences* (wyd. 2). Cambridge University Press. <https://doi.org/10.1017/9781108755610>
- Sutton, R. S., Barto, A. G. (2018). *Reinforcement learning: An introduction* (wyd. 2). The MIT Press.
- Wiecki, T. V., Poland, J., Frank, M. J. (2015). Model-based cognitive neuroscience approaches to computational psychiatry: Clustering and classification. *Clinical Psychological Science*, 3(3), 378–399. <https://doi.org/10.1177/2167702614565359>
- Williams, D. (2018). Predictive processing and the representation wars. *Minds and Machines*, 28(1), 141–172. <https://doi.org/10.1007/s11023-017-9441-6>
- Williams, D. (2020). Predictive coding and thought. *Synthese*, 197(4), 1749–1775. <https://doi.org/10.1007/s11229-018-1768-x>
- Williams, D. (2023). Hierarchical minds and the perception/cognition distinction. *Inquiry*, 66(2), 275–297. <https://doi.org/10.1080/0020174X.2019.1610045>
- Wilson, R. C., Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, 8, artykuł e49547. <https://doi.org/10.7554/eLife.49547>

MICHAŁ CHRUSZCZEWSKI  <https://orcid.org/0000-0002-3584-8680>
Wydział Psychologii, Uniwersytet Warszawski

Nuda: opis zjawiska, konotacje kliniczne i pomiar

Boredom: The description of the phenomenon, clinical connotations, and assessment

Autor korespondencyjny: Michał Chruszczewski
e-mail: mikael@psych.uw.edu.pl

Sposób cytowania: Chruszczewski, M. (2024). Nuda: opis zjawiska, konotacje kliniczne i pomiar. W: B. Zawadzki, K. Fronczyk i K. Bargiel-Matusiewicz (red.), *Wybrane problemy współczesnej diagnostyki klinicznej w psychologii* (s. 101–122). Wydawnictwo Liberi Libri. <https://doi.org/10.47943/lib.9788363487676.rozdzial04>

Streszczenie

W artykule charakteryzuje się zjawisko nudy w kategoriach psychologicznych. Przedstawia się różne podejścia do jej definiowania, wskazuje się na związki nudy z zaburzeniami psychicznymi oraz na liczne kwestionariuszowe metody pomiaru nudy. Szerzej omawia się cztery z nich: *Skalę podatności na nudę*, czyli polską adaptację *Boredom Proneness Scale* Farmera i Sundberga autorstwa Flakus, *Wielowymiarową skalę stanu nudy*, czyli polską adaptację *Multidimensional State Boredom Scale* Fahlman i in. w opracowaniu Chruszczewskiego, *Wielowymiarową skalę cechy nudy*, czyli przeróbkę *Multidimensional State Boredom Scale* przeprowadzoną również przez Chruszczewskiego oraz oryginalny *Boredom State Inventory* Baratty i Spence'a. Wymienione metody wyróżniają korzystne parametry psychometryczne.

Słowa kluczowe: nuda, zaburzenia psychiczne, jakość życia, miary kwestionariuszowe

Abstract

The paper refers to boredom characterised within the psychological terms. The different approaches to defining a boredom, its associations with the mental disorders and numerous boredom questionnaires are taken into account. Four questionnaire measures of boredom are discussed in a broader manner, i.e. the Polish adaptation of Farmer and Sundberg's Boredom Proneness Scale by Flakus, Polish adaptation of Fahlman et.al. Multidimensional State Boredom Scale done by Chruszczewski and called *Wielowymiarowa Skala Stanu Nudy*, the revamp of Multidimensional State Boredom Scale focused on the measuring the trait boredom and known as *Wielowymiarowa Skala Cechy Nudy*, done also by Chruszczewski, and original Baratta and Spence's Boredom State Inventory. By the definitely good psychometric properties all the boredom measures mentioned above are distinguished.

Keywords: boredom, mental disorders, quality of life, questionnaire measures

Problemem przez długi czas marginalnie obecnym w badaniach i dociekaniach psychologicznych (por. Chruszczewski, 2020) w kontraście z jego społecznym rozpowszechnieniem i oddziaływaniem na ludzi jest nuda. Jeżeli wyobrazimy sobie przez analogię, że do lat 70. czy 80. XX wieku nikt prawie spośród ludzi nauki nie zajmowałby się stresem, stresorami, radzeniem sobie ze stresem i schorzeniami psychosomatycznymi na tle przewlekłego stresu, a badania tej problematyki ruszyłyby z kopyta dopiero na przełomie stuleci, to otrzymamy w przybliżeniu obraz tego, co w psychologii działo się z nudą.

Trudna do zdefiniowania w kategoriach pozytywnych, to znaczy poprzez określenie posiadanych przez nią i specyficznych dla niej własności, jest nuda, niczym bezpostaciowa bestia, raczej brakiem bądź rozpadem, krachem lub choćby kryzysem. Czego? Przede wszystkim zaangażowania i poczucia sensu (Chruszczewski, 2020). Anektując kolejne sfery ludzkiego doświadczenia i utrwalając się, stając się na koniec składnikiem osobowości, nuda znacznie obniża jakość życia, zadowolenie z niego, zmniejsza możliwości sprawnego działania i prowadzi do (lub przynajmniej jest czynnikiem ryzyka) wielu negatywnych zachowań i problemów zdrowotnych (zob. przegląd: Chruszczewski, 2020). Nuda rozpełza się po społeczeństwach, w szczególności, choć wcale nie tylko, krajów rozwiniętych, dezintegrując je, i manifestuje się jako zjawisko istotne w skali socjologicznej, a nawet historycznej. Od rewolucji przemysłowej poziom ludzkiego znudzenia wzrasta, a dynamika tego wzrostu przyspiesza (Dalle Peze i Salzani, 2009; Healy, 1984). Z tych wszystkich powodów warto zwrócić uwagę na zagadnienie nudy.

PRÓBY OKREŚLENIA ZNACZENIA POJĘCIA NUDA

Na początek garść ustaleń definicyjnych i faktów empirycznych. Amerykańskie Towarzystwo Psychologiczne określa nudę jako „stan [...] wynikający z braku zajmowania się bodźcami w środowisku. Jest on ogólnie uważany za jeden z najmniej pożądanых stanów życia na co dzień i często brany za powód odczuwania przygnębienia. Może być widziany jako przeciwieństwo zaciekania lub zaskoczenia” (VandenBos, 2015, s. 139). Wspomniałem o tym, że trudno nudę ujmować w terminach pozytywnych, widać to i w przytoczonej propozycji definicyjnej, gdyż mowa w niej o deficycie stymulacji. Nuda jest tam stanem, lecz – jak wiele innych zjawisk w psychologii – wykazuje charakterystyczny dualizm: może być też cechą, swoistą podatnością czy wrażliwością.

Osoby o wyższym natężeniu nudy-cechy łatwiej, częściej, mocniej przeżywają stany znudzenia. Psychologowie w ostatnich kilku dekadach podjęli liczne próby zdefiniowania nudy, tak że można kategoryzować je odpowiednio do przyjętej teorii na psychodynamiczne, egzystencjalne, aktywacyjne, afektywne, poznawcze, dotyczące różnic indywidualnych, psychobiologiczne albo odpowiednio do przyjętych uwarunkowań nudy na środowiskowe, podmiotowe (m.in. dotyczące uwagi), funkcjonalno-interakcyjne (gdzie nuda mówi o wartości podejmowanych działań, zachodzi przy tym interakcja czynników podmiotowych ze środowiskowymi). Dodać trzeba jeszcze ujęcia syntetyzujące, jak powyższa definicja.

W podejściu psychodynamicznym traktuje się nudę jako efekt wewnętrznego konfliktu w postaci zablokowania popędu życia (Fenichel, 1951; Greenson, 1953) czy też swoistej obrony przed czekaniem (Philips, 1993). Poczucie braku sensu własnego życia i dominacja apatii to w skrócie obraz nerwicy egzystencjalnej, jak interpretował nudę Maddi (1967), wywodząc ją z alienacji społecznej i autoalienacji. Barbalet (1999) pojmował nudę jako emocję związaną z obawą przed (spodziewanym) brakiem znaczenia czy sensu we własnych czynnościach. Niski poziom aktywacji odpowiada nudzie według Russella (1980; tegoż zdania są Vogel-Walcutt i in., 2012). Natomiast według Fahlman i in. (2013) nudzie właściwa jest labilność na wymiarze pobudzenia: od apatii po poirytowanie. Jednoznacznie przedstawia się nuda ze stanowiska psychologii emocji: jest emocją negatywną, najpewniej łagodną postacią wstrętu (Miller, 1997; Plutchik, 1980) bądź, jeśli trwa dłużej, negatywnym nastrojem z wyraźnym rysem niskiego pobudzenia energetycznego. Leary i in. (1986) sądzą, że nuda jest skutkiem wytężonego skupiania uwagi na mało atrakcyjnym bodźcu. Cheyne i in. (2006) twierdzą, że niezdolność do podtrzymywania uwagi to zasadnicza cecha nudy (podobnie uważają Eastwood i in., 2012). Fisher (1993) pisała w związku z nudą o konflikcie między dekoncentracją a koncentracją. W perspektywie psychologii różnic indywidualnych dostrzega się dualizm przelotnego stanu znudzenia i cechy nudy jako osobowościowo zakorzenionej podatności. Zuckermanowi (1971) oraz Farmerowi i Sundbergowi (1986) zawdzięczamy pionierskie badania nad nudą jako cechą indywidualną. Przypuszcza się też (Merrifield i Danckert, 2014), że w nudzie dochodzi do aktywacji tak zwanej sieci wzbudzeń podstawowych w mózgu, co sprzyja błędzeniu myśli, marzeniom na jawie, inkoherencji, osłabieniu czujności.

Środowiskowe koncepcje definiowania nudy akcentują przede wszystkim monotonię otoczenia (o czym na przestrzeni stu lat pisało wielu autorów,

m.in. de Chenne, 1988 czy Todman, 2003). W podejściach nazwanych powyżej podmiotowymi rozważa się zarówno cechy jednostek (głównie wrażliwość czy podatność na nudę), jak i ich procesy poznawcze (głównie pracę uwagi). Bardziej interesujące, gdyż potencjalnie pełniejsze, zdają się podejścia funkcjonalno-interakcyjne. Zdaniem Geiwitza (1966) nuda wytwarza się w przestrzeni czterech zmiennych, które pozostają ze sobą w związku systemowym: niskiego pobudzenia korowego, subiektywnej powtarzalności (monotonii), poczucia ograniczenia (przymusu), nieprzyjemnego nastroju. Nuda według O'Hanlona (1981) jest wiązką reakcji emocjonalnych (złego samopoczucia), poznawczych (nieuważności) i motywacyjnych (zniechęcenia) na wspólnym podłożu biologicznym (chodzi o obniżenie pobudzenia korowego). Hill i Perkins (1985) twierdzą, że istnieją dwa podstawowe składniki nudy, czyli poznawczy (subiektywnie przeżywana monotonia) oraz afektywny (frustracja). Stan względnie niskiego pobudzenia i niezadowolenia, łączony w umyśle nudzącego się człowieka ze zbyt mało pobudzającą sytuacją to nuda w ujęciu Mikulasa i Vodanovicha (1993). Z kolei Baratta i Spence (2018) niskie natężenie pobudzenia (odczuwanego jako nieprzyjemne), nieuważność oraz niezaangażowanie przyjmują za trzy definicyjne elementy nudy. Znane jest też ewolucjonistyczne podejście do nudy (np. Elpidorou, 2018), zgodnie z którym nuda informuje o stopniu zrealizowania celu oraz reguluje zachowanie (modyfikując cele dla lepszej użyteczności).

Za szczególnie trafny i kompletny, a przy tym bardzo dobrze podbudowany empirycznie, uważam model nudy Fahlman (2008, Fahlman i in., 2013), w którym postuluje się istnienie pięciu elementów – jakościowo swoistych aspektów nudy – tworzących strukturę czynnikową z czynnikiem wyższego rzędu, który odpowiada zjawisku nudy w ogóle. Owa piątka to: niezaangażowanie, słabe pobudzenie (nazywane wcześniej przez autorkę przygnębieniem), silne pobudzenie (nazywane wcześniej przez autorkę rozdrażnieniem), nieuważność i odbiór upływu czasu (czas w nudzie przewleka się niemiłosiernie).

Kliniczny i subkliniczny kontekst nudy

Nuda, należy podkreślić: o ile jest przewlekła, obciąża prognozy kliniczne. Z pewnością jest czynnikiem ryzyka niektórych problemów ze zdrowiem psychicznym; nierzadko jej związek z psychopatologią jest głębszy niż tylko statystyczny. Przeżywanie nudy jest powszechnym doświadczeniem pacjentów

klinik neurologicznych, w szczególności tych dotkniętych urazowymi uszkodzeniami mózgu (Kreutzer i in., 2001). Nuda jest wtedy następstwem patologii organicznej, która upośledza tak zwane funkcje wykonawcze, co właśnie powoduje nudę (Gerritsen i in., 2014). Nudą nasycony jest emocjonalny koloryt wielu zaburzeń depresyjnych i związanych z lękiem (zob. też: Kreutzer i in., 2001; Oddy i in., 1978). Nuda współwystępuje z problemami somatyzacyjnymi i obsesyjno-kompulsyjnymi (Sommers i Vodanovich, 2000). Ciągłe odczucie pustki lub nudy było też jednym z podstawowych kryteriów diagnostycznych osobowości chwiejnej emocjonalnie typu borderline (World Health Organization, 1997). Stwierdza się nudę w przebiegu zespołu nadpobudliwości psychoruchowej z deficytem uwagi (Pironti i in., 2016) oraz w zaburzeniu z napadami objadania się (Stickney i Miltenberger, 1999). Nuda współwystępuje z nałogowym hazardem (Kuley i Jacobs, 1988; Mercer i Eastwood, 2010), a także używaniem i nadużywaniem substancji psychoaktywnych (zob. też: Carlson i in., 2010; Ferguson, 1973; LePera, 2011). W szczególności alkoholicy mają wyższy poziom wrażliwości na nudę niż niealkoholicy (Flory i in., 2011). Warto wiedzieć, że nuda jest jednym z głównych czynników, na podstawie których można przewidywać pojawienie się myśli samobójczych u pacjentów szpitali psychiatrycznych (Ben-Zeev i in., 2012). Należy dodać, że nieco szerszego przeglądu zagadnień związanych z klinicznymi uwarunkowaniami i okolicznościami nudy dokonała w polskiej literaturze Flakus (2016).

Poza tym nuda jest względnie silnie powiązana (współczynniki korelacji zwykle wynoszą od 0,40 do 0,50; niekiedy nawet więcej) z wieloma postaciami negatywnego samopoczucia, które same w sobie niekoniecznie świadczą o zaburzeniach psychicznych, ale często są obecne w przebiegu niektórych z nich. Mam tu na myśli poczucie beznadziejności, depresyjność, lęk, poczucie osamotnienia, apatię, poczucie bezsensu (por. Chruszczewski, 2020). Do tego dodać by należało jeszcze własność z innej kategorii, acz regularnie odnotowywaną u osób trwale znudzonych, czyli niską samokontrolę. Trzeba zauważyć, że nuda przewlekła z reguły prowadzi do nasilenia się jej stanu, więc wzrostu depresyjności, alienacji itd. (krótko mówiąc, obniża się nastrój), jak też do pogorszenia funkcjonowania psychospołecznego. A negatywne skutki stanu nudy są bardziej dojmujące dla osób o znacznej podatności na nudę (w sensie cechy psychicznej). Niestety skutki nudy rzutują na siebie, kreując pętle sprzężeń zwrotnych (por. Chruszczewski, 2020). Nuda może na przykład zarazem bezpośrednio i pośrednio – w tym drugim przypadku poprzez osłabienie uwagi, które obniża poziom wykonania pracy – zmniejszać zadowolenie z pracy

(Cummings i in., 2016; Mael i Jex, 2015; Thackray, 1981). Z kolei zmniejszona satysfakcja z pracy wpływa niekorzystnie na nastrój i tak już wcześniej z racji nudy pogorszony. Narasta stres i ryzyko zaburzeń psychosomatycznych. Zmniejsza się zaangażowanie w pracę, zwiększa się poczucie, że nie ma ona sensu. Stan nudy się jeszcze bardziej pogłębia. Działanie takich sprzężeń jest bardzo niekorzystne i trudne do przerwania, a efektem bywają zaburzenia psychiczne i choroby psychosomatyczne bądź nasilenie objawów, jeśli problemy zdrowotne były obecne przed wytworzeniem się nudy. U ludzi o silnej podatności na nudę i trwającym latami niezadowoleniu z życia trudno zresztą określić, co było pierwsze (nuda czy problemy psychiczne) i kwestia staje się „akademicka”.

ROZPOWSZECHNIENIE NUDY I JEJ POMIAR

A zatem waga zjawiska nudy, w tym potencjalnie bardzo złe następstwa nudy chronicznej, a także jej szerokie rozpowszechnienie (o czym poniżej), uzasadniają potrzebę prowadzenia badań dotyczących nudy. Badań empirycznych nie ma zaś bez mierzenia zjawisk obserwowalnych. W dalszej części tekstu skupię się na metodach pomiaru. Najpierw jednak, zgodnie z zapowiedzią, parę słów o częstości występowania nudy.

Nuda to częste uprzykrzenie życia. Na podstawie dokonanego przeglądu różnych badań dotyczących pracowania i spędzania czasu wolnego (Chruszczewski, 2020) sądzę, że obecnie mniej więcej jedna trzecia ludzi w krajach rozwiniętych doświadcza nadmiernej lub co gorsza przewlekłej nudy. W przebiegu życia okresem szczególnego zagrożenia nudą jest dorastanie i młodość; starcy statystycznie nudzą się najmniej. Istnieje przy tym wyraźna różnica związana z płcią: kobiety nudzą się mniej od mężczyzn (możliwe, że prawidłowość ta dotyczy tylko tak zwanych białych z krajów w miarę rozwiniętych; potrzebne byłyby dalsze badania). W czasie wolnym przewlekłe nudzi się co najmniej jedna osoba na osiem. Częściej nudzimy się jednak tam, gdzie mamy mniej swobody niż w czasie wolnym, to znaczy w pracy lub w szkole: w krajach anglosaskich ponad dwie piąte pracowników mocno nudzi się w pracy, podczas kiedy w szkołach i na uczelniach (nie tylko anglosaskich) nudzi się bez mała połowa uczniów i studentów, przy czym uczniowie nudzą się z reguły jeszcze bardziej niż studenci (odpowiednio około 50% i 40%).

Pomiar nudy ma charakter przede wszystkim kwestionariuszowy i na takim pomiarze oparta jest zdecydowana większość naszej wiedzy empirycznej na

temat zjawiska nudy. Inne metody to głównie obserwacja oraz techniki psychofizjologiczne. Porządkując zagadnienie kwestionariuszy mierzących nudę, warto wskazać na metody, w których ujmuje się ją jako stan oraz te, w których ujmuje się ją jako cechę. Na to, sytuacyjno-dyspozycyjne rozróżnienie, nałożyć warto byłoby drugie, dotyczące globalności albo specyficzności nudy. Dotyczy ono tego, czy nuda jest traktowana jako własność ogólna (niezależna od sytuacji), czy też związana ze specyficznym kontekstem (sferą życia). Otrzymujemy w efekcie wyobrażenie czteropolowej tabelki z kwestionariuszami do pomiaru nudy-stanu w ujęciu globalnym, nudy-stanu w ujęciu specyficznym, nudy-cechy w ujęciu specyficznym i nudy-cechy w ujęciu globalnym. Wszystkie pola są wypełnione konkretnymi metodami, lecz niektóre zawierają ich więcej. Miar specyficznych siłą rzeczy (chodzi wszak o różne środowiskowe konteksty nudy) jest więcej niż ogólnych, zaś najmniej jest miar globalnej nudy-cechy. Podkreślić muszę brak kwestionariuszy szacujących nudę, a przeznaczonych w szczególny sposób do użytku klinicznego.

Wykaz kwestionariuszowych miar nudy

Zawarte poniżej wyliczenie znanych kwestionariuszy, którymi mierzy się nudę, podaję wiernie za swoją książką (Chruszczewski, 2020), zmiany tekstu są nieliczne, głównie stylistyczne, choć wyliczenie zostało zaktualizowane według najnowszego stanu wiedzy (co znaczy, że dodałem wzmianki o dwóch nowych kwestionariuszach).

Dysponujemy dwiema metodami do pomiaru ogólnej nudy-cechy. Są to: *Skala wrażliwości na nudę* (Zuckerman Boredom Susceptibility Scale [ZBS]) w obrębie piątej wersji *Skali poszukiwania doznań* (Sensation Seeking Scale V [SSS-V]) Zuckermana (1979) i nowszy wariant tejże podskali w ramach *Kwestionariusza osobowości Zuckerman–Kuhlman–Aluja* (Zuckerman i Aluja, 2015) oraz *Skala podatności na nudę* (Boredom Proneness Scale [BPS]) Farmera i Sundberga (1986).

Znamy cztery skale, które są ogólnymi miarami nudy-stanu: *Wielowymiarową skalę stanu nudy* (Multidimensional State Boredom Scale [MSBS]) skonstruowaną przez Fahlman i in. (2013), *Miarę stanu nudy* (State Boredom Measure [SBM]) Todmana (2013), *Skalę doświadczenia nudy* (Boredom Experience Scale [BES]) van Tilburga i Igou (2012) oraz *Inwentarz stanu nudy* (Boredom State Inventory [BSI]) Baratty i Spence’a (2018).

Istnieje siedem skal mierzących nudę jako cechę, lecz ujętą w szczególnych kontekstach: *Skala radzenia sobie z nudą* (*Boredom Coping Scale* [BCS]) Hamilton i in. (1984), *Skala nudy wypoczynkowej* (*Leisure Boredom Scale* [LBS]) Iso-Aholi i Weissinger (1990), *Skala nudy czasu wolnego* (*Free Time Boredom Scale* [FTBS]) Ragheba i Meryditha (2001), *Skala nudy seksualnej* (*Sexual Boredom Scale* [SBS]) Watta i Ewing (1996), *Skala bezsensowności, niedostymulowania i nudy* (*Purposelessness, Understimulation and Boredom Scale* [PUB]) Passika i in. (2003), *Skala nudy w bliskim związku* (*Relational Boredom Scale* [RBS]) Harasymchuk i Fehr (2012) i *Szkolna skala podatności na nudę* (*School Boredom Proneness Scale* [SBPS]) autorstwa Carrington (2019).

Wiadomo też o ośmiu kontekstowo specyficznych miarach nudy jako stanu. Zaliczają się do tej kategorii: *Skala nudy w pracy* (*Job Boredom Scale* [JBS]) opracowana przez Grubba (1975), *Skala nudy w pracy Lee* (*Lee's Job Boredom Scale* [LJBS]) autorstwa Lee (1986), *Niderlandzka skala nudy* (*Dutch Boredom Scale* [DUBS]) opracowana przez Reijseger i in. (2013), *Skala radzenia sobie z nudą akademicką* (*Boredom Coping Scale – Academic* [BCS-A]) Nett i in. (2010) oraz podskale nudy w *Kwestionariuszu emocji związanych z osiągnięciami* (*Achievement Emotions Questionnaire – Boredom Scales* [AEQ-BS]) autorstwa Pekruna, Goetza, Frenzel i Perry'ego (2011; zob. też: Pekrun, Goetz, Frenzel, Barchfeld i Perry, 2011), czyli *Skala nudy związanej z lekcjami* (*Class-related Boredom Scale* [CRBS]) oraz *Skala nudy związanej z uczeniem się* (*Learning-related Boredom Scale* [LRBS]; nadto *Skala okoliczności nudy; Precursors to Boredom Scale* [PBS]) Daschmann i in. (2011), skala nudy ze *Skal emocji z epistemicznym odniesieniem* (*Epistemically-related Emotion Scales – Boredom* [EES-B]) Pekruna i in. (2016) oraz *Skala nudy uczelnianej* (*Academic Boredom Scale* [ABS]), której autorami są Acee i in. (2010). Poza tym wszystkim Sharp i in. (2021) ostatnio stworzyli *Narzędzie do pomiaru nudy akademickiej* (*The Academic Boredom Survey Instrument* [ABSI]), czyli kontekstowo specyficzne, które zawiera w sobie zarówno skale cechy, jak i stanu nudy. Istnieje też kilka innych kwestionariuszy, w małym stopniu opracowanych psychometrycznie i znanych z pojedynczych publikacji, czyli raczej niefunkcjonujących w szerszym obiegu naukowym. Pomijam je w wyliczeniu.

Istnieje polska wersja MSBS Fahlman i in. (Chruszczewski, 2017, 2020), jak też BPS Farmera i Sundberga (Flakus, 2018). Nadto jest do użytku *Wielowymiarowa skala cechy nudy*, czyli przeróbka MSBS dokonana w celu mierzenia nudy-cechy (Chruszczewski, 2017, 2020). Podejmowane znacznie wcześniej, jeszcze w latach 80. XX wieku, próby adaptacji *Skali poszukiwania doznań*

Zuckermana, więc i podskali wrażliwości na nudę, nie były udane, jak pisze Strelau (2002) z racji nadmiernej specyficzności kulturowej treści oryginalnego kwestionariusza. Od stosunkowo niedawna jest do dyspozycji polska adaptacja skróconej wersji *Kwestionariusza osobowości Zuckerman–Kuhlman–Aluja*, więc i podskali wrażliwości na nudę, w wykonaniu Ścigały (Aluja i in., 2020).

Przegląd wybranych kwestionariuszy mierzących nudę

Flakus (2018) opracowała tak pełną (czyli dwudziestoosiemopozycyjną), jak i znacznie skróconą (dwunastopozycyjną) polską wersję BPS. W badaniu udział wzięło ponad trzysta osób studiujących lub pracujących. Właściwości psychometryczne kwestionariusza można ogólnie określić jako co najmniej dobre. Problemów nastręczył tylko czynnikowy aspekt trafności teoretycznej (co nie jest w sumie takie rzadkie), w szczególności dla wersji pełnej. Ostatecznie autorka przedłożyła wersję skróconą nad pełną i dalsze badania prowadziła tylko nad tą pierwszą. W przypadku tejże wersji (skróconej) znaleziono sensowne teoretycznie i zasadne metodologicznie rozwiązanie dwuczynnikowe (klasyczne, tj. z czynnikiem stymulacji zewnętrznej i wewnętrznej). Wspomniane wyżej co najmniej dobre parametry psychometryczne to bardzo wysoka rzetelność w sensie zgodności wewnętrznej ($\alpha = 0,92$ dla całości, $\alpha = 0,85$ dla stymulacji zewnętrznej, $\alpha = 0,88$ dla stymulacji wewnętrznej) oraz zadowalające korelacje z wynikami SSS-V Zuckermana w – niezbyt udanej w przywołanej już akapit powyżej ocenie Strelaua – adaptacji Oleszkiewicz-Zsurzs (1985), *Inwentarza depresji Becka* (BDI) w polskiej adaptacji Parnowskiego i Jernajczyka (1977) oraz *Skali satysfakcji z życia* (SWLS) Dienera i in. (1985) w polskiej adaptacji Juczyńskiego (2009). W przypadku podskal SSS-V współczynniki korelacji przedstawiają się jak następuje: około 0,20 z podskalą poszukiwania grozy i przygód, około 0,35 z podskalą poszukiwania przeżyć, nieistotnie statystycznie z podskalą rozhamowania i nieprzeciętnie silne z podskalą wrażliwości na nudę – około 0,60. Stosowne współczynniki korelacji z *Inwentarzem depresji Becka* to około 0,30, a ze *Skalą satysfakcji z życia* około –0,35. Z wyjątkiem zaskakująco silnej zbieżności z Zuckermanowską wrażliwością na nudę jest to układ oczekiwany na gruncie wiedzy teoretycznej i świadczący korzystnie na rzecz trafności polskiej wersji kwestionariusza.

O ile Flakus przyjęła strategię adaptacji wykraczającą miejscami poza translację (ku trawestacji; zob. też: Drwal, 1990), o tyle ja w pracach nad MSBS,

które doprowadziły do powstania *Wielowymiarowej skali stanu nudy* (WSSN), przyjąłem strategię translacji, gdyż specyficzność kulturowa treści zdawała się niska. W obu przypadkach autorzy zlecili kilka niezależnych tłumaczeń na język polski, przeprowadzili tłumaczenie wsteczne (na angielski) i na podstawie tych zabiegów ustalili finalne brzmienie pozycji w języku polskim. Przyjąłem nazwy podskal, takie jak niezaangażowanie, podenerwowanie, nieuważność, przygnębienie, odbiór upływu czasu. Nie zdecydowałem się na nowszą nomenklaturę (z silnym pobudzeniem w miejsce rozdrażnienia i słabym w miejsce przygnębienia), uznając, że dotychczasowa doskonale opisuje symptomy nudy, a nowa jest zbyt abstrakcyjna. Z innych spraw „kosmetycznych” warto zaznaczyć, że punktację pozycji o formacie siedmiokategorialnym z zakresu 1–7 sprowadziłem do 0–6 jako bardziej naturalnego z racji zera; ma to znaczenie przy porównywaniu wyników badań zagranicznych. Badanie niespełna stu studentów anglistyki oryginałem i skalą po polsku dało wyniki niemal całkowicie zbieżne (od ideału dzieliły dwie pozycje z dwudziestu dziewięciu, w których średnie różniły się istotnie zależnie od języka). Badanie trafności teoretycznej w aspekcie czynnikowym (ponad tysiąc stu respondentów reprezentatywnych dla populacji ze względu na wiek, wykształcenie, miejsce zamieszkania) dało wyniki zdecydowanie dobre, choć nie w pełni zadowalające. Najlepiej dopasowane do danych (na tle jednoczynnikowego i czteroczynnikowego) okazały się w konfirmacyjnej analizie czynnikowej metodą ważonych najmniejszych kwadratów dwa modele pięcioczynnikowe (prosty i hierarchiczny), przy czym minimalnie lepiej, doprawdy śladowo¹, niemniej lepiej dopasowany był model prosty, czyli bez czynnika nadrzędnego nudy ogólnej. Ze względów praktycznych (dla wygodniejszej diagnozy) i z racji znacznego skorelowania podskal pozostałem jednak przy czynniku nadrzędnym. Rzetelność skal, bardzo wysoka, nie odbiega od rzetelności oryginału. W szczególności odnotowałem $\alpha = 0,93$ dla niezaangażowania (liczba pozycji: $k = 10$), $\alpha = 0,91$ dla podenerwowania ($k = 5$), $\alpha = 0,87$ dla nieuważności (ta podskala jest najkrótsza, $k = 4$), $\alpha = 0,89$ w przypadku przygnębienia ($k = 5$), $\alpha = 0,91$ w przypadku odbioru upływu czasu ($k = 5$), zaś $\alpha = 0,97$ dla wyniku ogólnego ($k = 28$). W dalszych badaniach (trafnościowych) ustaliłem w próbie niespełna trzystu studentów kierunków humanistycznych i społecznych, że wynik ogólny WSSN łączy się

¹ O czym świadczy χ^2 Satorry-Bentlera równe 1928,36 ($df = 367$) vs 1955,39 ($df = 372$), SRMR równy 0,048 vs 0,049 i przede wszystkim AIC równy 2064,36 vs 2081,39, przy identycznych wskaźnikach NNFI = 0,99 i RMSEA = 0,061.

korelacyjnie – zgodnie z oczekiwaniami – stosunkowo silnie z deklarowaną częstością nudzenia się ($r = 0,50$), podobnie, lecz ujemnie, z satysfakcją z życia ($r = -0,41$), zaś wcale i nieistotnie statystycznie z empatią, co wskazuje na trafność różnicową. Wynik ogólny WSSN korelował z impulsywnością w nikłym stopniu ($r = 0,11$) i na poziomie tzw. tendencji ($p \approx 0,07$), a ze skłonnością do ryzyka, niestety wbrew hipotezie, istotnej statystycznie korelacji nie było. Użyłem *Kwestionariusza impulsywności IVE* Eysencka i Eysenck (2006) w polskiej adaptacji Jaworowskiej (2011) i *Skali satysfakcji z życia* Dienera i in. (1985). W drugim badaniu trafnościowym z udziałem ponad trzystu młodych dorosłych zastosowałem *Kwestionariusz postrzegania czasu* (KPCz) Zimbardo i Boyda (2011) w polskiej adaptacji Stolarskiego (nieopublikowanej), *Inwentarz osobowości NEO-FFI* Costy i McCrae (1989) w polskiej adaptacji Zawadzkiego i in. (1998) oraz *Inwentarz stanu i cechy lęku* (STAI) Spielbergera i in. (1970) w polskiej adaptacji Spielbergera i in. (Wrześniewski i in., 2002). Udało się wykazać silne korelacje dodatnie wyniku ogólnego WSSN ze stanem ($r = 0,61$) oraz cechą ($r = 0,55$) lęku oraz z podstawowymi cechami osobowości (wyjątkiem była otwartość na doświadczenie: związek z nudą statystycznie nieistotny): w przypadku neurotyczności współczynnik korelacji to 0,49, dla ekstrawersji $-0,29$, ugodowości $-0,24$, sumienności $-0,31$. Wykazano też, zgodnie z hipotezą, dodatni związek stanu nudy z dezadaptacyjną perspektywą czasową (dominantą przeszłości negatywnej, niskiego natężenia przeszłości pozytywnej, terażniejszości fatalistycznej i niskiego natężenia przyszłości). Łącznie przemawia to zdecydowanie za trafnością metody.

Wielowymiarowa skala cechy nudy (WSCN) różni się od WSSN instrukcją; wprowadziłem też drobne zmiany językowe w tych nielicznych pozycjach, które ich wymagały z racji na zmieniony przedmiot pomiaru. Eksploracyjna analiza czynnikowa dała rezultaty zgodne z założeniami: wszystkie pozycje miały najwyższe ładunki czynnikowe w czynnikach odpowiadających skalom, do których je zaliczono, z reguły znacznie wyższe od ładunków tych pozycji w pozostałych czynnikach. Zbiór uczestników badania stanowiły dwie połączone próby opisane przy okazji charakteryzowania trafności WSSN, czyli ponad sześćuset młodych dorosłych. Rzetelność skal również była bardzo wysoka, to znaczy wynosiła: $\alpha = 0,90$ dla niezaangażowania ($k = 10$), $\alpha = 0,86$ dla podenerwowania ($k = 5$), $\alpha = 0,82$ dla nieuważności ($k = 4$), $\alpha = 0,88$ dla przygnębienia ($k = 5$), $\alpha = 0,89$ dla odbioru upływu czasu ($k = 5$), a $\alpha = 0,94$ w przypadku wyniku ogólnego ($k = 28$). WSCN w badaniach walidacyjnych skorelowałem z tymi samymi miarami, co WSSN oraz dodatkowo ze *Skalą*

samooceny Rosenberga, otrzymując silną korelację dodatnią łącznego wyniku WSCN z deklarowaną częstością nudzenia się ($r = 0,64$), bardzo słabą z impulsywnością ($r = 0,22$), dość silną ujemną z satysfakcją z życia ($r = -0,52$), nieistotną statystycznie i niemal zerową z empatią. Ze skłonnością do ryzyka, niezgodnie z hipotezą, znaleziono – minimalnej siły, niemniej statystycznie istotną – korelację ujemną ($r = -0,16$). Wynik łączny WSCN korelował umiarkowanie silnie ze stanem lęku ($r = 0,46$), a jeszcze silniej z cechą lęku ($r = 0,72$), co jest teoretycznie zasadne. Analogicznie jak WSSN, WSCN korelowała statystycznie istotnie ze wszystkimi podstawowymi cechami osobowości z wyjątkiem otwartości na doświadczenie; współczynnik korelacji WSCN z neurotycznością to 0,66, z ekstrawersją $-0,30$, z ugodowością $-0,26$, a z sumiennością $-0,43$ (za to ostatnie odpowiada przede wszystkim nieuważność). Cecha nudy wiązała się wyraźniej niż jej stan, choć w ten sam sposób, kiedy mowa o szczegółowych korelatach, z dezadaptacyjną perspektywą czasową. Związek cechy nudy z wynikiem polskiej adaptacji *Skali samooceny Rosenberga* (Dzwonkowska i in., 2008) był ujemny i raczej silny ($r = -0,54$). W ogromnej większości przypadków układ korelacji był zgodny ze szczegółowymi hipotezami, co świadczy o trafności metody. Natomiast związek WSSN z WSCN, zależnie od badania, wahał się w przedziale od około 0,60 do 0,70. Zazwyczaj właśnie tego rzędu są podawane w literaturze korelacje miar stanów z miarami cech.

W relacjonowanych badaniach udało mi się też pokazać szereg socjodemograficznych korelatów stanu nudy mierzonego WSSN. Mowa o stanie, więc odczuciach bezpośrednio z sytuacji badania. Jednak stan koreluje zdecydowanie silnie z cechą, więc z małym ryzykiem; w skrócie można powiedzieć, że najbardziej nudzą się osoby w wieku od 15 do 24 lat, zaś najmniej mające 55 lat lub więcej, nadto że osoby z wykształceniem wyższym magisterskim nudzą się trochę mniej od tych o niższym poziomie wykształcenia, mieszkańcy wsi i małych miast (mających do 20 000 mieszkańców) nudzą się nieco bardziej, acz różnica średnich była niewielka, niż ci mieszkańcy większych miejscowości, a respondenci pozostający w związkach małżeńskich nudzą się mniej niż ci o innym stanie cywilnym. Najbardziej, co nie zaskakuje, nudzili się ludzie bezrobotni, a zaraz po nich uczniowie i studenci. Najmniej – osoby pracujące, emeryci, renciści i osoby na urloпах rodzicielskich bądź wychowawczych. Dzieci okazują się remedium na nudę: istotnie bardziej nudzili się respondenci niemający dzieci w porównaniu z tymi, którzy mieli choćby jedno. Wreszcie stwierdziłem niezbyt silny dodatni związek nudy ze względным ubóstwem. To również jest znaczącym przyczynkiem do trafności WSSN.

Spośród licznych metod zagranicznych, nieprzyswojonych dotąd przez polską psychologię, omówić szerzej można BSI Baratty i Spence'a (2018). Jest on przykładem bardzo krótkiej skali pomiarowej: ma zaledwie 11 pozycji (za to z siedmiokategorialnym formatem odpowiedzi). Krótkość metody działa na niekorzyść rzetelności, lecz bywa zaletą w badaniach przesiewowych lub przeładowanych innymi kwestionariuszami. Racją teoretyczną kwestionariusza (stworzonego przy okazji doktoratu Baratty) autorzy uczynili swoją konkluzję z przeglądu piśmiennictwa, która głosiła istnienie czterech najbardziej podstawowych składników nudy, co do których panuje względnie duża zgodność pomiędzy różnymi autorami. Składnikami tymi są niezaangażowanie, negatywny afekt, niskie pobudzenie i nieuważność. Własnym wkładem koncepcyjnym było połączenie niskiego pobudzenia ze stanem afektu negatywnego; zredukowało to – co prawda w sensie bardziej technicznym niż teoretycznym – liczbę składników nudy do trzech. Walorem metodologicznym BSI, raczej znacznym, jest skonstruowanie go w oparciu o teorię odpowiedzi na pozycje testowe (IRT; *item response theory*) na dwutysięcznej próbie studentów, acz mężczyźni stanowili tylko jedną czwartą próby. Choć teoria ta wywodzi się z lat 40. XX wieku, w odniesieniu do pomiaru nudy BSI jest jedyną znaną mi metodą, która bazuje na IRT.

W konfirmacyjnej analizie czynnikowej wykazano dobre dopasowanie trójczynnikowej struktury nudy. Złożone współczynniki rzetelności konstruktów, oparte na udziałach wariancji standaryzowanych ładunków czynnikowych w wariancjach całkowitych, były bardzo wysokie: około 0,90 (zależnie od próby) dla niezaangażowania i przykrego niskiego pobudzenia oraz ponad 0,90 (we wszystkich próbach) dla nieuważności (Baratta i Spence, 2018). Rzetelność wyniku ogólnego BSI w sensie zgodności wewnętrznej także okazała się bardzo wysoka: $0,90 < \alpha \text{ Cronbacha} < 0,95$; wyjątkowo wysokie – zwłaszcza jak na tak krótkie skale – były też rzetelności podskal czynnikowych: współczynniki $\alpha \text{ Cronbacha}$ zawierały się między 0,92 a 0,97. Współzmiennosc podskal jest znaczna (korelacje od 0,40 do niespełna 0,80), co tłumaczy się obecnością czynnika wyższego rzędu, interpretowanego jako stan nudy.

Trafność BSI zweryfikowano w stopniu zdecydowanie zadowalającym (Baratta i Spence, 2018). Mam tu na myśli związki skal BSI z samoopisowym kwestionariuszem przeznaczonym do pomiaru symptomów zespołu nadpobudliwości psychoruchowej z deficytem uwagi (siły około 0,50–0,60). Niektóre inne korelaty BSI to negatywny afekt, depresyjność, lęk (współczynniki korelacji zazwyczaj średnie lub nawet silne: od 0,35 do około 0,70), afekt pozytywny

(korelacje rzędu $-0,45$), zaangażowanie w pracę (korelacje od około $-0,20$ do około $-0,50$, niekiedy jednak nieistotne statystycznie). Stan nudy mierzony BSI korelował też dość silnie ($r = 0,63$) z afektywnymi i poznawczymi aspektami nudy w pracy, które łącznie szacowała skala LJBS Lee (1986) radykalnie skrócona przez van Hooff i van Hoofta (2014) do pięciu pozycji. Badanie trafności różnicowej również przyniosło rezultaty pomyślne: BSI okazało się raczej niezależne od postaw respondentów wobec polityki rządu względem środowiska naturalnego oraz ich opinii na temat stosunku własnej organizacji (badano osoby pracujące) do dobrostanu konsumentów. Wyjątkiem w tej optymistycznej układance nie stała się bynajmniej trafność zbieżna: korelacje BSI z BPS Farmera i Sundberga (2018; czyli miarą nudy-cechy) znajdowano w przedziale od $0,64$ do $0,82$.

Napisałem kiedyś (Chruszczewski, 2020), że

BSI jest skalą dobrze i celnie mierzącą esencję stanu nudy, natomiast w sferze dyskusji teoretycznych pozostaje fakt, czy Baratta i Spence nie zawężili nadmiernie symptomatyki tego stanu [...]. W tym kontekście chcę przypomnieć swoiste pójście na skróty Baratty i Spence'a w budowaniu podwalin teoretycznych pod ich pojęcie stanu nudy: za najlepsze uznali to, co najczęściej rozważane przez innych. (s. 141)

BSI ma szereg mocnych stron metodologicznych i psychometrycznych, lecz jak mniemam, pełniejsze ujęcie syndromu nudy daje MSBS, który zawiera w sobie właśnie to, co BSI, lecz i coś więcej (silne pobudzenie i odbiór upływu czasu). BSI jawi mi się, na planie teoretycznym, jako mimowolny wyciąg z MSBS.

PERSPEKTYWY DALSZYCH BADAŃ NAD NUDĄ

Ze spojrzenia na piśmiennictwo światowe dotyczące nudy, a ukierunkowanego na stosowane kwestionariusze i liczbę prac, w których ich użyto, wnioskować wolno, że tematami najbardziej interesującymi współczesnych badaczy są nuda jako cecha (na tym polu królują BPS i, mimo wszystko, ZBS), nuda jako stan (MSBS), nuda wypoczynkowa (LBS) i nuda w życiu seksualnym (SBS). Coraz intensywniej bada się też (różnymi metodami) nudę szkolną i uczelnianą, co jest wyjściem naprzeciw problemowi społecznemu znacznej niestety rangi. Nuda w szkołach, nie wyłączając wyższych, nuda czasu wolnego, a nawet nuda odczuwana w sferze życia płciowego wzięte razem wystawiają niewesołe świa-

dectwo dominującemu (ilościowo) sposobowi życia i przeżywania go w ramach współczesnej cywilizacji nie tylko przecież północnoamerykańskiej. Trudno było powstrzymać się od tej końcowej uwagi, aczkolwiek cele napisania tego artykułu były ogólnopedagogiczne, autor nie zamierzał wchodzić w rozważania kulturoznawcze ani też krytykować kultury współczesnej. O tym piszą inni (Anderson, 2004; Dalle Peze i Salzani, 2009; Haller i in., 2013; Healy, 1984; Lehner, 2015; Leslie, 2009; Lombardo, 2017; Postman, 1995; Winter, 2012).

BIBLIOGRAFIA

- Acee, T. W., Kim, H., Kim, H. J., Kim, J.-I., Chu, H.-N. R., Kim, M., Cho, Y., Wicker, F. W., Bierer, L., Chung, W.-H., Daniel, S. R., Games, I. A., Lin, Y.-J., Peshva, A. V., Riekenberg, J. J. (2010). Academic boredom in under- and over-challenging situations. *Contemporary Educational Psychology*, 35(1), 17–27. <https://doi.org/10.1016/j.cedpsych.2009.08.002>
- Aluja, A., Rossier, J., Oumar, B., García, L. F., Bellaj, T., Ostendorf, F., Ruch, W., Wang, W., Kövi, Z., Ścigała, D., Čekrljia, Đ., Stivers, A. W., Di Blas, L., Valdivia, M., Ben Jemaa, S., Atitsogbe, K. A., Hansenne, M., Glicksohn, J. (2020). Multicultural validation of the Zuckerman–Kuhlman–Aluja Personality Questionnaire Shortened Form (ZKA-PQ/SF) across 18 countries. *Assessment*, 27(4), 728–748. <https://doi.org/10.1177/1073191119831770>
- Anderson, B. (2004). Time-stilled space-slowed: How boredom matters. *Geoforum*, 35(6), 739–754. <https://doi.org/10.1016/j.geoforum.2004.02.005>
- Baratta, P. L., Spence, J. R. (2018). Capturing the noonday demon: Development and validation of the State Boredom Inventory. *European Journal of Work and Organizational Psychology*, 27(4), 477–492. <https://doi.org/10.1080/1359432X.2018.1481830>
- Barbalet, J. M. (1999). Boredom and social meaning. *British Journal of Sociology*, 50(4), 631–646. <https://doi.org/10.1111/j.1468-4446.1999.00631.x>
- Ben-Zeev, D., Young, M. A., Depp, C. A. (2012). Real-time predictors of suicidal ideation: Mobile assessment of hospitalized depressed patients. *Psychiatry Research*, 197(1–2), 55–59. <https://doi.org/10.1016/j.psychres.2011.11.025>
- Carlson, S. R., Johnson, S. C., Jacobs, P. C. (2010). Disinhibited characteristics and binge drinking among university student drinkers. *Addictive Behaviors*, 35(3), 242–251. <https://doi.org/10.1016/j.addbeh.2009.10.020>
- Carrington, T. (2019). *Development of a school boredom proneness scale for children*. JMU Scholarly Commons. <https://commons.lib.jmu.edu/edspec201019/151>

- Cheyne, J. A., Carriere, J. S. A., Smilek, D. (2006). Absent-mindedness: Lapses of conscious awareness and everyday cognitive failures. *Consciousness and Cognition*, 15(3), 578–592. <https://doi.org/10.1016/j.concog.2005.11.009>
- Chruszczewski, M. H. (2017). Nuda w Polsce: adaptacja kanadyjskiej skali do pomiaru nudy. *Ruch Pedagogiczny*, 87(1), 5–24.
- Chruszczewski, M. H. (2020). *Nuda. Ujęcie psychologiczne i humanistyczne*. Wydawnictwa Uniwersytetu Warszawskiego. <https://doi.org/10.31338/uw.9788323546412>
- Costa, P. T., Jr, McCrae, R. R. (1989). *NEO-PI/FFI manual supplement*. Psychological Assessment Resources.
- Cummings, M. L., Gao, F., Thornburg, K. M. (2016). Boredom in the workplace: A new look at an old problem. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58(2), 279–300. <https://doi.org/10.1177/0018720815609503>
- Dalle Peze, B., Salzani, C. (2009). Introduction. The delicate monster: Modernity and boredom. W: B. Dalle Peze i C. Salzani (red.), *Critical studies: t. 31. Essays on boredom and modernity* (s. 5–33). Brill. https://doi.org/10.1163/9789042032125_002
- Daschmann, E. C., Goetz, T., Stupnisky, R. H. (2011). Testing the predictors of boredom at school: Development and validation of the precursors to boredom scales. *British Journal of Educational Psychology*, 81(3), 421–440. <https://doi.org/10.1348/000709910X526038>
- de Chenne, T. K. (1988). Boredom as a clinical issue. *Psychotherapy: Theory, Research, Practice, Training*, 25(1), 71–81. <https://doi.org/10.1037/h0085325>
- Diener, E., Emmons, R. A., Larsen, R. J., Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, 49(1), 71–75.
- Drwal, R. Ł. (1990). Problemy kulturowej adaptacji kwestionariuszy osobowości. W: A. Ciechanowicz (red.), *Kulturowa adaptacja testów* (s. 115–138). Polskie Towarzystwo Psychologiczne.
- Dzwonkowska, I., Lachowicz-Tabaczek, K., Łaguna, M. (2008). *SES. Samoocena i jej pomiar. Polska adaptacja skali SES M. Rosenberga. Podręcznik*. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Eastwood, J. D., Frischen, A., Fenske, M. J., Smilek, D. (2012). The unengaged mind: Defining boredom in terms of attention. *Perspectives on Psychological Science*, 7(5), 482–495. <https://doi.org/10.1177/1745691612456044>
- Elpidorou, A. (2018). The good of boredom. *Philosophical Psychology*, 31(3), 323–351. <https://doi.org/10.1080/09515089.2017.1346240>
- Eysenck, H. J., Eysenck, S. B. G. (2006). *Manual of the Eysenck Personality Scales (EPS Adult) comprising EPQ-Revised (EPQ-R), EPQ-R Short Scale, Impulsiveness (IVE) Questionnaire*. Hodder & Stoughton.

- Fahlman, S. A. (2008). *Development and validation of the Multidimensional State Boredom Scale* [niepublikowana rozprawa doktorska]. Faculty of Graduate Studies, York University.
- Fahlman, S. A., Mercer-Lynn, K. B., Flora, D. B., Eastwood, J. D. (2013). Development and validation of the Multidimensional State Boredom Scale. *Assessment*, 20(1), 68–85. <https://doi.org/10.1177/1073191111421303>
- Farmer, R., Sundberg, N. D. (1986). Boredom proneness – the development and correlates of a new scale. *Journal of Personality Assessment*, 50(1), 4–17. https://doi.org/10.1207/s15327752jpa5001_2
- Fenichel, O. (1951). On the psychology of boredom. W: D. Rapaport (red.), *Organization and pathology of thought: Selected sources* (s. 349–361). Columbia University Press. <https://doi.org/10.1037/10584-018>
- Ferguson, D. (1973). A study of occupational stress and health. *Ergonomics*, 16(5), 649–664. <https://doi.org/10.1080/00140137308924555>
- Fisher, C. D. (1993). Boredom at work: A neglected concept. *Human Relations*, 46(3), 395–417. <https://doi.org/10.1177/001872679304600305>
- Flakus, M. (2016). Kliniczne aspekty podatności na nudę. *Psychiatria po Dyplomie*, 13(5), 19–25.
- Flakus, M. (2018). Wstępna charakterystyka psychometryczna polskiej adaptacji Skali podatności na nudę (BPS). *Polskie Forum Psychologiczne*, 23(4), 783–802.
- Flory, J. D., Pytte, C. L., Hurd, Y., Ferrell, R. E., Manuck, S. B. (2011). Alcohol dependence, disinhibited behavior and variation in the prodynorphin gene. *Biological Psychology*, 88(1), 51–56. <https://doi.org/10.1016/j.biopsycho.2011.06.007>
- Geiwitz, P. J. (1966). Structure of boredom. *Journal of Personality and Social Psychology*, 3(5), 592–600. <https://doi.org/10.1037/h0023202>
- Gerritsen, C. J., Toplak, M. E., Sciaraffa, J., Eastwood, J. (2014). I can't get no satisfaction: Potential causes of boredom. *Consciousness and Cognition: An International Journal*, 27, 27–41. <https://doi.org/10.1016/j.concog.2013.10.001>
- Greenson, R. R. (1953). On boredom. *Journal of the American Psychoanalytic Association*, 1(1), 7–21. <https://doi.org/10.1177/000306515300100102>
- Grubb, E. A. (1975). Assembly line boredom and individual differences in recreation participation. *Journal of Leisure Research*, 7(4), 256–269. <https://doi.org/10.1080/00222216.1975.11970242>
- Haller, M., Hadler, M., Kaup, G. (2013). Leisure time in modern societies: A new source of boredom and stress? *Social Indicators Research*, 111(2), 403–434. <https://doi.org/10.1007/s11205-012-0023-y>
- Hamilton, J. A., Haier, R. J., Buchsbaum, M. S. (1984). Intrinsic enjoyment and boredom coping scales: Validation with personality, evoked potential, and atten-

- tion measures. *Personality and Individual Differences*, 5(2), 183–193. [https://doi.org/10.1016/0191-8869\(84\)90050-3](https://doi.org/10.1016/0191-8869(84)90050-3)
- Harasymchuk, C., Fehr, B. (2012). Development of a prototype-based measure of relational boredom. *Personal Relationships*, 19(1), 162–181. <https://doi.org/10.1111/j.1475-6811.2011.01346.x>
- Healy, S. D. (1984). *Boredom, self, and culture*. Fairleigh Dickinson University Press.
- Hill, A. B., Perkins, R. E. (1985). Towards a model of boredom. *British Journal of Psychology*, 76(2), 235–240. <https://doi.org/10.1111/j.2044-8295.1985.tb01947.x>
- Iso-Ahola, S. E., Weissinger, E. (1990). Perceptions of boredom in leisure: Conceptualization reliability, and validity of the Leisure Boredom Scale. *Journal of Leisure Research*, 22(1), 1–17. <https://doi.org/10.1080/00222216.1990.11969811>
- Jaworowska, A. (2011). IVE. Kwestionariusz impulsywności. Impulsywność, skłonność do ryzyka, empatia. Polska normalizacja. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Juczyński, Z. (2009). NPPZ. Narzędzia pomiaru w promocji i psychologii zdrowia. Podręcznik. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Kreutzer, J. S., Seel, R. T., Gourley, E. (2001). The prevalence and symptom rates of depression after traumatic brain injury: A comprehensive examination. *Brain Injury*, 15(7), 563–576. <https://doi.org/10.1080/02699050010009108>
- Kuley, N. B., Jacobs, D. F. (1988). The relationship between dissociative-like experiences and sensation seeking among social and problem gamblers. *Journal of Gambling Behavior*, 4(3), 197–207. <https://doi.org/10.1007/BF01018332>
- Leary, M. R., Rogers, P. A., Canfield, R. W., Coe, C. (1986). Boredom in interpersonal encounters: Antecedents and social implications. *Journal of Personality and Social Psychology*, 51(5), 968–975. <https://doi.org/10.1037/0022-3514.51.5.968>
- Lee, T. W. (1986). Toward the development and validation of a measure of job boredom. *Manhattan College Journal of Business*, 15(1), 22–28.
- Lehner, N. (2015). Is boredom inevitable? *Antae*, 2(3), 145–157.
- LePera, N. (2011). Relationships between boredom proneness, mindfulness, anxiety, depression, and substance use. *New School Psychology Bulletin*, 8(2), 15–25. <https://doi.org/10.1037/e741452011-003>
- Leslie, I. I. (2009). From idleness to boredom: On the historical development of modern boredom. W: B. Dalle Peze i C. Salzani (red.), *Critical studies: t. 31. Essays on boredom and modernity* (s. 35–59). Brill. https://doi.org/10.1163/9789042032125_003
- Lombardo, N. E. (2017). Boredom and modern culture. *Logos: A Journal of Catholic Thought & Culture*, 20(2), 36–59. <https://doi.org/10.1353/log.2017.0011>
- Maddi, S. R. (1967). The existential neurosis. *Journal of Abnormal Psychology*, 72(4), 311–325. <https://doi.org/10.1037/h0020103>

- Mael, F., Jex, M. (2015). Workplace boredom: An integrative model of traditional and contemporary approaches. *Group & Organization Management*, 40(2), 131–159. <https://doi.org/10.1177/1059601115575148>
- Mercer, K. B., Eastwood, J. D. (2010). Is boredom associated with problem gambling behaviour? It depends on what you mean by 'boredom'. *International Gambling Studies*, 10(1), 91–104. <https://doi.org/10.1080/14459791003754414>
- Merrifield, C., Danckert, J. (2014). Characterizing the psychophysiological signature of boredom. *Experimental Brain Research*, 232(2), 481–491. <https://doi.org/10.1007/s00221-013-3755-2>
- Mikulas, W. L., Vodanovich, S. J. (1993). The essence of boredom. *Psychological Record*, 43(1), 3–12.
- Miller, W. I. (1997). *The anatomy of disgust*. Harvard University Press. <https://doi.org/10.4159/9780674041066>
- Nett, U. E., Goetz, T., Daniels, L. M. (2010). What to do when feeling bored? Students' strategies for coping with boredom. *Learning and Individual Differences*, 20(6), 626–638. <https://doi.org/10.1016/j.lindif.2010.09.004>
- Oddy, M., Humphrey, M., Uttley, D. (1978). Subjective impairment and social recovery after closed head injury. *Journal of Neurology, Neurosurgery, and Psychiatry*, 41(7), 611–616. <https://doi.org/10.1136/jnnp.41.7.611>
- O'Hanlon, J. F. (1981). Boredom: Practical consequences and a theory. *Acta Psychologica*, 49(1), 53–82. [https://doi.org/10.1016/0001-6918\(81\)90033-0](https://doi.org/10.1016/0001-6918(81)90033-0)
- Oleszkiewicz-Zsurzs, E. (1985). Adaptacja skali poszukiwania wrażeń (SSS) M. Zuckermana do warunków polskich. *Przegląd Psychologiczny*, 28(4), 1123–1128.
- Parnowski, T., Jernajczyk, W. (1977). Inwentarz depresji Becka w ocenie nastroju osób zdrowych i chorych na choroby afektywne (ocena pilotażowa). *Psychiatria Polska*, 11(4), 417–421.
- Passik, S. D., Inman, A., Kirsh, K., Theobald, D., Dickerson, P. (2003). Initial validation of a scale to measure purposelessness, understimulation, and boredom in cancer patients: Toward a redefinition of depression in advanced disease. *Palliative and Supportive Care*, 1(1), 41–50. <https://doi.org/10.1017/S1478951503030062>
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, 36(1), 36–48. <https://doi.org/10.1016/j.cedpsych.2010.10.002>
- Pekrun, R., Goetz, T., Frenzel, A. C., Perry, R. P. (2011). *Academic Emotions Questionnaire (AEQ). User's manual* (wyd. 2). University of Munich. <https://doi.org/10.1037/t21196-000>

- Pekrun, R., Vogl, E., Muis, K. R., Sinatra, G. M. (2016). Measuring emotions during epistemic activities: The Epistemically-related Emotion Scales. *Cognition and Emotion*, 31(6), 1268–1276. <https://doi.org/10.1080/02699931.2016.1204989>
- Philips, A. (1993). *On kissing, tickling, and being bored. Psychoanalytic essays on the unexamined life*. Harvard University Press. <https://doi.org/10.4159/9780674417953>
- Pironti, V. A., Lai, M.-C., Müller, U., Bullmore, E. T., Sahakian, B. J. (2016). Personality traits in adults with attention-deficit hyperactivity disorder and their unaffected first-degree relatives. *BJPsych Open*, 2(4), 280–285. <https://doi.org/10.1192/bjpo.bp.116.003608>
- Plutchik, R. (1980). *Emotions: A psychoevolutionary synthesis*. Harper and Row.
- Postman, N. (1995). *Technopol: triumf techniki nad kulturą*. Państwowy Instytut Wydawniczy.
- Ragheb, M. G., Merydith, S. P. (2001). Development and validation of a unidimensional scale measuring free time boredom. *Leisure Studies*, 20(1), 41–59. <https://doi.org/10.1080/02614360122569>
- Reijseger, G., Schaufeli, W. B., Peeters, M. C. W., Taris, T. W., van Beek, I., Ouweneel, E. (2013). Watching the paint dry at work: Psychometric examination of the Dutch Boredom Scale. *Anxiety, Stress & Coping: An International Journal*, 26(5), 508–525. <https://doi.org/10.1080/10615806.2012.720676>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. <https://doi.org/10.1037/h0077714>
- Sharp, J. G., Zhu, X., Matos, M., Sharp, J. C. (2021). The Academic Boredom Survey Instrument (ABSI): A measure of trait, state and other characteristic attributes for the exploratory study of student engagement. *Journal of Further and Higher Education*, 45(9), 1253–1280. <https://doi.org/10.1080/0309877X.2021.1947998>
- Sommers, J., Vodanovich, S. J. (2000). Boredom proneness: Its relationship to psychological and physical-health symptoms. *Journal of Clinical Psychology*, 56(1), 149–155. [https://doi.org/10.1002/\(SICI\)1097-4679\(200001\)56:1<149::AID-JCLP14>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4679(200001)56:1<149::AID-JCLP14>3.0.CO;2-Y)
- Spielberger, C. D., Gorsuch, R. L., Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.
- Stickney, M. I., Miltenberger, R. G. (1999). Evaluating direct and indirect measures for the functional assessment of binge eating. *International Journal of Eating Disorders*, 26(2), 195–204. [https://doi.org/10.1002/\(SICI\)1098-108X\(199909\)26:2<195::AID-EAT9>3.0.CO;2-2](https://doi.org/10.1002/(SICI)1098-108X(199909)26:2<195::AID-EAT9>3.0.CO;2-2)
- Strelau, J. (2002). *Psychologia różnic indywidualnych*. Wydawnictwo Naukowe Scholar.
- Thackray, R. I. (1981). The stress of boredom and monotony: A consideration of the evidence. *Psychosomatic Medicine*, 43(2), 165–176. <https://doi.org/10.1097/00006842-198104000-00008>

- Todman, M. (2003). Boredom and psychotic disorders: Cognitive and motivational issues. *Psychiatry*, 66(2), 146–167. <https://doi.org/10.1521/psyc.66.2.146.20623>
- Todman, M. (2013). The dimensions of state boredom: Frequency, duration, unpleasantness, consequences, and causal attributions. *Educational Research International*, 1(1), 32–40.
- van Hooff, M. L. M., van Hooft, E. A. J. (2014). Boredom at work: Proximal and distal consequences of affective work-related boredom. *Journal of Occupational Health Psychology*, 19(3), 348–359. <https://doi.org/10.1037/a0036821>
- van Tilburg, W. A. P., Igou, E. R. (2012). On boredom: Lack of challenge and meaning as distinct boredom experiences. *Motivation and Emotion*, 36(2), 181–194. <https://doi.org/10.1007/s11031-011-9234-9>
- VandenBos, G. R. (red.). (2015). *APA dictionary of psychology* (wyd. 2). American Psychological Association. <https://doi.org/10.1037/14646-000>
- Vogel-Walcutt, J. J., Fiorella, L., Carper, T., Schatz, S. (2012). The definition, assessment, and mitigation of state boredom within educational settings: A comprehensive review. *Educational Psychology Review*, 24(1), 89–111. <https://doi.org/10.1007/s10648-011-9182-7>
- Watt, J. D., Ewing, J. E. (1996). Toward the development and validation of a measure of sexual boredom. *The Journal of Sex Research*, 33(1), 57–66. <https://doi.org/10.1080/00224499609551815>
- Winter, R. (2012). *Nuda w kulturze rozrywki. Poradnik*. Wydawnictwo WAM.
- World Health Organization (1997). *ICD-10. Międzynarodowa Statystyczna Klasyfikacja Chorób i Problemów Zdrowotnych. Rewizja dziesiąta*. Centrum Systemów Informacyjnych Ochrony Zdrowia.
- Wrześniewski, K., Sosnowski, T., Matusik, D. (2002). *STAI. Inwentarz stanu i cechy lęku. Polska adaptacja STAI. Podręcznik* (wyd. 2 zmien.). Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Zawadzki, B., Strelau, J., Szczepaniak, P., Śliwińska, M. (1998). *NEO-FFI. Inwentarz osobowości Paula T. Costy Jr. i Roberta R. McCrae. Adaptacja polska. Podręcznik*. Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Zimbardo, P. G., Boyd, J. (2011). *Paradoks czasu*. Wydawnictwo Naukowe PWN.
- Zuckerman, M. (1971). Dimensions of sensation seeking. *Journal of Consulting and Clinical Psychology*, 36(1), 45–52. <https://doi.org/10.1037/h0030478>
- Zuckerman, M. (1979). *Sensation seeking: Beyond the optimal level of arousal*. Lawrence Erlbaum Associates.
- Zuckerman, M., Aluja, A. (2015). Measures of sensation seeking. W: G. J. Boyle, D. H. Saklofske i G. Matthews (red.), *Measures of personality and social psychological constructs* (s. 352–380). Elsevier. <https://doi.org/10.1016/B978-0-12-386915-9.00013-9>

CZĘŚĆ II: PROBLEMY DIAGNOZY KLINICZNEJ I PRZEGLĄD NARZĘDZI BADAWCZYCH W ZAKRESIE NEUROPSYCHOLOGII ORAZ PSYCHOFIZJOLOGII

DOMINIKA ŻARNECKA  <https://orcid.org/0000-0002-4026-3900>

Wydział Psychologii, Uniwersytet Warszawski

EMILIA ŁOJEK  <https://orcid.org/0000-0002-0809-5601>

Wydział Psychologii, Uniwersytet Warszawski

Zaburzenia węchu w perspektywie neuropsychologicznej: znaczenie kliniczne i przegląd metod badania węchu

The neuropsychological perspective on olfactory disorders: Clinical significance and the review of olfactory testing methods

Autor korespondencyjny: Dominika Żarnecka

e-mail: d.zarnecka@student.uw.edu.pl

Publikacja finansowana ze środków Wydziału Psychologii Uniwersytetu Warszawskiego przyznanych przez MNiSW w formie subwencji na utrzymanie i rozwój potencjału badawczego w roku 2022 (501-D125-01-1250000 zlec*. 2022/501-D125-01-1250000 zlec 501000216).

Streszczenie

Zaburzenia funkcji węchu są problemem dotyczącym ok. 15% populacji ogólnej, zaś całkowita utrata węchu dotyka ok. 5%. Obraz kliniczny zaburzeń węchu jest zróżnicowany i może obejmować zmiany w zakresie: czułości węchu, zdolności różnicowania zapachów, zdolności identyfikowania zapachów. Do zaburzeń węchu zaliczana jest również tzw. phantasmia, czyli występowanie halucynacji węchowych. Struktury neuroanatomiczne zaangażowane w przetwarzanie informacji węchowych powiązane są również z procesami poznawczymi oraz procesami regulacji emocji, co stanowi prawdopodobne wytłumaczenie częstego występowania dysfunkcji węchu u osób z zaburzeniami neurologicznymi i psychicznymi. W pierwszej części rozdziału przedstawiono specyfikę zaburzeń węchu, obserwowanych u pacjentów z rozpoznaniem: choroby Alzheimera, choroby Parkinsona, otępienia z ciałami Lewy'ego, depresji, schizofrenii, zespołu stresu pourazowego oraz zespołu lęku uogólnionego. Szczegółowa wiedza na temat funkcjonowania węchu u tych pacjentów ma zastosowanie w procesie diagnostycznym, w planowaniu interwencji terapeutycznych, a także w prognozowaniu dalszego rozwoju zaburzeń. W drugiej części rozdziału szczegółowo omówiono dostępne metody badań węchu, uwzględniając cele poszczególnych narzędzi, ich specyfikę oraz opis procedury badawczej. Przegląd metod objął metody samoopisowe, testowe oraz fizjologiczne.

Słowa kluczowe: zaburzenia węchu, anosmia, hyposmia, diagnoza neuropsychologiczna

Abstract

Smell disorders affect about 15% of the general population, and total loss of smell affects about 5%. The clinical picture of olfactory disorders is diverse and may include changes in: olfactory threshold, odor discrimination and odor identification. Olfactory disorders also include phantasmia, defined as olfactory hallucinations. The neuroanatomical

structures involved in the olfactory processing are also associated with cognitive functions and emotional regulation, which is a likely explanation for the frequent occurrence of olfactory disorders in neurological and psychiatric patients. The first part of the chapter describes the characteristics of olfactory disorders observed in patients diagnosed with: Alzheimer's disease, Parkinson's disease, dementia with Lewy bodies, depression, schizophrenia, post-traumatic stress disorder and generalized anxiety disorder. Extensive knowledge of the olfactory functions in these patients is relevant for the diagnostic process, planning therapeutic interventions, and prognosis of a disease course. The second part of the chapter provides the detailed description of available olfactory testing methods, including the purpose, characteristics and conducting procedure for individual tools. The review of methods covered both self-descriptive methods, physiological methods and test methods.

Keywords: olfactory disorders, anosmia, hyposmia, neuropsychological diagnosis

ZNACZENIE WĘCHU W PERSPEKTYWIE NEUROPSYCHOLOGICZNEJ

Węch odgrywa kluczową rolę w wielu aspektach życia: wpływa na nasze decyzje dotyczące odżywiania, ostrzega przed niebezpieczeństwem, pomaga w ocenie przyjemnych i nieprzyjemnych bodźców, a także ma znaczenie dla relacji międzyludzkich (Marin i in., 2018). Zdecydowana większość badań oraz metod oceny węchu koncentruje się na tzw. węchu ortonosowym (*orthonasal olfaction*), związanym z przetwarzaniem bodźców zapachowych odbieranych na poziomie nozdrzy. Często pomijanym w literaturze zagadnieniem jest tzw. węch retronosowy (*retronasal olfaction*), który wiąże się z odbieraniem bodźców zapachowych na poziomie jamy ustnej, co ma miejsce podczas spożywania posiłków i napoi. Węch retronosowy odgrywa kluczową rolę w percepcji smaku (Rozin, 1982).

Prawidłowe funkcjonowanie węchu (tzw. normosmia) jest istotne dla utrzymania odpowiedniej jakości życia, a jego utrata wpływa negatywnie na

życie codzienne, nastrój i bezpieczeństwo jednostki (Athanassi i in., 2021). Według badań całkowitej utraty węchu (tzw. anosmii) doświadcza około 5% populacji ogólnej, zaś problem obniżonej sprawności węchu (tzw. hyposmii) dotyka ok. 15% (Hummel i Podlesek, 2021). Zaburzenia węchu nie ograniczają się jednak jedynie do jego osłabienia lub utraty, mogą mieć również charakter zmian jakościowych. O zmienionej percepcji zapachów (tzw. parosmii) mówimy w przypadku, kiedy obecny jest rzeczywisty bodziec zapachowy, jednak informacja na jego temat przetwarzana jest w sposób nieprawidłowy. Zdarza się też, że pacjent odczuwa wrażenia węchowe mimo braku obecności bodźca, wówczas mówimy o omamach węchowych i tzw. fantosmii (Altundag, 2023).

Przyczyny występowania tych zaburzeń mogą być różnorodne: od naturalnych procesów starzenia, przez chroniczne zapalenia i infekcje górnych dróg oddechowych, obecność polipów w nosie, urazy głowy, aż po zaburzenia neurodegeneracyjne i psychiatryczne oraz czynniki idiopatyczne (Hummel i Podlesek, 2021). W ostatnich latach badania zaczęły również uwzględniać nową przyczynę występowania zaburzeń węchu, tj. zakażenie wirusem SARS-CoV-2. Szacuje się, że około 35% zakażonych tym wirusem doświadcza całkowitej utraty węchu, zaś nieprawidłowości o różnym stopniu nasilenia dotyczą nawet 98% przypadków (Yan i in., 2020). Co więcej, u niektórych chorych zaburzenia te nie ustępują po wyzdrowieniu i mogą przejść w fazę chroniczną (Prem i in., 2022).

Zaburzenia węchu obserwowane są często u chorych z diagnozą: choroby Alzheimera (*Alzheimer's disease* [AD]), choroby Parkinsona (*Parkinson's disease* [PD]; Albers i in., 2006), otępienia z ciałami Lewy'ego (*Lewy Body Dementia* [LBD]; Olichney i in., 2005), depresji oraz schizofrenii (Rochet i in., 2018; Rupp, 2010). Są to więc pacjenci, z którymi niejednokrotnie w swojej pracy spotykają się psychologowie i neuropsychologowie. Wśród pacjentów z AD i idiopatyczną odmianą PD (*idiopathic Parkinson's disease* [iPD]) zaburzenia węchu często mogą być pierwszymi zauważalnymi objawami, poprzedzającymi wystąpienie klasycznych objawów (Albers i in., 2006). Ponadto ocena funkcji węchu pozwala na różnicowanie tych pacjentów od chorych na inne zaburzenia otępienne i neurologiczne (Doty, 2003; Katzenschlager i in., 2004; Müller i in., 2002). Z kolei trudności w zakresie identyfikacji zapachów mogą pomóc w różnicowaniu osób z wysokiej grupy ryzyka rozwinięcia schizofrenii i zaburzeń schizofrenopodobnych (Brewer i in., 2003). Najnowsze badania wskazują także na istnienie zależności między nasileniem zaburzeń węchu u osób, które chorowały na COVID-19, a trudnościami poznawczymi, jakich doświadczają (Di Stadio i in., 2022).

Możliwym wyjaśnieniem licznych zależności między funkcjami węchu a funkcjami poznawczymi oraz stanem emocjonalnym pacjentów jest fakt, że mają one wspólne podłoże neuroanatomiczne. Struktury zaangażowane zarówno w przetwarzanie informacji o bodźcach węchowych, jak i w regulację emocji i funkcjonowanie poznawcze to: opuszka węchowa, ciało migdałowe, hipokamp, kora oczodołowo-czołowa, kora zakrętu obręczy, wyspa, uzdeczka (Rochet i in., 2018). Co więcej, układ limbiczny i układ węchowy są uważane za „ewolucyjnie stare”. Systemy te wspólnie ewoluowały, co może tłumaczyć, dlaczego węch, tak istotny przy ocenie zagrożenia, poszukiwaniu pożywienia i wyczuwaniu innych osobników, pełni też istotną rolę w procesie regulacji motywacji, emocji i zachowań (Croy i Hummel, 2017).

Wszystko to przemawia za tym, aby funkcjonowanie węchu znalazło się w polu zainteresowań neuropsychologów i było uwzględniane podczas planowania oceny neuropsychologicznej pacjenta. W poniższym podrozdziale omówiono nieprawidłowości w funkcjonowaniu węchu, jakie mogą występować w niektórych zaburzeniach neurodegeneracyjnych i psychiatrycznych oraz zaprezentowano przegląd metod samoopisowych, testowych i fizjologicznych, które mogą być wykorzystywane podczas badań węchu.

DYSFUNKCJE WĘCHU W ZABURZENIACH NEURODEGENERACYJNYCH

Zaburzenia węchu w zakresie wykrywania, różnicowania oraz identyfikacji zapachów występują u 85–90% chorych na AD oraz ok. 90% chorych na iPD (z wyłączeniem przypadków związanych z mutacją w genie Parkin; Alber i in., 2006). Nieprawidłowości w zakresie funkcji węchu mogą pomóc w diagnozie różnicowej między osobami chorymi z AD a cierpiącymi na: otępienie naczyniowe, postępujące porażenie nadjądrowe lub depresję (Doty, 2003). Z kolei w przypadku pacjentów z iPD zaburzenia węchu mogą pomóc w różnicowaniu z następującymi zaburzeniami: zanikiem wieloukładowym (*Multiple System Atrophy* [MSA]; Müller i in., 2002), parkinsonizmem naczyniowym (Katzenschlager i in., 2004) i parkinsonizmem wywołanym metylofenyloitetrahydropirydynem (*1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine* [MPTP]) postępującym porażeniem nadjądrowym, zwyrodnieniem zwojowym korowo-podstawnym i drżeniem samoistnym (Albers i in., 2006).

Zaburzenia węchu u pacjentów z AD i iPD często pojawiają się wcześniej niż klasyczne objawy tych chorób (Albers i in., 2006). W podłużnym badaniu

Royall i in. (2002) zaobserwowano, że u osób starszych, które uzyskały niskie wyniki (≤ 18 na 40) w teście *University of Pennsylvania Smell Identification Test* (UPSIT lub SIT; omówionym szerzej w dalszej części rozdziału) w przeciągu następnych trzech lat wystąpiło pogorszenie funkcji poznawczych i pojawiły się problemy z pamięcią podobne do tych występujących w AD. Takiej zależności nie zaobserwowano u osób, które w UPSIT uzyskały wyniki powyżej 19 punktów.

Zaburzenia węchu są również powszechne w populacji osób cierpiących na LBD. Według niektórych badań anosmia występuje aż u 65% osób z rozpoznaniem LBD (podczas gdy dla chorych z AD odsetek ten wynosi 23%; Olichney i in., 2005). W ostatnich latach wysunięto hipotezę, według której obecność ciał Lewy'ego może korelować z występowaniem wcześniejszych lub bardziej nasilonych deficytów węchu (Albers i in., 2006). McShane i in. (2001) przeprowadzili badanie postmortem na 65 pacjentach z demencją. Wśród pacjentów, którzy za życia mieli problem z wyczuwaniem zapachów (co oceniano za pomocą prostej procedury z butelką wody lawendowej), zaobserwowano większą objętość korowych ciałek Lewy'ego, częstsze występowanie ciałek Lewy'ego w przednim zakręcie obręczy oraz większą utratę komórek w obrębie istoty czarnej.

Dane zebrane podczas obserwacji dysfunkcji węchu dostarczają potencjalnie istotnych informacji dla: wykrywania, diagnostyki oraz prognozy dalszego rozwoju zaburzeń neurodegeneracyjnych (Papazian i Pinto, 2021).

DYSFUNKCJE WĘCHU W ZABURZENIACH PSYCHICZNYCH

Niektóre spośród struktur mózgowych powiązanych z przetwarzaniem informacji węchowych należą jednocześnie do struktur układu limbicznego, odpowiadającego za regulację emocji i nastroju oraz odczuwanie lęku (Croy i Hummel, 2017). Są to: ciało migdałowe, hipokamp, kora oczodołowo-czołowa, kora przedczołowa, jądro półleżące, zakręt przyhipokampowy, wyspa, prążkowie brzuszne (Athanassi i in., 2021). Wspólne neuroanatomiczne podłoże tych procesów czyni węch szczególnie istotnym w obserwacji zaburzeń psychiatrycznych, takich jak zaburzenia lękowe, depresja, schizofrenia oraz choroba afektywna dwubiegunowa (Marin i in., 2023). Badania wskazują m.in. na nieprawidłowości w zakresie wykrywania i identyfikowania zapachów oraz na zwiększoną wrażliwość na bodźce zapachowe związane z doświadczoną traumą u osób z zespołem stresu pourazowego (*post-traumatic stress*

disorder [PTSD]; Cortese i in., 2017), a także na deficyty w zakresie czułości węchu oraz zdolności różnicowania i identyfikowania zapachów u pacjentów z zespołem lęku uogólnionego (*generalized anxiety disorder* [GAD]; Marin i in., 2023). Co więcej, wyniki uzyskane w badaniach prowadzonych przez Chen i in. (2021) wskazują na zależność między obniżeniem funkcji węchu a nasileniem somatycznych objawów lęku u chorych na GAD. Możliwym wytłumaczeniem tej zależności są dysfunkcje ciała migdałowatego – struktury jednocześnie odpowiedzialnej za przetwarzanie informacji zapachowych i odgrywającej kluczową rolę w zaburzeniach lękowych (Marin i in., 2023).

Relacje między dysfunkcjami węchu a zaburzeniami psychiatrycznymi zostały najszerzej opisane w przypadku depresji oraz schizofrenii.

Depresja

Badania wskazują na istnienie wyraźnej zależności między pogorszeniem funkcji węchu a występowaniem objawów depresyjnych (Rochet i in., 2018). Wśród pacjentów z diagnozą depresji obserwuje się osłabienie węchu w porównaniu z grupą kontrolną, z kolei u osób z dysfunkcjami węchu obserwuje się nasilenie objawów depresyjnych wprost proporcjonalnie do głębokości zaburzeń węchu (Kohli i in., 2016). Jakość życia osób chorych na depresję jest niższa wśród tych chorych, u których występują zmiany w zakresie przetwarzania bodźców węchowych (Athanassi i in., 2021). Co ciekawe, skuteczne leczenie depresji wiąże się z poprawą funkcji węchu (Athanassi i in., 2021). Według badań Croy i in. (2014) zarówno stosowanie antydepresantów, jak i korzystanie z psychoterapii wpływały na poprawę wcześniejszych deficytów w zakresie: postrzegania intensywności zapachów, różnicowania zapachów oraz oceny bodźców węchowych jako przyjemnych.

W ostatnich latach coraz więcej uwagi poświęcono zagadnieniu treningów węchowych i ich wpływu na nastrój osób badanych. Hummel i in. (2009) prowadzili 12-tygodniowy trening węchowy u osób cierpiących na zaburzenia węchu o zróżnicowanej etiologii. Metoda ta okazała się skuteczna: codzienny trening skutkował poprawą funkcji węchu. Podobny trening został wykorzystany w badaniu Wegener i in. (2018), przeprowadzonym w grupie osób starszych. Badanych podzielono na dwie grupy, jedna z nich przez okres pięciu miesięcy codziennie brała udział w treningu węchowym, zaś druga grupa codziennie wypełniała sudoku. Okazało się, że w pierwszej grupie zaobserwowano istot-

ne obniżenie nasilenia objawów depresyjnych (mierzonych za pomocą *Skali depresji Becka*) oraz poprawę w subiektywnej ocenie samopoczucia. Dokładny mechanizm wpływu zapachów na objawy depresyjne nie jest jeszcze znany i wymaga dalszych badań. Niewykluczone jednak, że w przyszłości trening węchowy będzie powszechnie stosowanym narzędziem terapeutycznym (Rochet i in., 2018).

Schizofrenia

Najczęściej opisywanym w literaturze deficytem węchu występującym w schizofrenii jest obniżona zdolność identyfikacji zapachów (*olfactory identification deficits* [OIDs]). Dysfunkcja ta pojawia się już na wczesnym etapie zaburzenia (Rupp, 2010). Problemy z identyfikacją zapachów zaobserwowano również wśród osób z grupy wysokiego ryzyka, u których później rozwinęła się schizofrenia. Sugeruje to, że OIDs mogą być czynnikiem przedchorobowym, poprzedzającym rozwój właściwego zaburzenia, nie ma jednak możliwości różnicowania na ich podstawie schizofrenii od innych zaburzeń afektywnych/psychoz (Brewer i in., 2003). Występowanie zaburzeń węchu w schizofrenii może również zależeć od czynników genetycznych, na co wskazują badania opisujące dysfunkcje węchowe u zdrowych krewnych oraz występowanie zarówno u członków rodziny, jak i u samych chorych specyficznej dla niektórych zapachów hyposmii (Turetsky i Moberg, 2009).

Według niektórych badań występowanie OIDs w schizofrenii koreluje z obecnością objawów negatywnych (Good i in., 2010), takich jak ograniczenie ekspresji afektu, ograniczenie przeżywanych emocji, ubóstwo mowy, utrata zainteresowań, obniżona potrzeba kontaktów społecznych i obniżone poczucie celu (Kirkpatrick i Galderisi, 2008). Z kolei badania Straussa i in. (2010) sugerują, że różnice w zakresie identyfikacji zapachów, obecne między chorymi, u których występują i u których nie występują objawy negatywne, zależą od rodzaju prezentowanych bodźców. W przypadku przyjemnych bodźców zapachowych pacjenci z objawami negatywnymi istotnie częściej oceniali je jako mniej przyjemne niż pacjenci bez objawów negatywnych. W przypadku nieprzyjemnych zapachów nie zaobserwowano istotnych różnic między dwiema grupami pacjentów.

Badania wskazują również na istnienie dodatniej korelacji między deklarowaną przez pacjentów anhedonią a nasileniem zaburzeń węchu u pacjentów w pierwszym epizodzie psychozy (Kamath i in., 2018). Co więcej, obecność OIDs podczas

pierwszego epizodu psychozy może być predyktorem gorszego funkcjonowania pacjenta w przyszłości w porównaniu z pacjentami, u których podczas pierwszego epizodu nie wystąpiły zaburzenia węchu (Good i in., 2010). Ocena funkcji węchu u pacjentów ze schizofrenią jest prostą i stosunkowo niedrogą procedurą, która jednak może dostarczyć bardzo istotnych informacji na temat prawdopodobnego kierunku rozwoju zaburzenia i pomóc wyselekcjonować tych pacjentów, którzy potrzebują intensywnej interwencji terapeutycznej (Good i in., 2010).

Dane na temat występowania, nasilenia i specyfiki dysfunkcji węchu u osób zmagających się z zaburzeniami psychiatrycznymi lub neurologicznymi pełnią istotną rolę w procesie diagnostycznym, podczas planowania działań terapeutycznych oraz przewidywania rozwoju zaburzenia i określania rokowania pacjentów. W celu oceny funkcjonowania węchu, badacze i klinicyści mogą sięgnąć po szereg różnorodnych metod zarówno kwestionariuszowych, jak i tych wykorzystujących rzeczywiste bodźce zapachowe. Dobór odpowiedniego narzędzia zależy od takich czynników, jak cel badania, poziom współpracy ze strony pacjenta, zdolność koncentracji uwagi osoby badanej czy ograniczenia czasowe. Wybrane metody badań węchu omówiono szerzej w dalszej części rozdziału.

PRZEGLĄD METOD BADANIA WĘCHU

Na potrzeby poniższego przeglądu metod badań węchu podzielono je na trzy główne grupy: (1) metody samoopisowe, obejmujące wywiady i kwestionariusze, (2) metody testowe wykorzystujące rzeczywiste bodźce zapachowe oraz (3) metody oceniające fizjologiczną reakcję organizmu na wystąpienie bodźca zapachowego.

W naukach medycznych powszechna jest również odmienna klasyfikacja, która w zależności od wymaganego poziomu współpracy osoby badanej wprowadza kategorie: metod subiektywnych, półobiektywnych oraz obiektywnych (Saussez i in., 2021). Terminologia ta może być myląca, ponieważ w naukach społecznych obiektywność jest jednym z kryteriów dobroci testu. W psychologicznym rozumieniu tego pojęcia, test jest obiektywny, jeżeli interpretacja jego wyników nie zależy od osoby, która tej interpretacji dokonuje. Istnieje jasno określony klucz obliczania i oceniania wyników, który pozwala na ograniczenie wpływu subiektywnych interpretacji osoby badającej (Hornowska, 2001). Według klasyfikacji medycznej badanie jest obiektywne wtedy, kiedy jego wynik nie zależy od poziomu zaangażowania i koncentracji uwagi osoby

badanej/pacjenta. Tak rozumiane metody obiektywne opierają się na rejestracji zmian w ośrodkowym układzie nerwowym (OUN) oraz występowania reakcji odruchowych wywołanych stymulacją bodźcami węchowymi (Dżaman, 2008). Najczęściej stosowaną metodą z tej grupy jest metoda potencjałów wywołanych przez bodźce zapachowe (*olfactory event-related potentials* [OERP]; Arana-Fernández i in., 2021; Dżaman, 2008). Metody samoopisowe nie są możliwe do przeprowadzenia bez współpracy osoby badanej, dlatego określane są jako metody subiektywne. Klasyfikacja medyczna uwzględnia też tzw. półobiektywne metody (*semi-objective*), które odnoszą się do metod będących pomiędzy obiektywnymi a subiektywnymi (Saussez i in., 2021). Oparte są one na rzeczywistych bodźcach zapachowych, jednak uzyskany podczas badania wynik zależy od współpracy osoby badanej, jej uwagi oraz zrozumienia polecenia (Rumeau i in., 2016).

W poniższym przeglądzie metod posłużono się klasyfikacją psychologiczną i podzielono metody na: samoopisowe, testowe oraz fizjologiczne. Zdecydowano się jednak na krótkie omówienie klasyfikacji medycznej, ponieważ jest ona powszechnie stosowana w literaturze anglojęzycznej.

Metody samoopisowe

Kwestionariusze wykorzystywane w badaniach węchu różnią się od siebie w zależności od celu, w jakim są wykorzystywane. Skonstruowano m.in. narzędzia służące ocenie: poziomu funkcjonowania węchu, znaczenia węchu, wpływu zaburzeń węchu na życie codzienne, zależności między węchem a procesami psychologicznymi (Han i in., 2021). Metody samoopisowe nie dostarczają równie wiarygodnych informacji nt. funkcjonowania węchu, co metody testowe wykorzystujące bodźce zapachowe oraz metody fizjologiczne. Są jednak mniej czasochłonne, a w konsekwencji też mniej obciążające dla osób badanych. Ponadto pozwalają na zebranie informacji dotyczących subiektywnego postrzegania funkcji i znaczenia węchu, które mogą wpływać na różnice indywidualne pacjentów w zakresie adaptacji do występujących objawów. Poniżej przedstawiono przegląd wybranych metod anglojęzycznych. Niestety, polskie adaptacje opisanych narzędzi nie są dostępne.

Questionnaire of Olfactory Disorders (QOD)

Kwestionariusz QOD złożony jest z listy 52 stwierdzeń, które można podzielić na trzy części: 39 stwierdzeń negatywnych, pięć stwierdzeń pozytywnych

oraz osiem stwierdzeń społecznie pożądaných. Pierwsza grupa stwierdzeń mierzy poziom cierpienia spowodowanego zaburzeniami węchu i odnosi się np. do: sytuacji z życia codziennego, które uległy pogorszeniu w związku z zaburzeniami węchu; sytuacji społecznych, których osoba badana zaczęła unikać w związku z pojawieniem się zaburzeń węchu; negatywnych emocji, jakie odczuwa osoba badana w związku z zaburzeniami węchu. Stwierdzenia pozytywne mierzą poziom akceptacji i adaptacji wobec występujących zaburzeń. Ostatnia grupa stwierdzeń dotyczy zachowań i postaw społecznie pożądaných (np. niespóźniania się) i pełni rolę tzw. *lie scale*, tzn. umożliwia ocenę, na ile osoba badana udzielała odpowiedzi zgodnych z oczekiwaniami społecznymi. Zadaniem osoby badanej jest ocena każdego ze stwierdzeń na trzystopniowej skali: *zgadzam się, częściowo się zgadzam, nie zgadzam się*. Za udzielone odpowiedzi przyznawane są kolejno 2, 1 i 0 punktów (Frasnelli i Hummel, 2005). Rzetelność dla podskali stwierdzeń negatywnych wynosi $\alpha = 0,93$, dla stwierdzeń pozytywnych $\alpha = 0,54$, zaś rzetelność test-retest wynosi między 0,71 a 0,78. Wyniki uzyskiwane w QOD korelowały istotnie z wynikami uzyskiwanymi przez osoby z zaburzeniami węchu w kwestionariuszach: *Beck's Depression Inventory* (BDI) i *Mood Inventory* (MI; Frasnelli i Hummel, 2005; Han i in., 2021).

The Multi-Clinic Smell and Taste Questionnaire (MCSTQ)

Kwestionariusz MCSTQ pozwala na ocenę subiektywnie postrzeganej czułości węchu oraz zaobserwowanych zmian w percepcji zapachów (parosmii). Zawiera także pytania dotyczące wpływu zaburzeń węchu na codzienne życie oraz pytania o strategie kompensacyjne. Dodatkową zaletą kwestionariusza jest uwzględnienie w nim pytań dotyczących fantosmii. MCSTQ pozwala zebrać informacje na temat występowania, stopnia nasilenia oraz czasu trwania halucynacji (Han i in., 2021). W praktyce klinicznej MCSTQ stosowany jest podczas badań populacji pacjentów ze zmienionym węchem, np. kobiet w ciąży (Nordin, Broman i in., 2004). Wadą narzędzia jest brak danych na temat jego trafności i rzetelności (Han i in., 2021).

Importance of Olfaction Questionnaire (IOQ)

Kwestionariusz IOQ pozwala na ocenę, na ile istotny w życiu codziennym osoby badanej jest węch. Złożony jest łącznie z 20 stwierdzeń, wobec których osoba badana ma za zadanie ustosunkować się za pomocą czterostopniowej skali: *zgadzam się, raczej się zgadzam, raczej się nie zgadzam, nie zgadzam się*. W obrębie kwestionariusza wyróżnić można trzy podskale, każdą złożoną

z sześciu stwierdzeń: *association-scale* dotyczącą wpływu zapachów na emocje, wspomnienia oraz ocenę; *application-scale*, która mierzy, na ile zapachy wpływają na życie codzienne osoby badanej oraz *consequence-scale* mierzącą wpływ węchu na procesy podejmowania decyzji. Na potrzeby praktyki klinicznej do ostatecznej wersji kwestionariusza dodano dwa pytania tworzące *aggravation scale*, służące ocenie, czy osoba badana ma tendencję do przeceniania znaczenia utraty węchu (Croy i in., 2010). Narzędzie cechuje dobra zgodność wewnętrzną $\alpha = 0,77$ (Croy i in., 2010). Autorzy nie podali wartości zgodności dla poszczególnych podskal, podają jednak wartości korelacji między każdą z podskal (*association-scale-application-scale*: $r = 0,5$; $p < 0,001$; *association-scale-consequence-scale*: $r = 0,4$; $p < 0,001$; *application-scale-consequence-scale*: $r = 0,5$; $p < 0,001$) oraz wartości korelacji każdej z podskal z wynikiem ogólnym (*association-scale*: $r = 0,81$; $p < 0,001$; *application-scale*: $r = 0,83$; $p < 0,001$; *consequence-scale*: $r = 0,77$; $p < 0,001$; Croy i in., 2010).

The Relational Scale of Olfaction (EROL)

EROL (*Escala Relacional sobre el Olfato*) to kwestionariusz służący ocenie wpływu zapachów na nastrój, zachowanie i funkcje poznawcze osoby badanej (Han i in., 2021). Skala złożona jest z 11 pytań (dotyczących np. wpływu noszenia perfum na poczucie pewności siebie, wpływu zapachów na zachowania seksualne i na postrzeganie przestrzeni). W zależności od pytania osoba badana jest proszona o określenie częstotliwości występowania danej sytuacji albo o określenie poziomu poczucia atrakcyjności lub pewności siebie w opisanej sytuacji. Kwestionariusz cechuje zgodność wewnętrzną na poziomie $\alpha = 0,76$ oraz rzetelność test-retest równa $r = 0,75$. Trafność narzędzia oceniono poprzez korelację z kwestionariuszami: *The Affective Impact of Odors Scale* ($r = 0,53$; $p < 0,001$) oraz *Odor Awareness Scale* (OAS; $r = 0,63$; $p < 0,001$; Burón i in., 2013).

The Self-administered Odor Questionnaire (SAOQ)

SAOQ to kwestionariusz, w którym osoba badana jest proszona o określenie, czy wybrane zapachy są dla niej: mocno odczuwalne, słabo odczuwalne, nieodczuwalne lub nieznane. SAOQ zawiera listę 20 różnych zapachów zarówno przyjemnych (np. kwiaty, truskawki), jak i nieprzyjemnych (np. pot, odchody, śmieci). Większość wymienionych zapachów związana jest z jedzeniem, napojami i przyprawami (np. curry, czosnek, pieczony chleb, pomarańcza, czekolada, kawa, zielona herbata). Narzędzie zostało skonstruowane w Japonii, na liście znajdują się więc również zapachy bardziej popularne w tamtejszym

kręgu kulturowym (np. sos sojowy, zupa miso, gotowany ryż, wodorosty; Takebayashi i in., 2011). Trafność narzędzia oceniono, porównując uzyskane wyniki z pomiarami czułości węchu wykonanymi olfaktometrem T&T (*Toyota and Takagi*) oraz wynikami wizualnej skali analogowej VAS (*visual analogue scale*). Skala VAS ma postać linii o długości 10 cm, na której badany jest poproszony o wskazanie, jak subiektywnie ocenia czułość swojego węchu, przy założeniu, że jeden z końców linii oznacza 0%, a przeciwny koniec 100% czułości. Korelacja wyników uzyskanych w SAOQ oraz T&T wyniosła $r = -0,58$; $p < 0,001$ w grupie osób badanych przed terapią zaburzeń węchu oraz $r = 0,59$; $p < 0,001$ w grupie badanych po terapii zaburzeń węchu. Korelacja wyników SAOQ oraz VAS wynosiła $r = 0,82$; $p < 0,001$ i $r = 0,83$; $p < 0,001$ odpowiednio dla badanych przed i po terapii zaburzeń węchu (Takebayashi i in., 2011). Autorzy nie podali danych dot. rzetelności narzędzia.

Hypsmia Rating Scale (HRS)

HRS to narzędzie, które powstało z myślą o pacjentach zmagających się z PD, ale jest skonstruowane w sposób pozwalający na wykorzystywanie go również w szerszej populacji. HRS zawiera listę sześciu przykładowych sytuacji, związanych z odczuwaniem różnego rodzaju zapachów (takich jak np. pójście do kwaciarni, poczucie zapachu perfum przytulającej nas osoby, przebywanie w pobliżu śmieci lub ścieków). Osoba badana proszona jest o określenie na pięciostopniowej skali, w jakim stopniu jest świadoma otaczających ją zapachów (Millar Verneti i in., 2012). Zgodność wewnętrzna narzędzia wynosi $\alpha = 0,89$. Trafność oceniono, porównując wyniki HRS do całkowitych wyników uzyskanych w teście *Sniffin' Sticks* przez osoby chore na PD oraz kontrolną grupę osób zdrowych ($r = 0,65$; $p < 0,001$; Millar Verneti i in., 2012).

The Affective Impact of Odor Scale (AIO)

AIO to narzędzie służące ocenie wpływu zapachów na postrzeganie nowych miejsc, osób, dań lub kosmetyków. Osoba badana proszona jest o określenie na czterostopniowej skali, jak często odczuwany zapach ma wpływ na jej pozytywną lub negatywną ocenę (odpowiedzi do wyboru: 0 – *nigdy*, 1 – *rzadko*, 2 – *czasami*, 3 – *często*; Wrzesniewski i in., 1999). Kwestionariusz AIO zawiera łącznie osiem pytań. Zgodność wewnętrzna wynosi 0,73–0,75. Wynik AIO koreluje z pamięcią zapachową ($r = 0,56$; $p < 0,01$), uwagą zapachową ($r = 0,3$; $p < 0,01$) oraz skojarzeniami zapachowymi ($r = 0,4$; $p < 0,01$; Wrzesniewski i in., 1999), nie ma jednak związku z rozpoznawaniem zapachów (Nguyen i in., 2012).

The Odor Awareness Scale (OAS)

OAS służy ocenie wpływu zapachów na uwagę, emocje, pamięć oraz zachowania konsumenckie. Dodatkowo zawiera pytania dotyczące czułości węchu i znaczenia zapachów w życiu codziennym (Smeets i in., 2008). Badania wykazały istotną zależność między wynikami uzyskanymi w OAS a ogólnym poziomem funkcjonowania węchu rozumianym jako całościowy wynik w testach z baterii *Sniffin' Sticks* (Kendall's $\tau = 0,13$; $p < 0,05$) oraz zdolnością identyfikacji zapachów mierzoną również tą baterią (Kendall's $\tau = 0,16$; $p < 0,01$; Nováková i in., 2014). Dodatkowo wyniki uzyskane w OAS mają związek z pamięcią węchową (Arshamian i in., 2011). Pewną wadą narzędzia może być jego długość: OAS złożone jest z 32 pytań, może być więc obciążające dla niektórych pacjentów (Han i in., 2021); 11 pytań dotyczy świadomości przyjemnych zapachów, zaś pozostałe 21 pytań dotyczy nieprzyjemnych zapachów. Zgodność wewnętrzna wynosi $\alpha = 0,77$ dla podskali świadomości przyjemnych zapachów oraz $\alpha = 0,80$ dla podskali świadomości nieprzyjemnych zapachów. Brak informacji na temat zgodności wewnętrznej całego narzędzia (Smeets i in., 2008).

Kwestionariusze służące ocenie wrażliwości na nieprzyjemne zapachy

Zapachy chemiczne

Skonstruowano kilka różnych kwestionariuszy pozwalających na ocenę wrażliwości osoby badanej na chemiczne zapachy. *The Chemical Odor Intolerance Index* (CII) jest krótkim narzędziem przesiewowym, w którym osoba badana jest proszona o określenie, jak często czuje się źle z powodu następujących pięciu zapachów: pestycydów, farby, perfum, spalin samochodowych, nowej wykładziny (Szarek i in., 1997). Podobnym narzędziem jest *The Chemical Odor Sensitivity Scale* (COSS). Kwestionariusz zawiera 11 opisów sytuacji związanych z reakcją na zapachy: farby, dymu, detergentów, benzyny, spalin oraz perfum. Osoba badana proszona jest o ocenę, na ile każda z opisanych sytuacji jej dotyczy za pomocą sześciostopniowej skali (Bailer i in., 2006). COSS jest narzędziem przesiewowym, powstałym w oparciu o złożony z 67 opisów sytuacji *Questionnaire of Chemical and General Environmental Sensitivity* (CGES). Nieco bardziej rozbudowanym narzędziem jest *The Chemical Sensitivity Scale* (CSS), złożonym z 21 stwierdzeń dotyczących m.in. wrażliwości na niektóre nieprzyjemne zapachy, zdolności koncentracji uwagi niezależnie od zapachów w otoczeniu, wpływu nieprzyjemnych zapachów na nastrój

(w szczególności na poziom irytacji; Nordin i in., 2003). Skonstruowano również skróconą wersję kwestionariusza *Chemical Sensitivity Scale for Sensory Hyperreactivity* (CSS-SHR), złożoną z 11 stwierdzeń i przeznaczoną dla grupy pacjentów z nadreaktywnością sensoryczną (Nordin, Millqvist i in., 2004). Dane nt. rzetelności i trafności wyżej wymienionych narzędzi prezentuje tabela 1.

Tabela 1

Kwestionariusze służące ocenie wrażliwości na zapachy chemiczne- rzetelność i trafność

Narzędzie	Rzetelność pomiaru	Trafność pomiaru
CII	$\alpha = 0,80-0,92$	Korelacje z CSS w grupie osób z astmą i alergiami: $r = 0,72; p < 0,001$; korelacje z CSS w grupie kontrolnej: $r = 0,53; p < 0,001$
COSS	$\alpha = 0,88-0,96$ rzetelność test-retest: $r = 0,90$	Korelacje z idiopatyczną nietolerancją środowiskową (<i>idiopathic environmental intolerance</i> [IEI]): $r = 0,53-0,81; p < 0,001$
CSS	$\alpha = 0,88$ rzetelność test-retest: $r = 0,87$	Korelacja z CII na poziomie $r = 0,65; p < 0,001$
CSS-SHR	$\alpha = 0,76-0,84$ rzetelność test-retest: $r = 0,87$	Korelacja z CII na poziomie $r = 0,74; p < 0,001$; korelacja z tolerancją na substancje zapachowe na poziomie $r = -0,72; p < 0,001$

Adnotacja. CII – *The Chemical Odor Intolerance Index*, COSS – *The Chemical Odor Sensitivity Scale*, CSS – *The Chemical Sensitivity Scale*, CSS-SHR – *Chemical Sensitivity Scale for Sensory Hyperreactivity*. Dane pochodzą z: Bailor i in. (2006), Han i in. (2021), Nordin i in. (2003), Nordin, Millqvist i in. (2004), Szarek i in. (1997).

Zapachy wydzielane przez człowieka

The Body Odor Disgust Scale (BODS). BODS służy ocenie wrażliwości na nieprzyjemne zapachy wydzielane przez człowieka. Wśród 12 opisów sytuacji rozróżnia się takie, kiedy źródłem nieprzyjemnych zapachów (m.in. zapachu potu, moczu lub oddechu) jesteśmy my sami, a kiedy ich źródłem są inne osoby z naszego otoczenia. Osoba badana proszona jest o określenie na skali od 1 do 5, na ile odczuwałaby wstręt w opisanych sytuacjach. Kwestionariusz BODS koncentruje się na roli węchu i zapachów w aktywacji behawioralnego układu odpornościowego. Zgodność wewnętrzną narzędzia wynosi $\alpha = 0,94$, a uzyskane wyniki istotnie korelują z wynikami uzyskiwanymi w innych skalach służących do pomiaru wstrętu (Liuzza i in., 2017).

Narzędzia przesiewowe

Zdarza się, że ze względu na ograniczenia czasowe, wysoką męczliwość pacjenta lub jego problemy z koncentracją uwagi konieczne jest sięgnięcie po narzędzia badawcze, których czas wykonania jest krótki. Przykładami takich narzędzi są m.in. skrócona wersja *Questionnaire of Olfactory Disorders* oraz *Self-reported Mini Olfactory Questionnaire*.

Skrócona wersja Questionnaire of Olfactory Disorders (QOD-NS)

Na podstawie listy 39 stwierdzeń negatywnych występujących w oryginalnej wersji QOD sformułowano kwestionariusz QOD-NS, składający się z 17 wybranych stwierdzeń. Jednak narzędzie to, chociaż wyraźnie krótsze od oryginalnego QOD, wciąż bywa obciążające dla pacjentów, którzy w krótkim czasie zobowiązani są do wypełnienia dużej liczby kwestionariuszy. Dlatego w 2019 roku stworzono nową, skróconą wersję QOD-NS, złożoną zaledwie z siedmiu stwierdzeń. Narzędzie to pozwala na szybką ocenę wpływu zaburzeń węchu na codzienne życie pacjenta (na jego zachowania w sytuacjach społecznych, nawyki żywieniowe, poczucie lęku oraz irytacji; Mattos i in., 2019). Skrócona wersja nazywana jest *Brief Questionnaire of Olfactory Disorders* (B-QOD), a zgodność wewnętrzna narzędzia wynosi $\alpha = 0,83\text{--}0,88$, zaś korelacja z wynikami uzyskiwanymi w oryginalnej wersji kwestionariusza QOD wynosi $r = 0,96\text{--}0,98$; $p < 0,05$ (Mattos i in., 2021).

Self-reported Mini Olfactory Questionnaire (Self-MOQ)

Self-MOQ pozwala na zebranie informacji na temat problemów z węchem, jakie napotyka osoba badana w codziennym życiu (np. trudności z czuciem zapachu kawy i świeżego chleba lub zapachów kwiatów w kwiaciarni; trudności z rozpoznaniem zapachu świeżo skoszonej trawy itd.). Narzędzie złożone jest z pięciu stwierdzeń. Osoba badana proszona jest o zaznaczenie *Tak*, jeżeli uważa, że dane stwierdzenie pasuje do jej dolegliwości albo *Nie*, jeżeli uważa, że opisana sytuacja jej nie dotyczy. Wynik uzyskany w Self-MOQ pozwala na określenie stanu węchu pacjenta jako: normosmię (prawidłowy węch), hyposmię (obniżony poziom funkcji węchu) lub anosmię (brak funkcji węchu). Kwestionariusz jest narzędziem o dobrej trafności i rzetelności; powinien być jednak stosowany tylko jako narzędzie przesiewowe (Zou i in., 2020). Zgodność wewnętrzna wynosi $\alpha = 0,84$, zaś korelacja z całkowitym wynikiem uzyskanym w *Sniffin' Sticks* wynosi $r = -0,60$; $p < 0,001$ (Zou i in., 2020).

Metody testowe

Metody testowe wykorzystują rzeczywiste bodźce zapachowe, które osoba badana ma za zadanie ocenić. Testy te koncentrują się przede wszystkim na następujących trzech funkcjach węchu: wyczuwanie zapachu, identyfikowanie zapachu, różnicowanie zapachów. Do najbardziej rozpowszechnionych metod testowych badania węchu zalicza się: bateria testów *Sniffin' Sticks Test* (SST) oraz UPSIT (Han i in., 2021). W większości opracowań anglojęzycznych testy te będą zaliczane do kategorii testów półobiektywnych, zgodnie z kryteriami klasyfikacji medycznej omawianej na początku rozdziału (Gong i in., 2021; Rumeau i in., 2016; Saussez, i in., 2021; Sonnet i in., 2017). Testy SST oraz UPSIT wykorzystują rzeczywiste bodźce sensoryczne (co jest elementem wspólnym z metodami obiektywnymi), jednak uzyskany wynik zależy od współpracy osoby badanej i jej zrozumienia polecenia (co z kolei jest cechą metod subiektywnych; Rumeau i in., 2016).

SST

Bateria testów SST wykorzystuje sztyfty zapachowe, umożliwiające ocenę: czułości węchu, zdolności różnicowania zapachów oraz identyfikacji zapachów. Czułość węchu badana jest za pomocą testu wykrywania, złożonego z 32 sztyftów bezzapachowych oraz 16 sztyftów zapachowych, nasączonych roztworem n-butanolu lub 2-fenyloetanolu o różnym stężeniu. Badanie podzielone jest na 16 poziomów, na każdym z nich prezentowane są dwa sztyfty bezzapachowe oraz jeden zapachowy. Osoba badana proszona jest o wskazanie sztyftu nasączonego substancją zapachową. W pierwszej kolejności prezentowane są sztyfty o najniższym stężeniu n-butanolu lub 2-fenyloetanolu, a następnie przechodzi się do sztyftów z coraz wyższym stężeniem. Kiedy osoba badana prawidłowo wskaże sztyft zapachowy, kolejność ekspozycji na bodźce zostaje odwrócona i prezentowane są sztyfty z coraz niższym stężeniem. Za każdym razem, kiedy osoba badana prawidłowo zidentyfikuje sztyft nasączony substancją zapachową, kolejność bodźców zostaje odwrócona. Procedura ta powtarzana jest siedmiokrotnie. Ostateczny wynik wylicza się na podstawie średniej z czterech ostatnich prawidłowych wskazań (Hummel i in., 1997).

Test różnicowania zapachów również złożony jest z 16 poziomów i 48 sztyftów, jednak w tym przypadku zamiast sztyftów bezzapachowych wykorzystywane są dwa sztyfty nasączone tą samą substancją zapachową (dystraktory). Na każdym z poziomów prezentowane są trzy sztyfty zapachowe: dwa identyczne

dyskraktery oraz trzeci sztyft o odmiennym zapachu. Osoba badana ma za zadanie wskazać, który spośród prezentowanych zapachów jest inny. Każda trójka zapachów prezentowana jest tylko raz. Wynik testu wyliczany jest na podstawie liczby prawidłowych wskazań (Hummel i in., 1997).

Test identyfikacji zapachów składa się z 16 sztyftów, z których każdy nasączony jest inną substancją zapachową. Sztyfty prezentowane są pojedynczo, a osoba badana ma za zadanie wybrać spośród czterech możliwych odpowiedzi właściwą nazwę dla każdego z zapachów. Prezentowane w teście zapachy to: pomarańcza, skóra, cynamon, mięta, banan, cytryna, lukrecja, terpentyna, czosnek, kawa, jabłko, goździki, ananas, róża, anyżek, ryba (Hummel i in., 1997). Rzetelność test-retest dla poszczególnych podskal SST wynosi $r = 0,61$ dla testu proggu wyczuwania zapachów, $r = 0,54$ dla testu różnicowania zapachów oraz $r = 0,73$ dla testu identyfikacji zapachów (Hummel i in., 1997).

W 2014 roku opublikowano wyniki badań nad adaptacją i normalizacją polskiej wersji SST, które wskazały na potrzebę wprowadzenia kilku zmian w polskich nazwach niektórych dyskraktorów (np. odpowiedź *skóra* została zamieniona na *skóra obuwnicza* w celu uniknięcia dwuznaczności; Sorokowska i Hummel, 2014).

SST – wersja przesiewowa

Na podstawie testu identyfikacji zapachów z baterii testów SST utworzono wersję przesiewową testu, złożoną z 12 sztyftów zapachowych. Zapachy wykorzystywane w tym teście to: cynamon, banan, cytryna, lukrecja, ananas, kawa, goździk, róża, skóra obuwnicza, ryba, pomarańcza, mięta. Procedura badawcza trwa około czterech minut (Hummel i Podlesek, 2021). Możliwe jest samodzielne przeprowadzenie testu przez osobę badaną (Mueller i Renner, 2006). Opracowano również warianty testu identyfikacji złożone z trzech zapachów (goździk, kawa, róża; Hummel i in., 2010) oraz z pięciu zapachów (pomarańcza, skóra, mięta, róża, ryba; Mueller i Renner, 2006). Zgodność wewnętrzna przesiewowej wersji SST wynosi $\alpha = 0,54$. Rzetelność narzędzia jest ograniczona, ale zdaniem autorów wystarczająca do celów przesiewowych (Hinz i in., 2019).

UPSIT

UPSIT służy ocenie zdolności identyfikowania zapachów i składa się z czterech książeczek, zawierających łącznie 40 próbek zapachowych (po 10 w każdej). Na dole każdej strony znajduje się pasek, który należy potrzeć ołówkiem w celu uwolnienia substancji zapachowej. Osoba badana jest pro-

szona o zidentyfikowanie zapachu i wskazanie jego właściwej nazwy spośród czterech możliwych odpowiedzi. Wynik testu (suma wszystkich prawidłowych odpowiedzi) można interpretować na podstawie norm uwzględniających wiek i płeć osoby badanej. UPSIT umożliwia różnicowanie poziomów dysfunkcji węchu, pozwala wykryć symulację oraz może zostać przeprowadzony samodzielnie przez osobę badaną (Doty i in., 1984). Rzetelność test-retest wynosi $r = 0,95$ dla pomiaru powtórzonego po upływie dwóch tygodni i $r = 0,92$ dla pomiaru powtórzonego po sześciu miesiącach w grupach młodych dorosłych i dorosłych w średnim wieku (Doty i in., 1984).

Brief Smell Identification Test (BSIT) i Pocket Smell Test (PST) – wersje przesiewowe

BSIT jest krótkim testem przesiewowym, powstałym w oparciu o UPSIT (nazywanym również SIT). BSIT składa się z 12 bodźców zapachowych (goździk, skóra, cytryna, bez, mentol, gaz ziemny, ananas, dym, mydło, truskawka, wiśnia, róża), a jego wykonanie zajmuje około pięciu minut (Menon i in., 2013). Korelacja wyników uzyskiwanych w UPSIT i BSIT wynosi $r = 0,89$; $p < 0,001$ (El Rassi i in., 2016). Skonstruowano także dwie wersje PST, narzędzia przesiewowego złożonego z czterech bodźców zapachowych (w wersji A są to: czekolada, truskawka, dym, skóra; w wersji B: mydło, winogrono, cebula, gaz ziemny). Jeżeli osoba badana nie rozpozna któregokolwiek z zapachów, zalecane jest przeprowadzenie badania z wykorzystaniem pełnej wersji UPSIT (Seasonics International, b.d.).

Analogicznym do BSIT narzędziem jest *Cross-cultural Smell Identification Test* (CCSIT). Test ten również jest złożony z 12 zapachów, uwalnianych pod wpływem potarcia kartki ołówkiem. Osoba badana ma za zadanie zidentyfikować zapach, wybierając jedną z czterech możliwych odpowiedzi do każdego z nich. Podobnie jak w przypadku BSIT, całe badanie trwa około pięciu minut. Wykorzystywane zapachy to: opona, czekolada, cynamon, benzyna, cytryna, cebula, rozpuszczalnik do farb, ananas, róża, mydło, dym i puder dla niemowląt. Autorzy CCSIT dobrali prezentowane zapachy w taki sposób, aby były one neutralne kulturowo (Doty i in., 1996). Rzetelność test-retest dla CCSIT wynosi $r = 0,71$ (Doty i in., 1996). Autorzy narzędzia porównali rozkład częstości dla wyników oraz średnie wyniki uzyskiwane w CCSIT i w odpowiednich 12 pozycjach UPSIT i nie zaobserwowali istotnych statystycznie różnic między nimi (Doty i in., 1996).

Pozostałe narzędzia testowe

SST i UPSIT są najbardziej rozpowszechnionymi, ale zdecydowanie nie jedynymi tego typu narzędziami do oceny węchu. Większość testów jest do

siebie podobna pod względem procedury badania, różni się jednak metodą prezentowania zapachów. W tym celu wykorzystywane mogą być: szklane buteleczki zawierające roztwory substancji zapachowych (jak np. w teście *European Test of Olfactory Capabilities*), specjalne dyskiety zapachowe (test *Zürcher Riechtest*) czy składane karty, które uwalniają zapach po ich rozłożeniu (test *Open Essence*; Hummel i Podlessek, 2021). Proste badanie przesiewowe *Alcohol Sniff Test* wykorzystuje saszetki z jednorazowymi gazikami nasączo-nymi alkoholem. Po otwarciu saszetki, zbliża się ją stopniowo coraz bliżej nozdrzy osoby badanej, której zadaniem jest zakomunikowanie, w którym momencie zaczyna czuć zapach alkoholu. Miarą czułości węchu w tej prostej próbie jest odległość gazika od nozdrzy w momencie, w którym zapach zaczął być wyczuwalny. Zaletą metody jest szybkość i łatwość wykonania (Davidson i Murphy, 1997). W czasie pandemii COVID-19 zaproponowano inny wariant tej metody, w którym w odległości 3 cm od nozdrzy prezentowane są gaziki nasączone roztworami alkoholu etylowego o różnym stężeniu, a osoba badana ma za zadanie wskazać gazik z roztworem o najniższym stężeniu (Calvo-Henriquez i in., 2020).

Badania węchu retronosowego

W ostatnich latach opracowano nową metodę badawczą, która dzięki wykorzystaniu specjalnych proszków zapachowych pozwala na ocenę funkcjonowania węchu retronosowego. W przeprowadzonym przez Yoshino i in. (2021) badaniu posłużono się 20 proszkami, zawierającymi substancje zapachowe odpowiadające następującym aromatom: kawa, pomarańcza, mięta, brzoskwinia, cebula, kokos, czarna porzeczka, cynamon, cola, ser, bazylia, miód, herbatnik, truskawka, melon, wanilia, wiśnia, czekolada, jogurt, cytryna. Poszczególne próbki proszku różniły się zapachem, jednak cechował je ten sam, trudno wyczuwalny, słodkawy smak. Procedurę badawczą rozpoczynano od zatkania nozdrzy osoby badanej, następnie umieszczano niewielką ilość proszku (ok. 0,05 g) na środkowej części grzbietowej języka, odblokowywano nozdrza i proszono osobę badaną o wykonanie wydechu przez nos. Po wykonaniu tej czynności badany był proszony o identyfikację zapachu i wybór właściwej nazwy spośród czterech możliwych odpowiedzi. Bódcze zapachowe prezentowane były w 30-sekundowych odstępach, a po każdej z prób osoba badana przeplukiwała usta wodą z kranu. Wynik testu liczony jest jako suma prawidłowych identyfikacji. Wykorzystanie zapachowych proszków jest dobrze zapowiadającą się metodą pozwalającą na ocenę węchu retronosowego bez

interferencji smaku. Jednak ze względu na stosunkowo niewielką próbę badawczą w eksperymencie Yoshino i in. (2021) metoda wymaga dalszych badań.

Metody fizjologiczne

Metody fizjologiczne pomagają na ocenę fizjologicznych reakcji organizmu na pojawienie się bodźców zapachowych (zgodnie z klasyfikacją medyczną zaliczane są więc do metod obiektywnych). Należą do nich takie techniki, jak węchowe potencjały wywołane (*olfactory event-related potentials* [OERP]), elektroolfaktogram (*electroolfactography* [EOG]), a także metody wykorzystujące funkcjonalny rezonans magnetyczny (*functional magnetic resonance imaging* [fMRI]) oraz pozytonową tomografię emisyjną (*positron emission tomography* [PET]; Arana-Fernández i in., 2021).

Najczęściej stosowaną metodą fizjologiczną oceny węchu jest technika potencjałów wywołanych przez bodźce zapachowe OERP. Jest to technika elektrofizjologiczna umożliwiająca obserwację sekwencji aktywacji obszarów mózgu związanych z węchem podczas przetwarzania bodźców zapachowych (obszary te to m.in. ciało migdałowe, przyśrodkowa część płata skroniowego, środkowa część kory oczodołowo-czołowej, kora wyspowa; Barresi i in., 2012). Metoda OERP pozwala na ocenę integralności centralnej ścieżki węchowej (Hummel i Podlesek, 2021), określenie intensywności bodźca zapachowego i różnicowanie zapachów aktywizujących nerw trójdzielny i nerw węchowy. Amplituda i czas latencji wywołanych potencjałów węchowych zależą m.in. od wieku i płci osoby badanej (Stuck i in., 2006). OERP jest metodą wymagającą niewielkiej współpracy ze strony osoby badanej, dzięki czemu może być wykorzystywana podczas badań z udziałem pacjentów o niskiej responsywności (Barresi i in., 2012). OERP stosuje się m.in. w badaniach nad węchem pacjentów z PD, AD, padaczką skroniową i stwardnieniem rozsianym (Barresi i in., 2012; Morgan i Murphy, 2002).

PODSUMOWANIE

Dysfunkcje węchu mogą być pierwszym objawem obecnym w niektórych zaburzeniach neurologicznych (w szczególności AD, iPD, LBD; Albers i in., 2006; Olichney i in., 2005) i psychicznych (depresji, schizofrenii, PTSD, GAD;

Chen i in., 2021; Cortese i in., 2017; Rochet i in., 2018; Rupp, 2010), poprzedzającym wystąpienie bardziej typowych objawów tych zaburzeń. Specyfika i nasilenie obserwowanych dysfunkcji dostarczają istotnych informacji w procesie diagnostycznym, które mogą być kluczowe dla diagnostyki różnicowej, pomagają w prognozowaniu dalszego rozwoju zaburzeń i mają znaczenie dla planowania podejmowanych interwencji terapeutycznych. Z tych względów badanie funkcji węchu powinno być uwzględniane podczas oceny neuropsychologicznej.

Przeprowadzając wywiad z pacjentem, warto nie ograniczać się do pojedynczego pytania o występowanie problemów z węchem i zamiast tego sięgnąć po specjalnie opracowane w tym celu kwestionariusze, które mają większe znaczenie kliniczne (Han i in., 2021). W miarę możliwości czasowych i możliwości pacjenta wskazane jest uzupełnienie badania o metody testowe lub fizjologiczne, wykorzystujące rzeczywiste bodźce zapachowe. Opisyane w rozdziale metody nie zostały wprowadzone zaadaptowane na warunki polskie (z wyjątkiem SST), wciąż jednak mogą być wykorzystywane w praktyce klinicznej. Przykładowo przegląd metod kwestionariuszowych może stanowić cenną wskazówkę dotyczącą doboru pytań podczas zbierania wywiadu z pacjentem.

Podczas dokonywania oceny węchu szczególnie istotne jest uwzględnienie takich czynników, jak dotychczasowe leczenie, przyjmowane leki, choroby współwystępujące, przebieg rozwoju objawów (czy rozwijały się stopniowo, czy pojawiły się nagle), czas trwania objawów i ich ewentualna zmienność, jakość życia, wywiad rodzinny, zwyczaje żywieniowe, alergie, palenie czy problemy zawodowe (Hummel i Podlesek, 2021). Ogromne znaczenie ma również badanie jamy nosowej w celu wykluczenia m.in. obecności polipów i innych czynników mogących zaburzać funkcje węchowe (Hummel i in., 2017).

W środowisku neuropsychologicznym istnieje potrzeba zarówno pogłębienia wiedzy na temat zaburzeń węchu, jak i podjęcia działań mających na celu zaadaptowanie na polskie warunki narzędzi badawczych pozwalających na ocenę węchu. Dalszych badań wymaga również wpływ treningu węchowego na funkcjonowanie różnych grup pacjentów. Obecnie metoda ta nie jest jeszcze wystarczająco przebadana, ale są podstawy przypuszczać, że mogłaby poprawić funkcjonowanie m.in. chorych na depresję i PD (Haehner i in., 2013; Wegener i in., 2018).

BIBLIOGRAFIA

- Albers, M. W., Tabert, M. H., Devanand, D. P. (2006) Olfactory dysfunction as a predictor of neurodegenerative disease. *Current Neurology and Neuroscience Reports*, 6, 379–386. <https://doi.org/10.1007/s11910-996-0018-7>
- Altundag, A. (2023). Parosmia and phantosmia: Managing quality disorders. *Current Otorhinolaryngology Reports*, 11(1), 19–26. <https://doi.org/10.1007/s40136-023-00441-w>
- Arana-Fernández, B., Santamaría-Gadea, A., Mariño-Sánchez, F., Mullol, J. (2021). Truths and fakes in the smell terminology during the COVID-19 outbreak. *Journal of Internal Medicine*, 290(2), 466–467. <https://doi.org/10.1111/joim.13295>
- Arshamian, A., Willander, J., Larsson, M. (2011). Olfactory awareness is positively associated to odour memory. *Journal of Cognitive Psychology*, 23(2), 220–226. <https://doi.org/10.1080/20445911.2011.483226>
- Athanassi, A., Dorado Doncel, R., Bath, K. G., Mandairon, N. (2021). Relationship between depression and olfactory sensory function: A review. *Chemical Senses*, 46, artykuł bjab044. <https://doi.org/10.1093/chemse/bjab044>
- Bailer, J., Witthöft, M., Rist, F. (2006). The Chemical Odor Sensitivity Scale: Reliability and validity of a screening instrument for idiopathic environmental intolerance. *Journal of Psychosomatic Research*, 61(1), 71–79. <https://doi.org/10.1016/j.jpsychores.2005.11.005>
- Barresi, M., Ciurleo, R., Giacoppo, S., Foti Cuzzola, V., Celi, D., Bramanti, P., Marino, S. (2012). Evaluation of olfactory dysfunction in neurodegenerative diseases. *Journal of the Neurological Sciences*, 323(1–2), 16–24. <https://doi.org/10.1016/j.jns.2012.08.028>
- Brewer, W. J., Wood, S. J., McGorry, P. D., Francey, S. M., Phillips, L. J., Yung, A. R., Anderson, V., Copolov, D. L., Singh, B., Velakoulis, D., Pantelis, C. (2003). Impairment of olfactory identification ability in individuals at ultra-high risk for psychosis who later develop schizophrenia. *The American Journal of Psychiatry*, 160(10), 1790–1794. <https://doi.org/10.1176/appi.ajp.160.10.1790>
- Burón, E., Bulbena, A., Barrada, J. R., Pailhez, G. (2013). EROL scale: A new behavioural olfactory measure and its relationship with anxiety and depression symptoms. *Actas Espanolas de Psiquiatria*, 41(1), 2–9.
- Calvo-Henriquez, C., Maldonado-Alvarado, B., Chiesa-Estomba, C., Rivero-Fernández, I., Sanz-Rodriguez, M., Villarreal, I. M., Rodriguez-Iglesias, M., Mariño-Sánchez, F., Rivero-de-Aguilar, A., Lechien, J. R., Martínez-Capoccioni, G., Saussez, S., Capasso, R., Karkos, P. D., Schriever, V., Martin-Martin, C., Alobid, I., Santamaría-Gadea, A., Fragola, C., [...] Marchan-López, Á. (2020). Ethyl alcohol threshold test: A fast, reliable and affordable olfactory assessment tool for COVID-19 pa-

- tients. *European Archives of Oto-Rhino-Laryngology*, 277(10), 2783–2792. <https://doi.org/10.1007/s00405-020-06131-3>
- Chen, X., Guo, W., Yu, L., Luo, D., Xie, L., Xu, J. (2021). Association between anxious symptom severity and olfactory impairment in young adults with generalized anxiety disorder: A case-control study. *Neuropsychiatric Disease and Treatment*, 17, 2977–2883. <https://doi.org/10.2147/NDT.S314857>
- Cortese, B. M., Schumann, A. Y., Howell, A. N., McConnell, P. A., Yang, Q. X., Uhde, T. W. (2017). Preliminary evidence for differential olfactory and trigeminal processing in combat veterans with and without PTSD. *NeuroImage: Clinical*, 17, 378–387. <https://doi.org/10.1016/j.nicl.2017.09.018>
- Croy, I., Buschhüter, D., Seo, H. S., Negoias, S., Hummel, T. (2010). Individual significance of olfaction: Development of a questionnaire. *European Archives of Oto-Rhino-Laryngology*, 267(1), 67–71. <https://doi.org/10.1007/s00405-009-1054-0>
- Croy, I., Hummel, T. (2017). Olfaction as a marker for depression. *Journal of Neurology*, 264(4), 631–638. <https://doi.org/10.1007/s00415-016-8227-8>
- Croy, I., Symmank, A., Schellong, J., Hummel, C., Gerber, J., Joraschky, P., Hummel, T. (2014). Olfaction as a marker for depression in humans. *Journal of Affective Disorders*, 160, 80–86. <https://doi.org/10.1016/j.jad.2013.12.026>
- Davidson, T. M., Murphy, C. (1997). Rapid clinical evaluation of anosmia. The Alcohol Sniff Test. *Archives of Otolaryngology-Head & Neck Surgery*, 123(6), 591–594. <https://doi.org/10.1001/archotol.1997.01900060033005>
- Di Stadio, A., Brenner, M. J., De Luca, P., Albanese, M., D’Ascanio, L., Ralli, M., Roccamatysi, D., Cingolani, C., Vitelli, F., Camaioni, A., Di Girolamo, S., Bernitsas, E. (2022). Olfactory dysfunction, headache, and mental clouding in adults with long-COVID-19: What is the link between cognition and olfaction? A cross-sectional study. *Brain Sciences*, 12(2), artykuł 154. <https://doi.org/10.3390/brainsci12020154>
- Doty, R. L. (2003). Odor perception in neurodegenerative diseases. W: R. L. Doty (red.), *Handbook of olfaction and gustation* (s. 850–890). CRC Press. <https://doi.org/10.1201/9780203911457-28>
- Doty, R. L., Marcus, A., William Lee, W. J. T. L. (1996). Development of the 12-item Cross-cultural Smell Identification Test (CC-SIT). *The Laryngoscope*, 106(3), 353–356. <https://doi.org/10.1097/00005537-199603000-00021>
- Doty, R. L., Shaman, P., Kimmelman, C. P., Dann, M. S. (1984). University of Pennsylvania Smell Identification Test: A rapid quantitative olfactory function test for the clinic. *The Laryngoscope*, 94(2), 176–178. <https://doi.org/10.1288/00005537-198402000-00004>
- Dżaman, K. (2008). Współczesne metody badania węchu i smaku. *Otorynolaryngologia*, 7(4), 173–177.

- El Rassi, E., Mace, J. C., Steele, T. O., Alt, J. A., Soler, Z. M., Fu, R., Smith, T. L. (2016). Sensitivity analysis and diagnostic accuracy of the Brief Smell Identification Test in patients with chronic rhinosinusitis. *International Forum of Allergy & Rhinology*, 6(3), 287–292. <https://doi.org/10.1002/alr.21670>
- Frasnelli, J., Hummel, T. (2005). Olfactory dysfunction and daily life. *European Archives of Oto-Rhino-Laryngology*, 262(3), 231–235. <https://doi.org/10.1007/s00405-004-0796-y>
- Gong, S. W., Ahmadi, S., Blackburn, S. L., Ulin, L., Citardi, M. J., Luong, A., Yao, W. C. (2021). Sniffin' Sticks to measure olfactory function and recovery following bilateral superior turbinate resection as part of endoscopic transsphenoidal approach. *The Annals of Otolaryngology, Rhinology, and Laryngology*, 130(6), 636–642. <https://doi.org/10.1177/0003489420965621>
- Good, K. P., Tibbo, P., Milliken, H., Whitehorn, D., Alexiadis, M., Robertson, N., Kopala, L. C. (2010). An investigation of a possible relationship between olfactory identification deficits at first episode and four-year outcomes in patients with psychosis. *Schizophrenia Research*, 124(1–3), 60–65. <https://doi.org/10.1016/j.schres.2010.07.010>
- Haehner, A., Tosch, C., Wolz, M., Klingelhofer, L., Fauser, M., Storch, A., Reichmann, H., Hummel, T. (2013). Olfactory training in patients with Parkinson's disease. *PLoS ONE*, 8(4), artykuł e61680. <https://doi.org/10.1371/journal.pone.0061680>
- Han, P., Su, T., Qin, M., Chen, H., Hummel, T. (2021). A systematic review of olfactory related questionnaires and scales. *Rhinology*, 59(2), 133–143. <https://doi.org/10.4193/Rhin20.291>
- Hinz, A., Luck, T., Riedel-Heller, S. G., Herzberg, P. Y., Rolfs, C., Wirkner, K., Engel, C. (2019). Olfactory dysfunction: Properties of the Sniffin' Sticks screening 12 test and associations with quality of life. *European Archives of Oto-Rhino-Laryngology*, 276(2), 389–395. <https://doi.org/10.1007/s00405-018-5210-2>
- Hornowska, E. (2001). *Testy psychologiczne. Teoria i praktyka*. Wydawnictwo Naukowe Scholar.
- Hummel, T., Pfetzing, U., Lötsch, J. (2010). A short olfactory test based on the identification of three odors. *Journal of Neurology*, 257(8), 1316–1321. <https://doi.org/10.1007/s00415-010-5516-5>
- Hummel, T., Podlesek, D. (2021). Clinical assessment of olfactory function. *Chemical Senses*, 46, artykuł bjab053. <https://doi.org/10.1093/chemse/bjab053>
- Hummel, T., Rissom, K., Reden, J., Hähner, A., Weidenbecher, M., Hüttenbrink, K. B. (2009). Effects of olfactory training in patients with olfactory loss. *The Laryngoscope*, 119(3), 496–499. <https://doi.org/10.1002/lary.20101>

- Hummel, T., Sekinger, B., Wolf, S. R., Pauli, E., Kobal, G. (1997). 'Sniffin' Sticks': Olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. *Chemical Senses*, 22(1), 39–52. <https://doi.org/10.1093/chemse/22.1.39>
- Hummel, T., Whitcroft, K. L., Andrews, P., Altundag, A., Cinghi, C., Costanzo, R. M., Damm, M., Frasnelli, J., Gudziol, H., Gupta, N., Haehner, A., Holbrook, E., Hong, S. C., Hornung, D., Hüttenbrink, K. B., Kamel, R., Kobayashi, M., Konstantinidis, I., Landis, B. N., [...] Welge-Luessen, A. (2017). Position paper on olfactory dysfunction. *Rhinology*, 56(1), 1–30. <https://doi.org/10.4193/Rhino16.248>
- Kamath, V., Lasutschinkow, P., Ishizuka, K., Sawa, A. (2018). Olfactory functioning in first-episode psychosis. *Schizophrenia Bulletin*, 44(3), 672–680. <https://doi.org/10.1093/schbul/sbx107>
- Katzenschlager, R., Zijlmans, J., Evans, A., Watt, H., Lees, A. J. (2004). Olfactory function distinguishes vascular parkinsonism from Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 75(12), 1749–1752. <https://doi.org/10.1136/jnnp.2003.035287>
- Kirkpatrick, B., Galderisi, S. (2008). Deficit schizophrenia: An update. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA)*, 7(3), 143–147. <https://doi.org/10.1002/j.2051-5545.2008.tb00181.x>
- Kohli, P., Soler, Z. M., Nguyen, S. A., Muus, J. S., Schlosser, R. J. (2016). The association between olfaction and depression: A systematic review. *Chemical Senses*, 41(6), 479–486. <https://doi.org/10.1093/chemse/bjw061>
- Liuzza, M. T., Lindholm, T., Hawley, C., Sendén, M. G., Ekström, I., Olsson, M. J., Larsson, M., Olofsson, J. K. (2017). The Body Odor Disgust Scale (BODS): Development and validation of a novel olfactory disgust assessment. *Chemical Senses*, 42(6), 499–508. <https://doi.org/10.1093/chemse/bjw107>
- Marin, C., Alobid, I., Fuentes, M., López-Chacón, M., Mullol, J. (2023). Olfactory dysfunction in mental illness. *Current Allergy and Asthma Reports*, 23, 153–164. <https://doi.org/10.1007/s11882-023-01068-z>
- Marin, C., Vilas, D., Langdon, C., Alobid, I., López-Chacón, M., Haehner, A., Hummel, T., Mullol, J. (2018). Olfactory dysfunction in neurodegenerative diseases. *Current Allergy and Asthma Reports*, 18(8), artykuł 42. <https://doi.org/10.1007/s11882-018-0796-4>
- Mattos, J. L., Bodner, T. E., Mace, J. C., Schlosser, R. J., Beswick, D. M., Ramakrishnan, V. R., Alt, J. A., Payne, S. C., Smith, T. L., Soler, Z. M. (2021). Psychometric properties of the brief version of the Questionnaire of Olfactory Disorders in patients with chronic rhinosinusitis. *International Forum of Allergy & Rhinology*, 11(10), 1436–1442. <https://doi.org/10.1002/alr.22800>

- Mattos, J. L., Edwards, C., Schlosser, R. J., Hyer, M., Mace, J. C., Smith, T. L., Soler, Z. M. (2019). A brief version of the Questionnaire of Olfactory Disorders in patients with chronic rhinosinusitis. *International Forum of Allergy & Rhinology*, 9(10), 1144–1150. <https://doi.org/10.1002/alr.22392>
- McShane, R. H., Nagy, Z., Esiri, M. M., King, E., Joachim, C., Sullivan, N., Smith, A. D. (2001). Anosmia in dementia is associated with Lewy bodies rather than Alzheimer's pathology. *Journal of Neurology, Neurosurgery, and Psychiatry*, 70(6), 739–743. <https://doi.org/10.1136/jnnp.70.6.739>
- Menon, C., Westervelt, H. J., Jahn, D. R., Dressel, J. A., O'Bryant, S. E. (2013). Normative performance on the Brief Smell Identification Test (BSIT) in a multi-ethnic bilingual cohort: A Project FRONTIER study. *The Clinical Neuropsychologist*, 27(6), 946–961. <https://doi.org/10.1080/13854046.2013.796406>
- Millar Verneti, P., Perez Lloret, S., Rossi, M., Cerquetti, D., Merello, M. (2012). Validation of a new scale to assess olfactory dysfunction in patients with Parkinson's disease. *Parkinsonism & Related Disorders*, 18(4), 358–361. <https://doi.org/10.1016/j.parkreldis.2011.12.001>
- Morgan, C. D., Murphy, C. (2002). Olfactory event-related potentials in Alzheimer's disease. *Journal of the International Neuropsychological Society*, 8(6), 753–763. <https://doi.org/10.1017/s1355617702860039>
- Mueller, C., Renner, B. (2006). A new procedure for the short screening of olfactory function using five items from the "Sniffin' Sticks" identification test kit. *American Journal of Rhinology*, 20(1), 113–116. <https://doi.org/10.1177/194589240602000121>
- Müller, A., Müngersdorf, M., Reichmann, H., Strehle, G., Hummel, T. (2002). Olfactory function in Parkinsonian syndromes. *Journal of Clinical Neuroscience*, 9(5), 521–524. <https://doi.org/10.1054/jocn.2001.1071>
- Nguyen, L. A., Ober, B. A., Shenaut, G. K. (2012). Odor recognition memory: Two encoding trials are better than one. *Chemical Senses*, 37(8), 745–754. <https://doi.org/10.1093/chemse/bjs060>
- Nordin, S., Broman, D. A., Olofsson, J. K., Wulff, M. (2004). A longitudinal descriptive study of self-reported abnormal smell and taste perception in pregnant women. *Chemical Senses*, 29(5), 391–402. <https://doi.org/10.1093/chemse/bjh040>
- Nordin, S., Millqvist, E., Löwhagen, O., Bende, M. (2003). The Chemical Sensitivity Scale: Psychometric properties and comparison with the Noise Sensitivity Scale. *Journal of Environmental Psychology*, 23(4), 359–367. [https://doi.org/10.1016/S0272-4944\(03\)00002-1](https://doi.org/10.1016/S0272-4944(03)00002-1)
- Nordin, S., Millqvist, E., Löwhagen, O., Bende, M. (2004). A short Chemical Sensitivity Scale for assessment of airway sensory hyperreactivity. *International Archives of*

- Occupational and Environmental Health*, 77(4), 249–254. <https://doi.org/10.1007/s00420-004-0504-7>
- Nováková, L., Varella Valentova, J., Havlíček, J. (2014). Engagement in olfaction-related activities is associated with the ability of odor identification and odor awareness. *Chemosensory Perception*, 7, 56–67. <https://doi.org/10.1007/s12078-014-9167-2>
- Olichney, J. M., Murphy, C., Hofstetter, C. R., Foster, K., Hansen, L. A., Thal, L. J., Katzman, R. (2005). Anosmia is very common in the Lewy body variant of Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 76(10), 1342–1347. <https://doi.org/10.1136/jnnp.2003.032003>
- Papazian, E. J., Pinto, J. M. (2021). Olfactory loss and aging: Connections with health and well-being. *Chemical Senses*, 46, artykuł bjab045. <https://doi.org/10.1093/chemse/bjab045>
- Prem, B., Liu, D. T., Besser, G., Sharma, G., Dultinger, L. E., Hofer, S. V., Matiasczyk, M. M., Renner, B., Mueller, C. A. (2022). Long-lasting olfactory dysfunction in COVID-19 patients. *European Archives of Oto-Rhino-Laryngology*, 279(7), 3485–3492. <https://doi.org/10.1007/s00405-021-07153-1>
- Rochet, M., El-Hage, W., Richa, S., Kazour, F., Atanasova, B. (2018). Depression, olfaction, and quality of life: A mutual relationship. *Brain Sciences*, 8(5), artykuł 80. <https://doi.org/10.3390/brainsci8050080>
- Royall, D. R., Chiodo, L. K., Polk, M. S., Jaramillo, C. J. (2002). Severe dysosmia is specifically associated with Alzheimer-like memory deficits in nondemented elderly retirees. *Neuroepidemiology*, 21(2), 68–73. <https://doi.org/10.1159/000048619>
- Rozin, P. (1982). “Taste-smell confusions” and the duality of the olfactory sense. *Perception & Psychophysics*, 31(4), 397–401. <https://doi.org/10.3758/bf03202667>
- Rumeau, C., Nguyen, D. T., Jankowski, R. (2016). How to assess olfactory performance with the Sniffin' Sticks Test. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 133(3), 203–206. <https://doi.org/10.1016/j.anorl.2015.08.004>
- Rupp, C. I. (2010). Olfactory function and schizophrenia: An update. *Current Opinion in Psychiatry*, 23(2), 97–102. <https://doi.org/10.1097/YCO.0b013e328336643f>
- Saussez, S., Hans, S., Chiesa-Estomba, C. M., Lechien, J. R. (2021). Psychophysical evaluations are important, usefulness and usability approaches to study the smell disorder related to COVID-19. *Journal of Internal Medicine*, 290(2), 468–469. <https://doi.org/10.1111/joim.13297>
- Seasonics International (b.d.). *The 4-item NHANES Pocket Smell Test (PST®)*. Pobrane 20 września 2024 z: <https://seasonics.com/product/the-4-item-nhanes-pocket-smell-test-pst>

- Smeets, M. A., Schifferstein, H. N., Boelema, S. R., Lensvelt-Mulders, G. (2008). The Odor Awareness Scale: A new scale for measuring positive and negative odor awareness. *Chemical Senses*, 33(8), 725–734. <https://doi.org/10.1093/chemse/bjn038>
- Sonnet, M. H., Nguyen, D. T., Nguyen-Thi, P. L., Arous, F., Jankowski, R., Rumeau, C. (2017). Fine-tuned evaluation of olfactory function in patients operated for nasal polyposis. *European Archives of Oto-Rhino-Laryngology*, 274, 2837–2843. <https://doi.org/10.1007/s00405-017-4580-1>
- Sorokowska, A., Hummel, T. (2014). Polska wersja testu Sniffin'Sticks – adaptacja i normalizacja. *Otolaryngologia Polska*, 68(6), 308–314. <https://doi.org/10.1016/j.otpol.2014.08.001>
- Strauss, G. P., Allen, D. N., Ross, S. A., Duke, L. A., Schwartz, J. (2010). Olfactory hedonic judgment in patients with deficit syndrome schizophrenia. *Schizophrenia Bulletin*, 36(4), 860–868. <https://doi.org/10.1093/schbul/sbn178>
- Stuck, B. A., Frey, S., Freiburg, C., Hörmann, K., Zahnert, T., Hummel, T. (2006). Chemosensory event-related potentials in relation to side of stimulation, age, sex, and stimulus concentration. *Clinical Neurophysiology*, 117(6), 1367–1375. <https://doi.org/10.1016/j.clinph.2006.03.004>
- Szarek, M. J., Bell, I. R., Schwartz, G. E. (1997) Validation of a brief screening measure of environmental chemical sensitivity: The chemical door intolerance index. *Journal of Environmental Psychology*, 17, 345–351. <https://doi.org/10.1006/jevp.1997.0071>
- Takebayashi, H., Tsuzuki, K., Oka, H., Fukazawa, K., Daimon, T., Sakagami, M. (2011). Clinical availability of a Self-administered Odor Questionnaire for patients with olfactory disorders. *Auris, Nasus, Larynx*, 38(1), 65–72. <https://doi.org/10.1016/j.anl.2010.05.013>
- Turetsky, B. I., Moberg, P. J. (2009). An odor-specific threshold deficit implicates abnormal intracellular cyclic AMP signaling in schizophrenia. *American Journal of Psychiatry*, 166(2), 226–233. <https://doi.org/10.1176/appi.ajp.2008.07071210>
- Wegener, B.-A., Croy, I., Hähner, A., Hummel, T. (2018). Olfactory training with older people. *International Journal of Geriatric Psychiatry*, 33(1), 212–220. <https://doi.org/10.1002/gps.4725>
- Wrzesniewski, A., McCauley, C., Rozin, P. (1999). Odor and affect: Individual differences in the impact of odor on liking for places, things and people. *Chemical Senses*, 24(6), 713–721. <https://doi.org/10.1093/chemse/24.6.713>
- Yan, C. H., Faraji, F., Prajapati, D. P., Ostrander, B. T., DeConde, A. S. (2020). Self-reported olfactory loss associates with outpatient clinical course in COVID-19. *International Forum of Allergy and Rhinology*, 10(7), 821–831. <https://doi.org/10.1002/alr.22592>

- Yoshino, A., Goektas, G., Mahmut, M. K., Zhu, Y., Goektas, O., Komachi, T., Okubo, K., Hummel, T. (2021). A new method for assessment of retronasal olfactory function. *The Laryngoscope*, 131(2), E324–E330. <https://doi.org/10.1002/lary.28698>
- Zou, L. Q., Linden, L., Cuevas, M., Metasch, M. L., Welge-Lüssen, A., Hähner, A., Hummel, T. (2020). Self-reported Mini Olfactory Questionnaire (Self-MOQ): A simple and useful measurement for the screening of olfactory dysfunction. *The Laryngoscope*, 130(12), E786–E790. <https://doi.org/10.1002/lary.28419>

MARTA SIEPSIAK  <https://orcid.org/0000-0001-8161-3367>

Wydział Psychologii, Uniwersytet Warszawski

WOJCIECH ŁUKASZ DRAGAN  <https://orcid.org/0000-0002-3349-6562>

Instytut Psychologii, Uniwersytet Jagielloński

NORA UGLIK-MARUCHA  <https://orcid.org/0000-0002-9030-8732>

Psychometrics and Measurement Lab, Biostatistics and Health Informatics
Department, Institute of Psychiatry, Psychology and Neuroscience,
King's College London

Diagnoza objawów mizofonii w badaniach naukowych i praktyce klinicznej – analiza porównawcza dwóch kwestionariuszy dostępnych w języku polskim – *MisoQuest* i *S-Five*

Diagnosis of symptoms of misophonia in scientific research and clinical practice – a comparative analysis of two questionnaires available in the Polish language – MisoQuest and S-Five

Autor korespondencyjny: Marta Siepsiak

e-mail: marta.siepsiak@psych.uw.edu.pl

Praca przygotowana w ramach stypendium Etiuda 7 2019/32/T/HS6/00219 i grantu Preludium 15 2018/29/N/HS6/01108.

Sposób cytowania: Siepsiak, M., Dragan, W. Ł., Uglik-Marucha, N. (2024). Diagnoza objawów mizofonii w badaniach naukowych i praktyce klinicznej – analiza porównawcza dwóch kwestionariuszy dostępnych w języku polskim – *MisoQuest* i *S-Five*. W: B. Zawadzki, K. Fronczyk i K. Bargiel-Matusiewicz (red.), *Wybrane problemy współczesnej diagnostyki klinicznej w psychologii* (s. 155–176). Wydawnictwo Liberi Libri. <https://doi.org/10.47943/lib.9788363487676.rozdzial06>

Streszczenie

Mizofonia to rodzaj nadwrażliwości dźwiękowej, charakteryzujący się negatywnymi reakcjami emocjonalnymi, fizjologicznymi i behawioralnymi w odpowiedzi na specyficzne, powtarzalne bodźce. Do najpopularniejszych bodźców mizofonicznych należą odgłosy wydawane przez człowieka, takie jak mlaskanie, oddychanie, pociąganie nosem, choć różne inne dźwięki i inne bodźce, np. wizualne, również mogą wywoływać reakcje mizofoniczne. Zaburzenie to obniża jakość życia, znacznie utrudniając codzienne funkcjonowanie. Choć mizofonii nie ma jeszcze w oficjalnych kryteriach diagnostycznych, dane opublikowane w ostatnich latach wyraźnie wskazują na jej odrębność od innych zaburzeń. Jako że nie ma jeszcze oficjalnych kryteriów diagnostycznych mizofonii, istotnym problemem jest jej definiowanie i pomiar zarówno w badaniach naukowych, jak i praktyce klinicznej. Odmienne sposoby operacjonalizacji tego zjawiska prowadzą do trudności z porównaniami wyników badań prowadzonych przez różnych naukowców oraz utrudniają diagnozę i terapię mizofonii. Dlatego niezwykle ważne jest rozwijanie i wykorzystywanie kwestionariuszy o dobrych wartościach psychometrycznych oraz zwalidowanych międzykulturowo. W rozdziale opisano i porównano dwa kwestionariusze dostępne w języku polskim – *MisoQuest* oraz *S-Five*. Oba narzędzia mają bardzo dobre właściwości psychometryczne określone w grupach polskojęzycznych, są dostępne w wielu innych językach i wykorzystywane w badaniach naukowych oraz praktyce klinicznej. Nie są one tożsame, dlatego zastosowanie danego kwestionariusza (lub obu na raz) powinno być uzależnione od celu badania i ram teoretycznych, w których prowadzona jest diagnoza.

Słowa kluczowe: mizofonia, *S-Five*, *MisoQuest*, diagnoza, nadwrażliwość dźwiękowa

Abstract

Misophonia is a type of sound over-responsivity characterized by negative emotional, physiological, and behavioral reactions in response to specific, repetitive stimuli. Among the most common triggers of misophonia, there are sounds made by humans, such as chewing, breathing, sniffing, although various other sounds and stimuli, including visual ones, can also evoke misophonic reactions. This disorder significantly reduces the quality of life, making daily functioning much more difficult. Although misophonia is not yet included in official diagnostic criteria, data published in recent years clearly indicate its distinctiveness from other disorders. Since there are no official diagnostic criteria for misophonia yet, a significant problem is its definition and measurement, both in scientific research and clinical practice. Different ways of operationalizing this phenomenon lead to difficulties in comparing the results of studies conducted by different researchers and hinder the diagnosis and treatment of misophonia. Therefore, it is extremely important to develop and utilize questionnaires with good psychometric properties and cross-cultural validation. This chapter describes and compares two questionnaires available in the Polish language – *MisoQuest* and *S-Five*. Both tools have very good psychometric properties determined in Polish-speaking groups, are available in many other languages, and are being used in scientific research and clinical practice. They are not identical, so the choice of a specific questionnaire (or both together) should depend on the research objectives and theoretical frameworks in which the diagnosis is conducted.

Keywords: misophonia, *S-Five*, *MisoQuest*, diagnosis, sound over-responsivity

DEFINICJA I KRÓTKA HISTORIA ROZWOJU ROZUMIENIA MIZOFONII

Mizofonia jest jednym z rodzajów nadwrażliwości dźwiękowych (Swedo i in., 2022). Osoby z tym zaburzeniem doświadczają silnych, awersyjnych reakcji w odpowiedzi na powtarzalne, niezależnie od ich głośności, dźwięki i inne związane z nimi bodźce. Najczęściej są to dźwięki związane ze spożywaniem pokarmów, wydawane przez człowieka. Głównymi, choć nie jedynymi, emocjami pojawiającymi się w przebiegu reakcji mizofoniczej są złość, panika, niepokój, czasami obrzydzenie (Brout i in., 2018; Jager i in., 2020; Vitoratou i in., 2021). Objawy pojawiają się zazwyczaj we wczesnym dzieciństwie i prawdopodobnie utrzymują się przez całe życie. Choć odgłosy mlaskania wydają się wywoływać nieprzyjemne emocje u wielu osób (YouGov, 2022), mizofonia zdaje się być związana z krańcowym nasileniem takich reakcji. Innymi słowy, mizofonia nie jest terminem opisującym często spotykane „niełubienie” pewnych odgłosów, nie jest preferencją czy cechą, ale zaburzeniem sprawiającym cierpienie, znacząco wpływającym na dobrostan, utrudniającym codzienne funkcjonowanie. Choć mizofonia bardzo często współwystępuje z innymi zaburzeniami, głównie z różnymi zaburzeniami lękowymi lub depresyjnymi (np. Erfanian i in., 2019; Rosenthal i in., 2022; Siepsiak i in., 2022), nie towarzyszy stale żadnemu specyficznemu zaburzeniu ani nie wpasowuje się w objawy żadnego konkretnego zaburzenia. Może też występować jako jedyne zaburzenie u danej osoby. Mimo istnienia bogatej literatury, wskazującej na odrębność mizofonii od innych zaburzeń, nie jest ona jeszcze włączona w oficjalne systemy diagnostyczne, tj. do *Międzynarodowej Statystycznej Klasyfikacji Chorób i Problemów Zdrowotnych (International Classification of Diseases, 11th Revision [ICD-11])* czy *Diagnostycznego i Statystycznego Podręcznika Zaburzeń Psychiczych (Diagnostic and Statistical Manual of Mental Disorders, 5th Edition [DSM-5])*. Jest to związane głównie z faktem, iż jest to stosunkowo nowo opisane zjawisko. Badania naukowe nad mizofonią trwają od nieco ponad 10 lat, większość z nich została opublikowana po 2020 roku, a więc nie miała ona jeszcze szans na oficjalne uwzględnienie w klasyfikacjach chorób i zaburzeń. Nie zostało też nadal ustalone, jakiego rodzaju jest to zaburzenie ani w jakim obszarze diagnostycznym powinno zostać umieszczone – chorób i zaburzeń neurologicznych, neurorozwojowych, psychicznych czy audiologicznych. Na pewno jednak jest to zaburzenie złożone, wymagające multidyscyplinarnego podejścia.

Termin *mizofonia* po raz pierwszy został użyty przez Jastreboffów (M. M. Jastreboff i Jastreboff, b.d.; P. J. Jastreboff i Jastreboff, 2000). Z kolei pierwsze

badania nad mizofonią opublikowano dopiero 11 lat później (Edelstein i in., 2013). Od tego czasu, pod wpływem badań i obserwacji klinicystów, rozumienie mizofonii podlega zmianom. Zaproponowano różne sposoby konceptualizacji i mierzenia tego zaburzenia (np. Jager i in., 2020; Rosenthal i in., 2021; Schröder i in., 2013; Siepsiak, Śliwerski i Dragan, 2020; Vitoratou i in., 2021; Wu i in., 2014), co z dużym prawdopodobieństwem wpływa na uzyskiwane wyniki badań. Takie koleje rzeczy w przypadku nowego zjawiska czy zaburzenia wydają się nieuniknione. Mnogość perspektyw pozwala na lepsze poznanie danego problemu. Różne sposoby konceptualizacji mizofonii pozwoliły również formułować hipotezy dotyczące heterogeniczności tego zaburzenia. Badanie przeprowadzone w Polsce (Siepsiak i in., 2022) na osobach z różnymi nadwrażliwościami dźwiękowymi oraz bez problemów tego rodzaju wykazało, że psychiczne korelaty mizofonii są różne, w zależności od tego, w jaki sposób zostanie ona zdefiniowana. Wyniki tych analiz podkreślają konieczność badania mizofonii definiowanej w różny sposób, również znacznie węższy, niż zostało to zaproponowane przez autorów nazwy tego zaburzenia (P. J. Jastreboff i Jastreboff, 2014), na przykład poprzez kontrolę rodzaju bodźców wywołujących objawy reakcji mizofonicznej, jak i rodzaj pojawiających się w odpowiedzi na te bodźce emocji.

Znane są już zarówno psychofizjologiczne (Edelstein i in., 2013; Kumar i in., 2017; Siepsiak i in., 2023), jak i neurobiologiczne korelaty mizofonii (Eijsker i in., 2021; Kumar i in., 2017; Schröder i in., 2019). Neasciu i in. (2022) zauważyli, że choć mizofonia w pewnych warunkach dzieli z innymi zaburzeniami korelaty neurobiologiczne, to jednak duża część z nich ma charakter unikatowy tylko dla niej. W ostatnim czasie opublikowano ponadto szereg badań (Heller i Smith, 2022; Samermit i in., 2022; Savard i in., 2022; Siepsiak i in., 2023) wskazujących na istotne znaczenie przetwarzania poznawczego czy kontekstu w nasileniu reakcji mizofonicznych. Warto przy tym podkreślić, że nie ma dowodów, które wskazywałyby na pojawianie się reakcji mizofonicznych w konsekwencji występowania nieadaptacyjnych myśli, przekonań, doświadczeń czy warunkowania klasycznego, tak jak to przyjmuje się w klasycznych modelach poznawczo-behawioralnych, dotyczących zaburzeń lękowych czy depresyjnych. Ocena poznawcza bodźca i sytuacji może jednak nasilać lub obniżać intensywność i czas trwania reakcji, np. sprawiać, że emocje, wraz z pobudzeniem psychofizjologicznym, utrzymują się nawet po ustaniu dźwięku.

W ostatnich trzech latach opublikowano na świecie więcej badań dotyczących mizofonii niż przez wszystkie poprzednie lata łącznie. Choć nadal nie ma

jednej, oficjalnej definicji mizofonii, za historyczny krok na drodze rozwoju jej rozumienia można uznać na pewno konsensus opublikowany w 2022 roku (Swedo i in., 2022), w którym podkreślono jednak tymczasowość ustaleń i konieczność dalszego badania tego zaburzenia.

Konsekwencją rozwoju badań dotyczących mizofonii jest opracowanie szeregu narzędzi służących do pomiaru tego zaburzenia. Pierwszym opublikowanym i do niedawna najczęściej wykorzystywanym kwestionariuszem mizofonii jest *Misophonia Questionnaire* (MQ) autorstwa Wu i in. (2014). Jako że w momencie tworzenia tej skali nie istniały jeszcze żadne konkretne kryteria mizofonii, narzędzie to ma naturalnie swoje ograniczenia. Mizofonia jest tu rozumiana jako nadwrażliwość na bardzo różnorodne bodźce dźwiękowe, z różnorodnymi reakcjami emocjonalnymi (np. smutek lub przygnębienie bez obecności innych emocji w reakcji na szeleszczenie), niekoniecznie wskazującymi na występowanie tego zaburzenia w dzisiejszym rozumieniu. Głównym ograniczeniem tego narzędzia jest jednak to, że skala nasilenia objawów mizofonii, na podstawie której określa się punkt odcięcia, została stworzona na podstawie skali przeznaczonej do oceny objawów zaburzeń obsesyjno-kompulsyjnych. Dziś już wiadomo, że te dwa zaburzenia, pomimo pewnych punktów wspólnych, znacznie się od siebie różnią. Niemniej jednak MQ odegrał niezwykle istotną rolę w rozwoju badań nad mizofonią. Kwestionariusz ten został przetłumaczony na język polski z wykorzystaniem metody tłumaczenia zwrotnego (*back-translation*), we współpracy z autorką narzędzia dr Moniką Wu, i był wykorzystywany w badaniach dotyczących mizofonii w Polsce. Jak dotąd nie zostały opublikowane dane dotyczące właściwości psychometrycznych polskiej wersji MQ. Narzędzie to jest jednak dostępne do celów badań naukowych u pierwszej autorki tego rozdziału.

W ostatnim czasie powstało wiele nowych narzędzi do pomiaru mizofonii, które są coraz bardziej precyzyjne i rozbudowane, np. *Duke Misophonia Questionnaire* (Rosenthal i in., 2021), *Berlin Misophonia Questionnaire-Revised* (Remmert, Schmidt i in., 2022), *Misophonia Response Scale* (Dibb i in., 2021). Warty uwagi są też testy psychoakustyczne do identyfikowania mizofonii i hiperakuzji (Enzler, Fournier i Noreña, 2021; Enzler, Lorient i in., 2021), w których osobom badanym prezentowane są różne odgłosy, a ich zadaniem jest określenie swoich reakcji. Niedawno został też opublikowany ustrukturyzowany wywiad diagnostyczny *Duke Misophonia Interview* (Guetta i in., 2022), który został już przetłumaczony na j. polski metodą tłumaczenia zwrotnego, we współpracy z autorami oryginalnej wersji narzędzia. Obecnie jest on wykorzystywany w ba-

daniach w Polsce i jest dostępny u pierwszej autorki tego rozdziału zarówno do celów naukowych, jak i do praktyki klinicznej. Szczegółowe opisanie i wymienianie wszystkich metod pomiaru objawów mizofonii dostępnych na świecie nie jest możliwe w tej publikacji, a powyższy opis ma na celu jedynie zarysowanie różnorodności narzędzi, sposobów diagnozowania i rozumienia mizofonii.

Celem niniejszego rozdziału jest omówienie specyfiki dwóch kwestionariuszy do pomiaru mizofonii, dostępnych w języku polskim: *S-Five* (Uglik-Marucha i in., 2024; Vitoratou i in., 2021) oraz *MisoQuest* (Siepsiak, Sobczak i in., 2020; Siepsiak, Śliwerski i Dragan, 2020).

CHARAKTERYSTYKA DWÓCH NARZĘDZI DO POMIARU MIZOFONII DOSTĘPNYCH OBECNIE W JĘZYKU POLSKIM

Tworzenie, charakterystyka i znaczenie *MisoQuest*

MisoQuest (Siepsiak, Sobczak i in., 2020; Siepsiak, Śliwerski i Dragan, 2020) był pierwszym na świecie w pełni zwalidowanym narzędziem do pomiaru mizofonii. Punktem wyjścia procesu konstrukcji *MisoQuest* były kryteria diagnostyczne zaproponowane przez Schrödera i in. (2013). Ostateczna wersja tego narzędzia uwzględnia modyfikacje wspomnianych kryteriów, będące wynikiem analizy literatury oraz rezultatów wieloetapowego procesu walidacji *MisoQuest*. Kryteria diagnostyczne Schrödera i in. (2013), na podstawie których stworzono wyjściowe pozycje kwestionariusza, przedstawiają się następująco:

- A. Obecność lub oczekiwanie specyficznych dźwięków wytwarzanych przez ludzi (np. dźwięki jedzenia, dźwięki oddychania) powodują impulsywną awersyjną reakcję somatyczną (*impulsive aversive physical reaction*), która zaczyna się irytacją albo zniesmaczeniem i natychmiast przeradza się w złość.
- B. Zaistniała złość powoduje poczucie utraty kontroli, a czasami wybuchy agresji.
- C. Dana osoba ocenia złość i wstręt jako reakcje nadmierne, nieuzasadnione czy nieproporcjonalne do okoliczności.
- D. Dana osoba stara się unikać sytuacji związanych z bodźcami awersyjnymi lub jeśli ich nie unika, ponosi koszty związane z wysokim dyskomfortem, złością i wstrętem.
- E. Złość, wstręt i unikanie powodują u danej osoby znaczny niepokój (np. przeszkadzają jej) lub mają wpływ na jej codzienne życie. Na przy-

kład złość czy wstręt mogą sprawiać, że wykonywanie zadań w pracy, spotykanie się ze znajomymi, uczestniczenie w lekcjach czy wchodzenie w interakcje międzyludzkie jest dla danej osoby trudne.

- F. Złość, wstręt i unikanie nie mogą być lepiej wytłumaczone przez inne zaburzenia, takie jak zaburzenia obsesyjno-kompulsywne (np. odraza wraz z obsesyjną obawą zatrucia się) czy zaburzenie stresowe pourazowe (np. unikanie bodźców związanych z traumą; Schröder i in., 2013; tłumaczenie: Siepsiak i Dragan, 2019, s. 449)

Kryteria przedstawione przez Schrödera i in., choć szybko stały się przedmiotem dyskusji i obecnie nie są kryteriami akceptowanymi przez wszystkich (P. J. Jastreboff i Jastreboff, 2014), w 2018 roku, kiedy powstawał kwestionariusz *MisoQuest*, były jedynym spójnym opisem tego zaburzenia. W 2020 roku opublikowano zaktualizowane kryteria Schrödera (Jager i in., 2020), gdzie do bodźców mogących wywoływać reakcję mizofoniczną dodano również bodźce wizualne lub inne doznania sensoryczne, przy zachowaniu konieczności występowania nadwrażliwości na bodźce dźwiękowe pochodzące z ust lub nosa; obok złości, do typowych emocjonalnych reakcji mizofonicznych dodano irytację i obrzydzenie; podkreślono, że (rzadkie) zachowania agresywne są związane z nagłą, awersyjną reakcją fizjologiczną i poczuciem braku kontroli.

Po stworzeniu puli pozycji testowych, sędziowie kompetentni (dwóch psychologów-psychoterapeutów różnych nurtów, z doświadczeniem pracy z osobami z mizofonią) oceniali poszczególne propozycje oraz proponowali ich poprawki pod kątem zgodności z kryteriami Schrödera i in. (2013), rezultatami ówczesnych badań dotyczących mizofonii oraz zrozumiałości, a ostateczna decyzja dotycząca pozycji została podjęta przez pierwszą autorkę narzędzia, w porozumieniu z sędziami kompetentnymi i współautorami. Tak uzyskaną pulą pytań przebadano 383 osoby, wśród których były osoby z samodiagnozowaną mizofonią, tzn. identyfikujące się jako osoby reagujące silnymi emocjami na selektywne, ciche, codzienne odgłosy wydawane przez usta lub nos człowieka, oraz osoby bez żadnych nadwrażliwości dźwiękowych. Dane uzyskane w tym badaniu zostały przeanalizowane pod kątem właściwości psychometrycznych narzędzia. W konsekwencji liczba twierdzeń składająca się na *MisoQuest* została zredukowana do 21. W ostatniej fazie badań przeanalizowano dane od kolejnych 322 osób, które poza odpowiadaniem na 21 twierdzeń z kwestionariusza, otrzymały listę twierdzeń pochodzących bezpośrednio z kryteriów diagnostycznych Schrödera i in. (2013), w celu wstępnego zweryfikowania trafności zewnętrznej *MisoQuest*. Dodatkowo określono stabilność

bezwzględna narzędzia, analizując dane zebrane na grupie 97 osób podczas dwukrotnego pomiaru w odstępie pięciu tygodni.

W ostatecznej wersji *MisoQuest* znalazło się 14 twierdzeń z 5-stopniową skalą odpowiedzi (od *zdecydowanie nie zgadzam się* do *zdecydowanie zgadzam się*), tworzących jedną skalę, z doskonałą spójnością wewnętrzną ($\alpha = 0,96$), wysoką stabilnością w czasie i bardzo dobrą trafnością zewnętrzną, wstępnie określoną poprzez porównanie dwóch grup – osób raportujących wystąpienie każdego z objawów diagnostycznych mizofonii wg kryteriów Schrödera i in. (2013) oraz osób niespełniających tych kryteriów (d Cohena = 2,13).

Jak wspomniano wyżej, *MisoQuest* został stworzony na podstawie zmodyfikowanych kryteriów diagnostycznych Schrödera i in. (2013). Oryginalne kryteria obejmują obecność awersyjnych reakcji na dźwięki wydawane tylko przez ludzi, podczas gdy *MisoQuest* uwzględnia obecność wszystkich rodzajów dźwięków jako bodźców mizofonicznych zarówno ludzkich, jak i niepocho- dzących od ludzi. Pozycja dotycząca dźwięków wydawanych przez ludzi miała lepsze właściwości psychometryczne od pozycji dotyczącej negatywnych reakcji na odgłosy wydawane przez przedmioty pod względem: (1) wyższego ładunku czynnikowego, (2) wyższej korelacji między skalą a pozycją, (3) wyższej stabilności w czasie oraz (4) lepszej dyskryminacji między osobami o różnym stopniu mizofonii (wyższy parametr dyskryminacyjny w analizie teorii odpowiadania na pozycje testowe; *item response theory* [IRT]). Jednak obie pozycje znalazły się w finalnej wersji *MisoQuest*, jako że coraz więcej badań wskazywało ówczesznie na to, że bodźce mizofoniczne, choć najczęściej są wydawane przez ludzi, mogą być bardziej różnorodne. Druga modyfikacja miała miejsce w przypadku określenia zakresu emocji doświadczanych w związku z bodźcami mizofonicznymi. W kryteriach Schrödera i in. (2013) wskazano, że do diagnozy mizofonii konieczne jest występowanie złości. W puli pozycji *MisoQuest* kilka z nich odnosi się bezpośrednio do doświadczania złości. Kilka innych pozycji dotyczy doświadczania innych specyficznych emocji oraz emocji ogólnie. W przypadku pozycji dotyczących natychmiastowości reakcji, ta, która odnosi się do doświadczania złości, miała lepsze właściwości psychometryczne niż pozycja odnosząca się do nieprzyjemnych emocji ogólnie. Dlatego też ta druga nie została włączona do finalnej wersji narzędzia. Inaczej sytuacja wyglądała w przypadku pozycji nieodnoszących się do natychmiastowości doznawania reakcji mizofonicznej oraz pozycji dotyczącej antycypacji wystąpienia bodźca mizofonicznego. W tym przypadku pozycje odnoszące się do doświadczania innych emocji niż złość lub odnoszące się do nieprzyjemnych emocji bez

określenia ich nazwy miały podobnie dobre wartości psychometryczne. Z tego względu wszystkie zostały włączone do finalnej wersji narzędzia. *MisoQuest* wyraźnie uwypukla więc znaczenie odgłosów wydawanych przez ciało ludzkie oraz emocję złości. Może to wpływać na ograniczenie liczby osób diagnozowanych jako mizofonicy przy użyciu tego narzędzia w porównaniu do opisanych niżej *S-Five* (Vitoratou i in., 2021) czy wyżej – MQ (Wu i in., 2014), co zostało zaobserwowane m.in. w badaniach we Francji (Enzler, Lorient i in., 2021). Innymi słowy, osoby niedoświadczające lub nieraportujące złości podczas ekspozycji na odgłosy wydawane przez ludzi z niższym prawdopodobieństwem osiągną punkt odcięcia dla mizofonii (Siepsiak, Sobczak i in., 2020), choć nie jest to wykluczone przy wysokim nasileniu innych objawów.

Kolejnym ważnym, ale i kontrowersyjnym, aspektem mizofonii, który występuje w kryteriach zarówno Schrödera i in. (2013), jak i Jager i in. (2020) jest występowanie zachowań agresywnych. W analizie konstrukcyjnej *MisoQuest* pozycja dotycząca zachowań agresywnych nisko łąduje czynnik ogólny kwestionariusza, ma bardzo niską korelację skala-pozycja, niską stabilność w czasie i bardzo niską wartość dyskryminacyjną w IRT. Ponadto uczestnicy badań walidacyjnych w komentarzach po badaniu wyrażali zdezorientowanie obecnością takiego zagadnienia. Wobec powyższych, a także w związku z brakiem wystarczających danych w literaturze na temat obecności zachowań agresywnych w mizofonii, podjęto decyzję o niewłączaniu pozycji dotyczącej zachowań agresywnych do kwestionariusza. Decyzja ta wydaje się właściwa w świetle najnowszych badań, wskazujących na to, że agresja nie jest charakterystycznym objawem tego zaburzenia (Vitoratou i in., 2021), przynajmniej nie u osób dorosłych. W diagnozie klinicznej warto przy tym pamiętać, że osoby z mizofonią, mówiąc o agresji, często opisują w ten sposób myśli o treści agresywnej, mogące pojawiać się podczas ekspozycji na bodźce mizofoniczne, np. myśli o tym, że komuś zrobią krzywdę, które nie mają odzwierciedlenia w zachowaniu.

Cechą charakterystyczną *MisoQuest* jest również podkreślenie natychmiastowości reakcji (jedna z pozycji odnosi się do natychmiastowości reakcji, a kolejne dwie dotyczą doświadczania emocji złości w konsekwencji antycypacji ekspozycji na bodźce mizofoniczne), co jest zgodne z kryteriami Schrödera i in. (2013). Z jednej strony jest to zaletą tego narzędzia, ponieważ umożliwia odsianie osób, które zaczynają doświadczać trudnych emocji dopiero po dłuższym czasie od nieustawiania dźwięku, co z kolei może być związane z innymi zaburzeniami tudzież wynikać ze struktury temperamentu (Strelau i Zawadzki, 1995). Z drugiej strony z tego powodu *MisoQuest* może wskazywać

falszywie negatywne wyniki u osób, które z różnych powodów (np. wysokich umiejętności radzenia sobie, przebytej terapii czy aktualnie sprzyjającego środowiska) reagują na odgłosy mizofoniczne łagodniej, mają obecnie niskie nasilenie objawów mizofonii czy są w stanie znieść pobudzenie psychofizjologiczne dłużej. Należy przy tym zaznaczyć, że zgodnie z konsensusem (Swedo i in., 2022) odgłosy mizofoniczne często charakteryzują się powtarzalnością, co teoretycznie może oznaczać, że wiele osób z mizofonią może doświadczać reakcji mizofonicznej dopiero po dłuższym czasie od nieustawiania odgłosu. Ani w konsensusie (Swedo i in., 2022), ani w innych źródłach nie określono precyzyjnie, jak szybka powinna być reakcja mizofoniczna. Ponadto w konsensusie zaznaczono również, że „kiedy tylko bodziec jest dostrzeżony, osoba z mizofonią może mieć trudność z odwróceniem od niego uwagi [...]” (Swedo i in., 2022, s. 10), co mogłoby sugerować natychmiastowość reakcji mizofonicznej. Trudno zatem określić, jak dalece pod tym względem *MisoQuest* jest zgodny z konsensusem. W literaturze brakuje nadal precyzyjnych danych dotyczących natychmiastowości/odroczenia pojawiającej się reakcji emocjonalnej. *MisoQuest* może zatem okazać się użyteczny w badaniach nad znaczeniem i związkami szybkości reagowania na odgłosy mizofoniczne. Niewykluczone jednak, że *MisoQuest* będzie musiał być w przyszłości zmodyfikowany pod tym kątem.

Kolejnym aspektem mizofonii, który należy wziąć pod uwagę, diagnozując to zaburzenie, jest stopień świadomości doświadczanych reakcji mizofonicznych. Zgodnie z konsensusem tylko „niektóre osoby z mizofonią” (Swedo i in., 2022, s. 10) są świadome nadmiarowości swoich reakcji na dźwięki. W *MisoQuest* zakłada się natomiast, że jednym ze wskaźników mizofonii jest świadomość nieadekwatności swoich reakcji mizofonicznych (*Uważam, że moje reakcje na dźwięki są wyolbrzymione*). Należy jednak pamiętać, że w momencie tworzenia tego kwestionariusza dostępne dane wskazywały na taką charakterystykę mizofonii. W pracy klinicznej, w szczególności w sytuacji wysokiego nasilenia objawów, ale nie na tyle, by osiągnąć punkt odcięcia, warto mieć na uwadze to, iż obecne rozumienie mizofonii nie obejmuje konieczności, by osoba z tym zaburzeniem określała swoje reakcje jako wyolbrzymione.

Choć w badaniach walidacyjnych *MisoQuest* nie ustalono granicy klinicznego nasilenia objawów mizofonii, to w późniejszych projektach prowadzonych na Wydziale Psychologii Uniwersytetu Warszawskiego przyjęto wstępnie granicę 61 punktów z 71 jako tymczasowy, umowny punkt odcięcia dla objawów mizofonii o znaczeniu klinicznym. Kierowano się różnicą pomiędzy średnią ($M = 65,72$) a różnicą jednego odchylenia standardowego ($SD = 4,3$) w wynikach

MisoQuest osób z mizofonią w badaniach walidacyjnych (Siepsiak, Śliwerski i Dragan, 2020). Taki próg przyjęto również w badaniach na grupie pacjentów oddziałów całodobowych z diagnozą depresji, gdzie porównano wyniki *MisoQuest* z wynikami diagnoz przeprowadzonych metodą stacjonarnego wywiadu indywidualnego, na podstawie kryteriów Schrödera i in. (2013). Badania te wykazały bardzo wysoką specyficzność *MisoQuest* (96,34%) przy dość niskiej czułości. Oznacza to, że jest mało prawdopodobne uzyskanie wyniku fałszywie pozytywnego, ale istnieje duże ryzyko wyniku fałszywie negatywnego, czyli pominięcia osoby, która w wywiadzie okazałaby się mieć klinicznie istotne nasilenie objawów mizofonii. Podobne wyniki uzyskali Enzler, Lorient i in. (2021), którzy badali osoby ze społeczności internetowych, gdzie *MisoQuest* okazał się mieć bardzo wysoki stopień specyficzności (99%), ale znacznie niższy stopień czułości (45%), w stosunku do samodiagnoz. Dane z powyższych badań wskazują na konieczność ostrożnego interpretowania wyników negatywnych, w szczególności w praktyce klinicznej. Wskazują one również na konieczność dalszych badań nad określeniem, czym charakteryzują się osoby, które nie mieszczą się w kryteriach diagnostycznych Schrödera i in. (2013), nawet po ich poszerzeniu – zidentyfikowane na podstawie *MisoQuest*, ale nie należące też do osób bez nadwrażliwości dźwiękowych.

MisoQuest jest wykorzystywany w badaniach naukowych w Polsce i w innych krajach (np. Bagrowska i in., 2022; Enzler, Lorient i in., 2021; Heller i Smith, 2022; Remmert, Schmidt i in., 2022; Savard i in., 2022). Dostępna jest między innymi adaptacja w języku angielskim, niemieckim, tureckim, francuskim, perskim, portugalskim (Brazylia), indonezyjskim i japońskim. Badania walidacyjne wersji angielskiej w Kanadzie (Raymond, 2023), Wielkiej Brytanii (Kula i in., 2024), Brazylii (Silva i in., 2024) i Turcji (Ay i in., 2024) potwierdziły bardzo dobre właściwości psychometryczne tego narzędzia (choć w Turcji *MisoQuest* okazał się mieć nieco inną – 3-czynnikową – strukturę).

Tworzenie, charakterystyka i znaczenie S-Five

S-Five (Uglik-Marucha i in., 2024; Vitoratou i in., 2021), czyli 5 „S” (*Selective Sound Sensitivity Syndrome Scale*; *Skala selektywnej nadwrażliwości na dźwięki*) jest wielowymiarowym narzędziem do identyfikacji i pomiaru nasilenia mizofonii. Jest to obecnie jedyne na świecie wielowymiarowe narzędzie tego rodzaju, które jest zwalidowane międzykulturowo. Umożliwia to między in-

nymi porównywanie danych dotyczących wielu różnych obszarów mizofonii z różnych, również pozaeuropejskich, krajów. *S-Five* zostało przetłumaczone i zwalidowane w Wielkiej Brytanii (Vitoratou i in., 2023), Polsce (Uglik-Marucha i in., 2024), Niemczech (Remmert, Schmidt i in., 2022), Chinach (Vitoratou i in., 2022) oraz Portugalii (Hayes i in., 2024). Toczą się również prace walidacyjne we Włoszech, Turcji, Grecji, Indiach i Singapurze.

Skala *S-Five* przeszła proces iteracyjnej konstrukcji, który obejmował cztery fale pobierania próbek danych w populacji anglojęzycznej doświadczającej symptomów mizofonii. Ponad trzy tysiące odpowiedzi na 118 pozycji zostało przeanalizowanych przez zespół badawczy składający się zarówno z ekspertów klinicznych, jak i psychometrów. Pierwszy etap obejmował opracowanie puli 50 wstępnych pozycji opartych na literaturze dotyczącej mizofonii. Zostały one następnie ocenione przez 853 osoby, które oprócz wypełnienia ankiety udzieliły również szczegółowych informacji zwrotnych na temat tego, jak je rozumieją i co o nich myślą. Na tym etapie do zespołu dołączył psycholog kliniczny z rozległym doświadczeniem w leczeniu pacjentów z mizofonią do pomocy w jakościowej analizie danych. W kolejnym kroku, na podstawie analizy jakościowej i ilościowej, pula stwierdzeń została zmodyfikowana i rozszerzona do 118 pozycji. Zidentyfikowano wtedy 12 wymiarów mizofonii: uwaga i skupienie, strategie radzenia sobie, eksternalizacja, wpływ na życie, wgląd, internalizacja, interpretacja, moderatory dotyczące kontekstu, wybuchy emocji, reakcje fizjologiczne, obawy przed konsekwencjami społecznymi i zagrożenie.

W drugim etapie tworzenia kwestionariusza zebrano dane od 815 osób. Eksploracyjna analiza czynnikowa ujawniła pięcioczynnikową strukturę kwestionariusza (internalizacja, eksternalizacja, wpływ na życie, wybuchy emocji, zagrożenie), wykazującą silną spójność wewnętrzną w obrębie każdej podskali. W wyniku dalszych analiz usunięto pozycje związane z uwagą i skupieniem, strategiami radzenia sobie, wglądem, interpretacją, modulatorami, reakcją fizjologiczną i obawami przed konsekwencjami społecznymi. W rezultacie 68 pozycji zostało poddanych testom w trzeciej fali pobierania próbek. Skala odpowiedzi została również zmieniona na skalę 0–10, aby umożliwić bardziej szczegółową ocenę nasilenia doznań mizofonicznych.

W trzecim etapie testów zrewidowany kwestionariusz został wypełniony przez 393 uczestników. Analiza czynnikowa wykazała taką samą pięcioczynnikową strukturę, jak stwierdzoną w drugim etapie konstrukcji narzędzia. Wykorzystując ściśle kryteria psychometryczne oraz dopasowanie modelu do danych, kwestionariusz skrócono do 25 pozycji.

Ostatnim etapem tworzenia narzędzia była pełna walidacja pięcioczynnikowego modelu w próbie 828 osób. Zakładana struktura została potwierdzona. Skala *S-Five* wykazała bardzo dobre własności psychometryczne pod względem stabilności w czasie ($\psi = 0,83\text{--}0,89$; $\text{ICC} = 0,87\text{--}0,90$), zgodności wewnętrznej (α dla skal = $0,83\text{--}0,88$; α dla skali ogólnej = $0,90$) i trafności treściowej ($r = 0,50\text{--}0,70$ z A-MISO-S; *Amsterdam Misophonia Scale*). Nie stwierdzono ponadto tendencji pomiaru kwestionariusza w odniesieniu do tożsamości płciowej i wieku.

W 2022 i 2023 roku kwestionariusz *S-Five* został przetłumaczony na język polski z wykorzystaniem metody tłumaczenia zwrotnego we współpracy z autorami oryginalnej wersji kwestionariusza. Badania walidacyjne w grupie 288 osób z i bez mizofonii potwierdziły pięcioczynnikową strukturę narzędzia oraz wskazały na ich bardzo dobrą stabilność, tzn. odtwarzalność pięcioczynnikowej struktury w innych próbach (Uglik-Marucha i in., 2024). Wersja polska narzędzia charakteryzuje się również bardzo dobrymi właściwościami psychometrycznymi, takimi jak stabilność w czasie ($\psi = 0,76\text{--}0,86$; $\text{ICC} = 0,84\text{--}0,88$), zgodność wewnętrzna (ω dla skal = $0,82\text{--}0,92$; ω dla skali ogólnej = $0,92$) i trafność teoretyczna ($r = 0,70$ z *MisoQuest*; $r = 0,53$ z *MQ*). W polskiej wersji *S-Five* nie ustalono jeszcze punktu odcięcia dla klinicznego nasilenia mizofonii. Wstępnych danych na ten temat dostarczyły jednak niedawno badania na populacji ogólnej w Wielkiej Brytanii (Vitoratou i in., 2023). Na podstawie analizy krzywej ROC (*receiver operating characteristic*) oraz porównania wyników *S-Five* z wywiadami diagnostycznymi określono punkt odcięcia dla mizofonii na 87 punktów (z 250), gdzie narzędzie wykazywało najbardziej zrównoważoną czułość (84%) i specyficzność (72%).

W finalnej wersji narzędzie składa się z 25 pozycji odnoszących się do różnych doświadczeń związanych z mizofonią, z których każde określane jest na 10-stopniowej skali interwałowej (gdzie 0 oznacza *zupełnie nie zgadzam się*, a 10 – *całkowicie zgadzam się*). Tworzą one pięć skal, odpowiadających pięciu aspektom mizofonii: internalizacja (*Sposób, w jaki reaguję na pewne dźwięki, sprawia, że zastanawiam się, czy w głębi duszy jestem złą osobą*), eksternalizacja (*Ludzie nie powinni wydawać niektórych dźwięków, nawet jeśli nie zdają sobie sprawy z nadwrażliwości innych osób na te dźwięki*), wpływ na życie (*Moje możliwości zatrudnienia są ograniczone ze względu na moje reakcje na niektóre dźwięki*), wybuchy emocji (*Obawiam się, że mogę zacząć zachowywać się agresywnie i gwałtownie, gdy nie mogę znieść niektórych dźwięków, które ktoś wydaje*), zagrożenie (*Niektóre dźwięki stanowią dla mnie duży dyskomfort/są dla mnie udręką*).

Dwie pierwsze skale opisują sposób oceny i atrybucji reakcji mizofonicznych i źródła dźwięku. Wysokie wyniki w skali eksternalizacji wskazują bardziej na szukanie przyczyny reakcji w innych ludziach (np. związanych z oceną ludzi wydających awersyjne dźwięki jako bezmyślnych, źle wychowanych czy niemających szacunku wobec innych). Natomiast wysokie wyniki w skali internalizacji wskazują na obwinianie i osądzanie samego siebie za pojawianie się reakcji mizofonicznych, wyrażające się np. w poczuciu, że jest się złą osobą czy np. w obniżonym szacunku do samego siebie przez to, w jaki sposób reaguje się na dźwięki mizofoniczne. Skala zagrożenia odnosi się bezpośrednio do reakcji i postrzegania momentu ekspozycji na dźwięki mizofoniczne, doświadczanych emocji lęku, bezradności, poczucia osaczenia oraz obawy przed potencjalnym wybuchem emocji. Skala wybuchów emocji jest natomiast związana z faktycznymi wybuchami agresji w odpowiedzi na bodźce mizofoniczne, choć jedna pozycja z tej skali opisuje tylko obawy o potencjalny wybuch, dlatego ważna jest, w szczególności w kontekście klinicznym, jakościowa analiza odpowiedzi pacjenta/klienta. Ostatnia skala – wpływ na życie, jak nazwa wskazuje, odnosi się do stopnia, w jakim objawy mizofoniczne dezorganizują codzienne funkcjonowanie danej osoby, ale również do obaw o przyszłość, w kontekście reakcji na pewne dźwięki.

S-Five zawiera również opcjonalną listę 37 bodźców awersyjnych (*S-Five-T*), które są oceniane na dwóch skalach: nominalnej, która prezentuje dziewięć emocji, które mogą być wywoływane przez dany bodziec, oraz interwałowej, która ocenia stopień nasilenia danej emocji w odpowiedzi na dany bodziec mizofoniczny (na skali od 0 do 10). Przykładowo pacjent/klient jest proszony o wskazanie jednej z dziewięciu podanych emocji lub reakcji fizjologicznych (brak reakcji, irytacja, rozpacz, obrzydzenie, złość, panika, inne uczucie: negatywne, inne uczucie: pozytywne, inne: reakcja fizjologiczna), która jest wywoływana przez dany dźwięk, na przykład połykanie/przelykanie. Następnie pacjent/klient ocenia intensywność wybranej emocji (np. obrzydzenie) na ten sam dźwięk. Jak podają autorzy (Vitoratou i in., 2023), lista bodźców awersyjnych może być modyfikowana, skracana lub wydłużana, w zależności od potrzeb. Dzięki tak skonstruowanej skali możliwe jest obliczenie 4 wskaźników: (1) ilość bodźców awersyjnych (TC; *trigger count*), (2) ilość danej reakcji emocjonalnej w odpowiedzi na bodziec (RC; *reaction count*), (3) sumę intensywności reakcji w odpowiedzi na wszystkie wybrane bodźce awersyjne (FIRS; *frequency/intensity of reaction scores*), (4) uśrednioną intensywność reakcji w odpowiedzi na bodziec (RIRS; *relative intensity of reaction scores*). Pięć skal *S-Five* i wskaźniki bodźców awersyjnych są spójne z literaturą dotyczącą mizofonii.

S-Five jest w stanie wychwycić nasilenie mizofonii na poziomie zaburzenia, zgodnie z konsensusem Swedo i in. (2022). Różni się to od zjawiska zmniejszonej tolerancji na określone dźwięki, które może, ale nie musi, skutkować poziomem dystresu niezbędnym do rozpoznania zaburzenia. Ponadto *S-Five* wykazuje znaczną spójność w odniesieniu do 10 domen wywodzących się bezpośrednio z definicji konsensusu (ogólny opis, bodźce wyzwalające, reakcje emocjonalne, reakcje fizjologiczne, reakcje behawioralne, reakcje skupienia uwagi, wpływ na reakcje, spostrzeżenia i świadomość związane z reakcjami, upośledzenie czynnościowe, początek i charakter zaburzenia). Niemniej część domen mizofonii zawartych w konsensusie (Swedo i in., 2022), choć zostały uwzględnione na pierwszych etapach konstrukcji *S-Five* (tj. uwaga i skupienie, strategie radzenia sobie, wgląd, interpretacja, modulatory, reakcje fizjologiczne, obawy przed konsekwencjami społecznymi), miały niskie wartości psychometryczne i nie zostały włączone do kwestionariusza. Dane pochodzące z badań walidacyjnych *S-Five* wstępnie wskazują więc, że choć te aspekty mizofonii są istotne, mogą być mniej lub niebezpośrednio związane z mizofonią. Problem ten wymaga jednak dalszych badań.

Dzięki swojej wielowymiarowości *S-Five* jest szczególnie przydatna w obszarze pracy klinicznej, np. do ewaluacji efektywności oddziaływań psychotherapeutycznych. Kwestionariusz wychwytuje poznawcze aspekty mizofonii, które mogą być celem modyfikacji w psychoterapii poznawczo-behawioralnej, takie jak tendencja do negatywnego oceniania siebie, w świetle reakcji mizofonicznych (skala internalizacja) oraz przewidywania katastrofalnych konsekwencji, w przypadku braku możliwości ucieczki od dźwięków, takich jak poczucie uwięzienia lub utrata kontroli (skala zagrożenia).

Objawy eksternalizacyjne lub procesy myślowe dotyczące atrybucji zewnętrznych, takie jak zwracanie komuś uwagi czy opisywanie bardzo negatywnych, czy wręcz brutalnych myśli o osobach wydających dźwięki mizofoniczne zdają się być dość powszechnie opisywane w internetowych grupach wsparcia. Opisy różnych, mało konstruktywnych, sposobów wyrażania złości w odpowiedzi na bodźce mizofoniczne są też obecne w doniesieniach medialnych. Kształtuje to poniekąd niepełny, zniekształcony i bardzo negatywnie nacechowany obraz mizofonii w opinii publicznej oraz wśród samych osób z mizofonią i ich rodzin. W tym kontekście niezwykle ważnym i nowym obszarem w badaniach mizofonii jest zwrócenie uwagi na istotę internalizacji w tym zaburzeniu. Badania przy pomocy *S-Five* wykazały, że to właśnie wysokie nasilenie na skali internalizacji, nie eksternalizacji, jest związane z wyższym nasileniem

objawów lęku i depresji. Co istotne, to również wyniki w skali internalizacji, nie eksternalizacji, są najwyżżej skorelowane (pozytywnie) z wpływem na życie oraz – mniej intuicyjnie – z wybuchami emocji (Remmert, Jebens i in., 2022; Vitoratou i in., 2021, 2022, 2023). Są to więc również dane, które mogą mieć bezpośrednie znaczenie kliniczne.

***MisoQuest* i *S-Five* w badaniach naukowych i praktyce klinicznej – PODSUMOWANIE**

Zarówno *MisoQuest* (Siepsiak, Sobczak i in., 2020; Siepsiak, Śliwerski i Dragan, 2020), jak i *S-Five* (Uglik-Marucha i in., 2024; Vitoratou i in., 2021) są narzędziami o bardzo dobrych właściwościach psychometrycznych, określonych w badaniach na grupach polskojęzycznych. Oba narzędzia nadają się do wykorzystywania zarówno w praktyce klinicznej, jak i w badaniach naukowych. Są to kwestionariusze w dużej mierze komplementarne, lecz co istotne – nietożsame.

MisoQuest jest krótką (14 pozycji, wypełnienie zajmuje około 4 minuty), pojedynczą skalą, skonstruowaną w celu szybkiego przesiewu i identyfikacji osób, które z bardzo dużym prawdopodobieństwem mają mizofonię. Nie został on stworzony z myślą o diagnozowaniu stopnia nasilenia mizofonii u osób z mizofonią ani stopnia nasilenia konkretnych objawów mizofonii u osób bez klinicznego nasilenia mizofonii – choć z braku innych narzędzi wielokrotnie był wykorzystywany w tym celu, dlatego jego użyteczność w tym aspekcie powinna jeszcze zostać zweryfikowana. *S-Five* natomiast jest narzędziem rozbudowanym, a więc wymagającym znacznie więcej czasu na jego wykonanie (około 10 minut bez opcjonalnej części *S-Five-T*, która zajmuje dodatkowe 10 minut). Kwestionariusz ten zawiera pięć skal odpowiadających pięciu aspektom mizofonii, wraz z dodatkowymi, ale opcjonalnymi, wskaźnikami pozwalającymi na ocenę konkretnych emocji i ich intensywności w odpowiedzi na konkretne bodźce dźwiękowe. Jest to zatem narzędzie pozwalające nie tylko na szczegółowe określenie związków specyficznych aspektów mizofonii z innymi zmiennymi w badaniach naukowych, ale jest też przydatny w praktyce klinicznej, np. do monitorowania postępów w terapii czy diagnozy obszarów funkcjonowania. Wydaje się, że w szczególności w praktyce klinicznej badanie powyższymi kwestionariuszami warto połączyć z wywiadem diagnostycznym *Duke Misophonia Interview* (Guetta i in., 2022), który jest już dostępny w języku polskim.

Ważną różnicą między tymi dwoma narzędziami jest zakres definicji mizofonii, na podstawie której zostały skonstruowane. *S-Five* jest narzędziem o szerokim zakresie definiowania mizofonii, bez faworyzowania określonych bodźców czy emocji. *MisoQuest* natomiast, choć dopuszcza uwzględnienie osób z różnorodnymi bodźcami mizofonicznymi i emocjami, podkreśla znaczenie emocji złości oraz odgłosów wydawanych przez usta człowieka. W konsekwencji, używając tych dwóch kwestionariuszy, opierając się na diagnozie *MisoQuest*, znacznie mniej osób zostanie zidentyfikowanych jako osoby z mizofonią niż w przypadku oparcia diagnozy o *S-Five*.

Narzędzia te na pewno mogą być więc wykorzystywane razem lub osobno, w zależności od konkretnych potrzeb badacza czy klinicysty. Co istotne, ani *MisoQuest*, ani *S-Five* nie służą do samodzielnej diagnozy mizofonii i, jak w przypadku każdego innego kwestionariusza, nigdy nie powinny być stosowane jako jedyny sposób diagnozy zaburzenia u konkretnej osoby. Co ważne, nie są znane jeszcze polskie normy dla objawów nasilenia mizofonii, a wspomniane punkty odcięcia w przypadku obu narzędzi są obecnie jedynie sugestią, wynikającą ze wstępnych analiz – nie powinny być traktowane kategorycznie, w szczególności w praktyce klinicznej.

BIBLIOGRAFIA

- Ay, E., Huviyetli, M., Çakmak, E. (2024). The mediating role of anxiety in the relationship between misophonia and quality of life: Findings from the validated Turkish version of MisoQuest. *Frontiers in Psychology*, 15, artykuł 1361645. <https://doi.org/10.3389/fpsyg.2024.1361645>
- Bagrowska, P., Pionke-Ubych, R., Gawęda, Ł. (2022). Do they make these sounds to hurt me? The mediating role of emotion regulation, anxiety and hostile attributions in the relationship between misophonia and paranoia-like thoughts. *Schizophrenia Research*, 250, 137–142. <https://doi.org/10.1016/j.schres.2022.11.005>
- Brout, J. J., Edelstein, M., Erfanian, M., Mannino, M., Miller, L. J., Rouw, R., Kumar, S., Rosenthal, M. Z. (2018). Investigating misophonia: A review of the empirical literature, clinical implications, and a research agenda. *Frontiers in Neuroscience*, 12, artykuł 36. <https://doi.org/10.3389/fnins.2018.00036>
- Dibb, B., Golding, S. E., Dozier, T. H. (2021). The development and validation of the Misophonia Response Scale. *Journal of Psychosomatic Research*, 149, artykuł 110587. <https://doi.org/10.1016/j.jpsychores.2021.110587>

- Edelstein, M., Brang, D., Rouw, R., Ramachandran, V. S. (2013). Misophonia: Physiological investigations and case descriptions. *Frontiers in Human Neuroscience*, 7, artykuł 296. <https://doi.org/10.3389/fnhum.2013.00296>
- Eijssker, N., Schröder, A., Liebrand, L. C., Smit, D. J. A., van Wingen, G., Denys, D. (2021). White matter abnormalities in misophonia. *NeuroImage: Clinical*, 32, artykuł 102787. <https://doi.org/10.1016/j.nicl.2021.102787>
- Enzler, F., Fournier, P., Noreña, A. J. (2021). A psychoacoustic test for diagnosing hyperacusis based on ratings of natural sounds. *Hearing Research*, 400, artykuł 108124. <https://doi.org/10.1016/j.heares.2020.108124>
- Enzler, F., Lorient, C., Fournier, P., Noreña, A. J. (2021). A psychoacoustic test for misophonia assessment. *Scientific Reports*, 11, artykuł 11044. <https://doi.org/10.1038/s41598-021-90355-8>
- Erfanian, M., Kartsonaki, C., Keshavarz, A. (2019). Misophonia and comorbid psychiatric symptoms: A preliminary study of clinical findings. *Nordic Journal of Psychiatry*, 73(4–5), 219–228. <https://doi.org/10.1080/08039488.2019.1609086>
- Guetta, R. E., Cassiello-Robbins, C., Anand, D., Rosenthal, M. Z. (2022). Development and psychometric exploration of a semi-structured clinical interview for misophonia. *Personality and Individual Differences*, 187, artykuł 111416. <https://doi.org/10.1016/j.paid.2021.111416>
- Hayes, C., Gregory, J., Aziz, R., Cerejeira, J., Cruz, M., Simões, J.A., Vitoratou, S. (2024). Psychometric evaluation and misophonic experience in a Portuguese-speaking sample. *Behavioral Sciences*, 14(2), artykuł 107. <https://doi.org/10.3390/bs14020107>
- Heller, L. M., Smith, J. M. (2022). Identification of everyday sounds affects their pleasantness. *Frontiers in Psychology*, 13, artykuł 894034. <https://doi.org/10.3389/fpsyg.2022.894034>
- Jager, I., de Koning, P., Bost, T., Denys, D., Vulink, N. (2020). Misophonia: Phenomenology, comorbidity and demographics in a large sample. *PLoS ONE*, 15(4), artykuł e0231390. <https://doi.org/10.1371/journal.pone.0231390>
- Jastreboff, M. M., Jastreboff, P. J. (b.d.). *Components of decreased sound tolerance: Hyperacusis, misophonia, phonophobia*. Pobrane 6 września 2024 z: https://www.tinnitus.org/DST_NL2_PJM.pdf
- Jastreboff, P. J., Jastreboff, M. M. (2000). Tinnitus Retraining Therapy (TRT) as a method for treatment of tinnitus and hyperacusis patients. *Journal of the American Academy of Audiology*, 11(3), 162–177.
- Jastreboff, P. J., Jastreboff, M. M. (2014). Treatments for decreased sound tolerance (hyperacusis and misophonia). *Seminars in Hearing*, 35(02), 105–120. <https://doi.org/10.1055/s-0034-1372527>

- Kula, F. B., Cropley, M., Aazh, H. (2024). *Evaluating the validity and reliability of the English version of the MisoQuest (a Misophonia Questionnaire) and its relationships with measures of hyperacusis, tinnitus, hearing impairment, sleep, fatigue and mental health* [maszynopis wysłany do publikacji]. University of Surrey, United Kingdom.
- Kumar, S., Tansley-Hancock, O., Sedley, W., Winston, J. S., Callaghan, M. F., Allen, M., Cope, T. E., Gander, P. E., Bamiou, D. E., Griffiths, T. D. (2017). The brain basis for misophonia. *Current Biology*, 27(4), 527–533. <https://doi.org/10.1016/j.cub.2016.12.048>
- Neacsiu, A. D., Szymkiewicz, V., Galla, J. T., Li, B., Kulkarni, Y., Spector, C. W. (2022). The neurobiology of misophonia and implications for novel, neuroscience-driven interventions. *Frontiers in Neuroscience*, 16, artykuł 893903. <https://doi.org/10.3389/fnins.2022.893903>
- Raymond, K. (2023). *Assessing misophonia: Cross-cultural validation of the MisoQuest and its relation with cognitive functioning* [praca magisterska, Western University]. Electronic Thesis and Dissertation Repository. <https://ir.lib.uwo.ca/etd/9407>
- Remmert, N., Jebens, A., Gruzman, R., Gregory, J., Vitoratou, S. (2022). A nomological network for misophonia in two German samples using the S-Five model for misophonia. *Frontiers in Psychology*, 13, artykuł 902807. <https://doi.org/10.3389/fpsyg.2022.902807>
- Remmert, N., Schmidt, K. M. B., Mussel, P., Hagel, M. L., Eid, M. (2022). The Berlin Misophonia Questionnaire Revised (BMQ-R): Development and validation of a symptom-oriented diagnostical instrument for the measurement of misophonia. *PLoS ONE*, 17(6), artykuł e0269428. <https://doi.org/10.1371/journal.pone.0269428>
- Rosenthal, M. Z., Anand, D., Cassiello-Robbins, C., Williams, Z. J., Guetta, R. E., Trumbull, J., Kelley, L. D. (2021). Development and initial validation of the Duke Misophonia Questionnaire. *Frontiers in Psychology*, 12, artykuł 709928. <https://doi.org/10.3389/fpsyg.2021.709928>
- Rosenthal, M. Z., McMahon, K., Greenleaf, A. S., Cassiello-Robbins, C., Guetta, R., Trumbull, J., Anand, D., Frazer-Abel, E. S., Kelley, L. (2022). Phenotyping misophonia: Psychiatric disorders and medical health correlates. *Frontiers in Psychology*, 13, artykuł 941898. <https://doi.org/10.3389/fpsyg.2022.941898>
- Samermitt, P., Young, M., Allen, A. K., Trillo, H., Shankar, S., Klein, A., Kay, C., Mahzouni, G., Reddy, V., Hamilton, V., Davidenko, N. (2022). Development and evaluation of a sound-swapped video database for misophonia. *Frontiers in Psychology*, 13, artykuł 890829. <https://doi.org/10.3389/fpsyg.2022.890829>
- Savard, M.-A., Sares, A. G., Coffey, E. B. J., Deroche, M. L. D. (2022). Specificity of affective responses in misophonia depends on trigger identification. *Frontiers in Neuroscience*, 16, artykuł 879583. <https://doi.org/10.3389/fnins.2022.879583>

- Schröder, A., Vulink, N., Denys, D. (2013). Misophonia: Diagnostic criteria for a new psychiatric disorder. *PLoS ONE*, 8(1), artykuł e54706. <https://doi.org/10.1371/journal.pone.0054706>
- Schröder, A., Wingen, G. van, Eijsker, N., San Giorgi, R., Vulink, N. C., Turbyne, C., Denys, D. (2019). Misophonia is associated with altered brain activity in the auditory cortex and salience network. *Scientific Reports*, 9(1), artykuł 7542. <https://doi.org/10.1038/s41598-019-44084-8>
- Siepsiak, M., Dragan, W. (2019). Mizofonia – przegląd badań i głównych koncepcji teoretycznych. *Psychiatria Polska*, 53(2), artykuł 447. <https://doi.org/10.12740/PP/92023>
- Siepsiak, M., Rosenthal, M. Z., Raj-Koziak, D., Dragan, W. (2022). Psychiatric and audiologic features of misophonia: Use of a clinical control group with auditory over-responsivity. *Journal of Psychosomatic Research*, 156, artykuł 110777. <https://doi.org/10.1016/j.jpsychores.2022.110777>
- Siepsiak, M., Sobczak, A. M., Bohaterewicz, B., Cichocki, Ł., Dragan, W. Ł. (2020). Prevalence of misophonia and correlates of its symptoms among inpatients with depression. *International Journal of Environmental Research and Public Health*, 17(15), artykuł 5464. <https://doi.org/10.3390/ijerph17155464>
- Siepsiak, M., Śliwerski, A., Dragan, W. (2020). Development and psychometric properties of MisoQuest – A new self-report questionnaire for misophonia. *International Journal of Environmental Research and Public Health*, 17(5), artykuł 1797. <https://doi.org/10.3390/ijerph17051797>
- Siepsiak, M., Vrana, S. R., Rynkiewicz, A., Rosenthal, M. Z., Dragan, W. Ł. (2023). Does context matter in misophonia? A multi-method experimental investigation. *Frontiers in Neuroscience*, 16, artykuł 880853. <https://doi.org/10.3389/fnins.2022.880853>
- Silva, L. de A., Noll, M., Siqueira, G. C., Barbosa, A. K. N. (2024). Assessing misophonia in young adults: The prevalence and psychometric validation of the MisoQuest Questionnaire. *Healthcare*, 12(18), artykuł 1888. <https://doi.org/10.3390/healthcare12181888>
- Strelau, J., Zawadzki, B. (1995). The Formal Characteristics of Behaviour—Temperament Inventory (FCB—TI): Validity studies. *European Journal of Personality*, 9(3), 207–229. <https://doi.org/10.1002/per.2410090304>
- Swedo, S. E., Baguley, D. M., Denys, D., Dixon, L. J., Erfanian, M., Fioretti, A., Jastreboff, P. J., Kumar, S., Rosenthal, M. Z., Rouw, R., Schiller, D., Simner, J., Storch, E. A., Taylor, S., Vander Werff, K. R., Altimus, C. M., Raver, S. M. (2022). Consensus definition of misophonia: A Delphi study. *Frontiers in Neuroscience*, 16, artykuł 841816. <https://doi.org/10.3389/fnins.2022.841816>
- Uglik-Marucha, N., Siepsiak, M., Zielińska, J., Dragan, W. Ł., Gregory, J., Vitoratou, S. (2024). Beyond sound irritation: Cross-cultural evidence on the robustness

- of the five aspects of misophonic experience measured by the S-Five in a Polish sample. *Frontiers in Psychology*, 15, artykuł 1372870. <https://doi.org/10.3389/fpsyg.2024.1372870>
- Vitoratou, S., Hayes, C., Uglik-Marucha, N., Pearson, O., Graham, T., Gregory, J. (2023). Misophonia in the UK: Prevalence and norms from the S-Five in a UK representative sample. *PLoS ONE*, 18(3), artykuł e0282777. <https://doi.org/10.1371/journal.pone.0282777>
- Vitoratou, S., Uglik-Marucha, N., Hayes, C., Gregory, J. (2021). Listening to people with misophonia: Exploring the multiple dimensions of sound intolerance using a new psychometric tool, the S-Five, in a large sample of individuals identifying with the condition. *Psych*, 3(4), 639–662. <https://doi.org/10.3390/psych3040041>
- Vitoratou, S., Wang, J., Hayes, C., Wang, Q., Stefanatou, P., Gregory, J. (2022). Evidence of cross-cultural consistency of the S-Five model for misophonia: Psychometric conclusions emerging from the Mandarin version. *Frontiers in Psychology*, 13, artykuł 879881. <https://doi.org/10.3389/fpsyg.2022.879881>
- Wu, M. S., Lewin, A. B., Murphy, T. K., Storch, E. A. (2014). Misophonia: Incidence, phenomenology, and clinical correlates in an undergraduate student sample. *Journal of Clinical Psychology*, 70(10), 994–1007. <https://doi.org/10.1002/jclp.22098>
- YouGov (17 listopada 2022). *Irritating and relaxing sounds*. <https://today.yougov.com/topics/society/articles-reports/2022/11/17/irritating-relaxing-sounds-yougov-polls-oct-2022>

CZĘŚĆ III: PROBLEMY PSYCHOMETRYCZNE POMIARU KLINICZNEGO W ODNIESIENIU DO ZABURZENIA STRESOWEGO POURAZOWEGO (PTSD)

MARIA CYNIAK-CIECIURA  <https://orcid.org/0000-0002-6293-8729>

Centrum Badań Klinicznych i Doskonalenia Psychoterapii,
Instytut Psychologii, Uniwersytet SWPS¹

AGNIESZKA POPIEL  <https://orcid.org/0000-0001-9990-4971>

Centrum Badań Klinicznych i Doskonalenia Psychoterapii,
Instytut Psychologii, Uniwersytet SWPS

PATRYK BIELAK  <https://orcid.org/0000-0002-5364-506X>

Wydział Psychologii, Uniwersytet Warszawski

VICTORIA CAMINO

Universidad de Buenos Aires, Argentina

EUN JUNG CHA

Wydział Psychologii, Chung-Ang University, Korea Południowa

YUNKYUNG CHO

Wydział Psychologii, Chung-Ang University, Korea Południowa

JULIE K. CREMEANS-SMITH

Kent State University, USA

DAVID B. FRUEHSTORFER

Youngstown State University, USA

MARINA GALARREGUI

Facultad de Psicología, Universidad de Buenos Aires, Argentina

ROCÍO GOLDFARB

Facultad de Psicología, Universidad de Buenos Aires, Argentina

MYOUNG HO HYUN

Wydział Psychologii, Chung-Ang University, Korea Południowa

ZHANNA KALININA

School of Psychology, University College Dublin, Irlandia

¹ Poza pierwszymi Autorkami pracy, wszyscy Współpracownicy są wymienieni w porządku alfabetycznym.

Sposób cytowania: Cyniak-Cieciura, M., Popiel, A., Bielak, P., Camino, V., Cha, E. J., Cho, Y., Cremeans-Smith, J. K., Fruehstorfer, D. B., Galarregui, M., Goldfarb, R., Hyun, M. H., Kalinina, Z., Keegan, E., Mambetalina, A., McHugh, L., Miracco, M., Oshio, A., Park, C., Partarrieu, A., [...] Zawadzki, B. (2024). Czy pomiar objawów PTSD za pomocą PTSD Checklist dla DSM-5 (PCL-5) jest równoważny kulturowo? Raport z badania przeprowadzonego w ośmiu kulturach w czasie pandemii COVID-19. W: B. Zawadzki, K. Fronczyk i K. Bargiel-Matusiewicz (red.), *Wybrane problemy współczesnej diagnostyki klinicznej w psychologii* (s. 179–215). Wydawnictwo Liberi Libri. <https://doi.org/10.47943/lib.9788363487676.rozdzial07>

EDUARDO KEEGAN

Facultad de Psicología, Universidad de Buenos Aires, Argentina

ALIYA MAMBETALINA

Gumilyov Eurasian National University, Kazakhstan

LOUISE MCHUGH

School of Psychology, University College Dublin, Ireland

MARIANA MIRACCO

Facultad de Psicología, Universidad de Buenos Aires, Argentina

ATSUSHI OSHIO

Faculty of Letters, Arts and Sciences, Waseda University, Japan

CHOWON PARK

Wydział Psychologii, Chung-Ang University, Korea Południowa

ANDRÉS PARTARRIEU

Facultad de Psicología, Universidad de Buenos Aires, Argentina

LORENA DE ROSA

Facultad de Psicología, Universidad de Buenos Aires, Argentina

RAIKHAN SABIROVA

Karaganda Buketov University, Kazakhstan

ADIL SAMEKIN

School of Liberal Arts, M. Narikbayev KAZGUU University, Kazakhstan

EMILIANO SÁNCHEZ

Facultad de Psicología, Universidad de Buenos Aires, Argentina

MARÍA SARNO

Facultad de Psicología, Universidad de Buenos Aires, Argentina

CECILIA TARRUELLA

Facultad de Psicología, Universidad de Buenos Aires, Argentina

GULMIRA TUYAKOVNA TOPANOVA

Kazakh National Women's Pedagogical University, Kazakhstan

GULMIRA M. TULEKOVA

Toraighyrov University, Kazakhstan

BOGDAN ZAWADZKI  <https://orcid.org/0000-0002-6578-8412>

Wydział Psychologii, Uniwersytet Warszawski

Czy pomiar objawów PTSD za pomocą *PTSD Checklist dla DSM-5 (PCL-5)* jest równoważny kulturowo? Raport z badania przeprowadzonego w ośmiu kulturach w czasie pandemii COVID-19

Is the measurement of PTSD symptoms with the use of the PTSD Checklist for DSM-5 (PCL-5) culturally equivalent? Report on a study conducted in eight cultures during the COVID-19 pandemic

Autor korespondencyjny: Maria Cyniak-Cieciura
e-mail: mcyniak-cieciura@swps.edu.pl

Streszczenie

Zaburzenie stresowe pourazowe opisuje chroniczną i utrudniającą codzienne funkcjonowanie reakcję na ekstremalny stres. Jednym z najbardziej znanych i najczęściej stosowanych narzędzi do pomiaru objawów zaburzenia stresowego pourazowego jest *PTSD Checklist dla DSM-5*. Narzędzie to zostało zaadaptowane do wielu różnych kultur, jednak jego równoważność pomiarowa była przedmiotem tylko jednego badania, obejmującego wyłącznie kraje europejskie. Celem przedstawionych analiz było zweryfikowanie pełnej równoważności pomiarowej *PTSD Checklist dla DSM-5*. Uwzględniono dane z ośmiu państw ($N > 200$): Polski, USA, Japonii, Argentyny, Korei Południowej, Irlandii, Wielkiej Brytanii i Kazachstanu, zebrane w czasie drugiej, trzeciej i czwartej fali pandemii COVID-19. Model struktury objawów oparty na DSM-5 okazał się dobrze dopasowany do danych, jednak najlepsze dopasowanie do danych wykazał hybrydowy model siedmioczynnikowy; weryfikację równoważności pomiarowej przeprowadzono względem tego modelu. Uzyskano potwierdzenie dla równoważności pomiarowej w zakresie jednakowej struktury oraz jednakowych ładunków czynnikowych (poziom konfiguralny i metryczny), nie potwierdzono natomiast równoważności w zakresie stałych regresji i błędów pomiarów (poziom skalarny i rezydualny). Objawy zaburzenia stresowego pourazowego były we wszystkich próbach dodatnio skorelowane z poziomem objawów depresyjnych, lękowych i stresu (silnie lub umiarkowanie), z poziomem stresorów związanych z COVID-19 oraz z reaktywnością emocjonalną (umiarkowanie lub słabo). Uzyskane wyniki pozwalają stwierdzić, że *PTSD Checklist dla DSM-5* mierzy uniwersalne kulturowo wskaźniki zaburzenia stresowego pourazowego, jednak w badaniach międzykulturowych należy każdorazowo zweryfikować równoważność pomiaru w uwzględnionych grupach, w celu potwierdzenia ekwiwalentności na poziomie skalarnym. W kolejnych badaniach nad równoważnością pomiarową *PTSD Checklist dla DSM-5* warto zadbać o dobór równolicznych i jednorodnych demograficznie prób.

Słowa kluczowe: zaburzenie stresowe pourazowe, *PTSD Checklist dla DSM-5*, model hybrydowy, równoważność pomiarowa

Abstract

Post-traumatic stress disorder describes a chronic and impairing response to extreme stress that affects daily functioning. One of the most well-known and commonly used tools for measuring post-traumatic stress disorder symptoms is the PTSD Checklist for DSM-5. This tool has been adapted to many different cultures, but measurement equivalence across cultures has only been examined in a single study of European countries. The aim of these analyses was to verify full measurement equivalence of the PTSD Checklist for DSM-5. Data from eight countries ($N > 200$) were included: Poland, USA, Japan, Argentina, South Korea, Ireland, United Kingdom, and Kazakhstan, collected during the second, third, and fourth waves of the COVID-19 pandemic. A symptom structure model based on DSM-5 presented good fit to the data, but the best fit was found for a hybrid seven-factor model; measurement equivalence was verified with respect to the latter model. Confirmation was obtained for equivalence in terms of identical structure and factor loadings (configural and metric invariance), but not in terms of intercepts and residuals (scalar and residual invariance). Post-traumatic stress disorder symptoms were positively correlated with the levels of depressive, anxiety, and stress symptoms (strongly or moderately) and with COVID-19-related stressors as well as emotional reactivity (moderately or weakly) in all samples. The results allow us to conclude that the PTSD Checklist for DSM-5 measures culturally universal indicators of post-traumatic stress disorder, but measurement equivalence should be verified in each cross-cultural study to confirm scalar invariance. In future research on PTSD Checklist for DSM-5 measurement equivalence, efforts should be made to select equally sized and demographically homogeneous samples.

Keywords: post-traumatic stress disorder, PTSD Checklist for DSM-5, hybrid model, measurement invariance

STRUKTURA OBJAWÓW ZABURZENIA STRESOWEGO POURAZOWEGO

Zaburzenie stresowe pourazowe (*posttraumatic stress disorder* [PTSD]) opisuje chroniczną i utrudniającą codzienne funkcjonowanie reakcję na ekstremalny stres związany z zagrożeniem życia i/lub zdrowia własnego, bliskich osób lub z byciem świadkiem (także poprzez wykonywaną pracę) takiego wydarzenia. Diagnoza PTSD wymaga, aby w okresie ponad miesiąca od przeżycia traumy obecne były objawy związane z ponownym przeżywaniem traumatycznego wydarzenia, unikaniem czynników i okoliczności przypominających o doświadczonej traumie, odczuwaniem fizjologicznych objawów stresu i pobudzenia autonomicznego układu nerwowego oraz z doświadczaniem zmian w zakresie nastroju i przekonań na temat siebie i świata (American Psychiatric Association, 2013; w tabeli 1 zamieszczono listę dwudziestu objawów PTSD zgodną z aktualną klasyfikacją zaburzeń – DSM-5; *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*).

Chroniczność nieleczonych objawów oraz klinicznej postaci PTSD (Bramsen i van der Ploeg, 1999; Goenjian i in., 2000; McCormack, 2009) oraz ich znaczący negatywny wpływ na zdrowie, kondycję psychiczną, funkcjonowanie społeczne i zawodowe (Arcaya i in., 2017; Davidson, 2000; Rao i in., 2015) uzasadniają wieloletnie poszukiwania najlepszego sposobu ujęcia typowych potraumatycznych reakcji (American Psychiatric Association, 1980, 2000, 2013) oraz czynników predestynujących do ich rozwoju (Brewin in., 2000; Ozer i in., 2003; Tortella-Feliu i in., 2019).

Struktura objawów PTSD ujmowanych w poszczególnych klasyfikacjach zaburzeń psychicznych Amerykańskiego Towarzystwa Psychiatrycznego (American Psychiatric Association, 1980, 2000, 2013) wielokrotnie stanowiła przedmiot dyskusji i nadal uważana jest za kontrowersyjną. W związku z tym, iż pomimo zmian wprowadzanych w kolejnych klasyfikacjach, wciąż nie znajduje jednoznacznego poparcia w wynikach badań (Armour i in., 2016); kolejne alternatywne próby pogrupowania objawów PTSD wykazywały coraz lepsze dopasowanie do danych empirycznych.

Tabela 1*Lista oraz struktura objawów PTSD zgodna z klasyfikacją DSM-5 oraz modele alternatywne*

Objawy PTSD wg DSM-5	4-czynnikowy model wg DSM-5	4-czynnikowy model z dysforią	5-czynnikowy model z dysforią i pobudzeniem	6-czynnikowy model z negatywnymi przekonaniami i odrętwieniem	6-czynnikowy model z anhedonią	7-czynnikowy model hybrydowy
B1: intruzyjne wspomnienia	R	R	R	R	R	R
B2: powtarzające się niepokojące sny	R	R	R	R	R	R
B3: reakcje dysocjacyjne	R	R	R	R	R	R
B4: reakcje emocjonalne na bodźce przypominające o traumie	R	R	R	R	R	R
B5: reakcje fizjologiczne na bodźce przypominające o traumie	R	R	R	R	R	R
C1: unikanie myśli	A	A	A	A	A	A
C2: unikanie czynników przypominających o traumie	A	A	A	A	A	A
D1: niemożność przypomnienia sobie	NAM	D	NAM	N	NA	NA
D2: negatywne przekonania o sobie/ świecie	NAM	D	NAM	NAM	NA	NA
D3: obwinianie siebie lub innych	NAM	D	NAM	NAM	NA	NA
D4: uporczywe negatywne emocje związane z traumą	NAM	D	NAM	NAM	NA	NA

Ciąg dalszy tabeli na następnej stronie

Ciąg dalszy tabeli z poprzedniej strony

Objawy PTSD wg DSM-5	4-czynnikowy model wg DSM-5	4-czynnikowy model z dysforią	5-czynnikowy model z dysforią i pobudzeniem	6-czynnikowy model z negatywnymi przekonaniem i odrętwieniem	6-czynnikowy model z anhedonią	7-czynnikowy model hybrydowy
D5: obniżone zainteresowanie i zaangażowanie w ważne aktywności	NAM	D	NAM	N	AN	AN
D6: uczucie odcięcia	NAM	D	NAM	N	AN	AN
D7: ograniczony afekt: niemożność odczuwania pozytywnych emocji	NAM	D	NAM	N	AN	AN
E1: wybuchy irytacji/gniewu	AR	D	DA	DA	DA	EB
E2: lekkomyślne lub autodestrukcyjne zachowanie	AR	D	DA	DA	DA	EB
E3: nadmierne pobudzenie	AR	AA	AA	AA	AA	AA
E4: przesadna reakcja na zaskoczenie	AR	AA	AA	AA	AA	AA
E5: trudności w koncentracji uwagi	AR	D	DA	DA	DA	DA
E6: zaburzenia snu	AR	D	DA	DA	DA	DA

Adnotacja. AR – zmiany w zakresie pobudzenia i reaktywności, AA – pobudzenie lękowe, AN – pobudzenie lękowe, A – unikanie, D – dysforia, DA – pobudzenie dysforyczne, EB – zachowania eksternalizacyjne, N – odrętwienie, NA – negatywny afekt, NAM – negatywne zmiany w zakresie przekonani i nastroju, R – ponowne przeżywanie.

Tabela 1 zawiera porównanie najbardziej popularnych modeli grupujących objawy PTSD. Model bazujący na klasyfikacji DSM-5 zakłada wyodrębnienie czterech czynników: ponownego przeżywania, unikania, negatywnych zmian w zakresie przekonań i nastroju oraz w zakresie pobudzenia i reaktywności. Kolejny model, czteroczynnikowy z dodatkowo wyodrębnioną dysforią, bazuje na wynikach analiz Simmsa i in. (2002) i poza ponownym przeżywaniem oraz unikaniem, grupuje objawy w czynnik dysforii oraz lękowego pobudzenia. Pięcoczynnikowy model zaproponowany przez Elhai i in. (2011) również wyodrębnia ponowne przeżywanie i unikanie, lecz pozostałe objawy grupuje w czynniki odrętwienia, dysforii i pobudzenia lękowego. Dwa popularne modele sześcioczynnikowe uwzględniają ponowne przeżywanie i unikanie, a oprócz nich także negatywny afekt, anhedonię, pobudzenie dysforyczne i lękowe (Liu i in., 2014) lub odrętwienie, zachowania eksternalizacyjne, pobudzenie lękowe i dysforyczne (Tsai i in., 2014). W kolejnych badaniach powyższe modele okazały się niejednokrotnie prezentować słabsze dopasowanie do danych w porównaniu z tzw. siedmioczynnikowym modelem hybrydowym (Armour i in., 2015), który wyodrębnia czynniki ponownego przeżywania, unikania, negatywnego afektu, anhedonii, zachowań eksternalizacyjnych, pobudzenia lękowego i dysforycznego. Model ten od kilku lat uznawany jest za najbardziej adekwatnie ujmujący objawy PTSD (Armour i in., 2016).

PTSD CHECKLIST DLA DSM-5

Jednym z najbardziej znanych i najczęściej stosowanych narzędzi do pomiaru objawów PTSD oraz do postawienia przesiewowej diagnozy klinicznej postaci zaburzenia jest *PTSD Checklist dla DSM-5 (PTSD Checklist for DSM-5 [PCL-5]*; Blevins i in., 2015; Weathers i in., 2013), opracowane przez National Center of Posttraumatic Stress Disorder w Stanach Zjednoczonych. Narzędzie jest samoopisowe i składa się z 20 stwierdzeń odnoszących się do 20 objawów PTSD zgodnie z klasyfikacją DSM-5, ocenianych ze względu na dokuczliwość objawów na pięciostopniowej skali typu likertowskiego, gdzie 0 oznacza *wcale*, a 4 – *ogromnie*. PCL-5 zostało zaadaptowane do wielu różnych kultur, m.in. brytyjskiej i francuskiej (Ashbaugh i in., 2016), duńskiej (van Praag i in., 2020), niemieckiej (Krüger-Gottschalk i in., 2017), portugalskiej (Carvalho i in., 2020), japońskiej (Ito i in., 2019), kurdyjskiej i arabskiej (Ibrahim i in., 2018), wstępne doniesienia dotyczą także adaptacji polskiej (Ogińska-Bulik

i in., 2018) i włoskiej (Di Tella i in., 2022), i wykazują dobre właściwości psychometryczne narzędzia. Pomiar objawów PTSD w różnych kulturach okazał się cechować zadowalającą rzetelnością, trafnością teoretyczną w zakresie zbieżnym i różnicowym oraz trafnością kryterialną (Forkus i in., 2023). Silne dodatnie korelacje odnotowano pomiędzy objawami PTSD oraz poziomem doświadczanych stresorów, objawami depresyjnymi i lękowymi, negatywnymi przekonaniami, dysocjacją i myślami samobójczymi (Forkus i in., 2023). Czteroczynnikowa struktura objawów zakładana przez DSM-5 okazała się relatywnie dobrze dopasowana do danych, jednakże lepsze wskaźniki dopasowania uzyskiwano dla wspomnianych powyżej modeli sześć- i siedmoczynnikowych (Forkus i in., 2023). Pomimo tak dużej popularności PCL-5, w momencie opracowywania tego rozdziału znaleźliśmy tylko jedno badanie poświęcone wprost weryfikacji równoważności pomiarowej różnych wersji kulturowych tego narzędzia.

RÓWNOWAŻNOŚĆ POMIAROWA

Analiza równoważności pomiarowej (*measurement invariance*) ma na celu sprawdzenie, czy pomiar danym narzędziem jest porównywalny w różnych grupach osób badanych (np. płciowych, wiekowych, kulturowych, stosujących różne wersje tego samego narzędzia). Uzyskanie dowodów na rzecz równoważności pomiarowej uprawnia do dokonywania porównań międzygrupowych, w tym wnioskowania o różnicach średnich oraz do badania relacji pomiędzy zmienną badaną danym narzędziem a innymi zmiennymi (Lubiewska i Głogowska, 2018; Putnick i Bornstein, 2016; Schmitt i Kuljanin, 2008). Weryfikacja równoważności pomiarowej polega na przeprowadzeniu serii confirmacyjnych analiz czynnikowych dla zagnieżdżonych modeli, z nałożeniem kolejnych ograniczeń dotyczących najpierw jednakowej struktury danego konstruktów w badanych grupach (tzw. równoważność konstruktów lub konfiguralna; *configural invariance*), potem także jednakowych ładunków czynnikowych (tzw. równoważność metryczna; *metric invariance*), jednakowych stałych regresji (tzw. równoważność skalarna; *scalar invariance*) oraz jednakowych błędów pomiaru (tzw. równoważność rezydualna; *residual invariance*). Możliwe jest także wprowadzenie ograniczenia w zakresie jednakowych średnich, wariancji i kowariancji, lecz w praktyce wykazanie równoważności na tym poziomie jest zadaniem niemalże nieosiągalnym ze względu na liczbę nałożonych

restrykcji (Lubiewska i Głogowska, 2018; Putnick i Bornstein, 2016; Schmitt i Kuljanin, 2008). Porównanie kolejnych modeli następuje na podstawie istotności różnic pomiędzy statystykami chi-kwadrat (χ^2) dla poszczególnych modeli (istotne wartości świadczą o braku równoważności pomiarowej) i/lub różnic w wartościach pozostałych wskaźników dopasowania: CFI, RMSEA, SRMR. Ze względu na wrażliwość statystyki χ^2 na wielkość badanych prób, niejednokrotnie tylko wnioskowanie oparte na różnicach w wartościach CFI, RMSEA i SRMR pozostaje uzasadnione. Aby móc wnioskować o każdym kolejnym poziomie równoważności, spełnione muszą być dla niego kryteria interpretacji wskaźników dopasowania, analogiczne do tych stosowanych w tradycyjnej konfirmacyjnej analizie czynnikowej, np. za: Hu i Bentler (1999). Wartości $CFI \geq 0,90$; $TLI \geq 0,90$; $RMSEA$ i $SRMR \leq 0,08$ wskazują na dobre dopasowanie modelu do danych; wartość różnicy pomiędzy χ^2 powinna być nieistotna statystycznie, zaś różnice w wartościach pozostałych wskaźników powinny spełniać poniższe kryteria: $\Delta CFI \leq -0,020$ ($\leq -0,010$ w przypadku równoważności skalarnej), $\Delta RMSEA \leq 0,030$ ($\leq 0,010$ w przypadku równoważności skalarnej) i $\Delta SRMR \leq 0,030$ (Chen, 2007; Rutkowski i Svetina, 2014). Spełnienie powyższych wymagań wobec niższego poziomu równoważności uprawnia do wnioskowania o kolejnym poziomie (Lubiewska i Głogowska, 2018; Putnick i Bornstein, 2016; Schmitt i Kuljanin, 2008).

Potwierdzenie równoważności pomiarowej na poziomie konstruktu oznacza, że struktura badanego konstruktu w analizowanych grupach jest jednakowa, z tą samą liczbą czynników latentnych wskazywanych przez te same zmienne we wszystkich uwzględnionych grupach. Równoważność na poziomie metrycznym oznacza, że każda zmienna ładuje ten sam czynnik latentny w porównywalnym stopniu we wszystkich grupach, co jest jednoznaczne z pomiarem tej samej zmiennej w tych grupach. Potwierdzenie równoważności na poziomie konfiguralnym i metrycznym uzasadnia analizowanie związków pomiędzy zmiennymi w poszczególnych grupach, jednak nie pozwala na testowanie różnic średnich latentnych pomiędzy grupami. Ostatni rodzaj analiz jest uzasadniony tylko w przypadku potwierdzenia równoważności skalarnej (Lubiewska i Głogowska, 2018; Putnick i Bornstein, 2016; Schmitt i Kuljanin, 2008).

W jedynym odnalezionym przez nas badaniu równoważności pomiarowej PCL-5, przeprowadzonym przez Bockhop i in. (2022), dokonano porównania pomiaru za pomocą sześciu różnych wersji kulturowych narzędzia (duńskiej, angielskiej, fińskiej, włoskiej, norweskiej i hiszpańskiej, co istotne zatem – były to tylko kraje osadzone w kulturze europejskiej). Próby liczyły między

212 (próba brytyjska) a 586 badanych (próba duńska) i były ograniczone do osób, które doznały poważnego urazu mózgu. Porównanie dotyczyło modelu czteroczynnikowego opartego na klasyfikacji DSM-5 oraz równoważności konfiguralnej i metrycznej. Zarówno na podstawie nieistotnych wartości różnicy statystyki χ^2 , jak i różnic wartości pozostałych wskaźników dopasowania stwierdzono równoważność konfiguralną i metryczną pomiaru PCL-5. Wynik ten uprawnia do wniosku, że struktura objawów wg DSM-5 mierzona przy pomocy PCL-5 jest taka sama i jednakowo ładowana przez poszczególne objawy w badanych kulturach, lecz nie uprawnia do porównań międzykulturowych (poziom równoważności skalarnej nie został w tym badaniu zaraportowany).

Warto wspomnieć o dwóch innych badaniach dotyczących równoważności pomiarowej PCL-5. Pierwsze z nich przeprowadzone zostało przez Contractor i in. (2018) w celu porównania pomiaru PTSD w grupie osób, które doświadczyły jednej traumy z osobami, które doświadczyły więcej traum. W badaniu tym dane najlepiej opisywał model hybrydowy, uzyskano także potwierdzenie dla równoważności pomiarowej na poziomie konfiguralnym, metrycznym i skalarnym. Natomiast w badaniu Caldas i in. (2020) pełną równoważność pomiarową (na poziomie konfiguralnym, metrycznym, skalarnym i rezydualnym) potwierdzono pomiędzy grupą studentów a grupą osób zrekrutowanych za pomocą platformy MTurk. Obydwa badania przeprowadzone zostały w obrębie jednego kraju.

CELE I HIPOTEZY BADAWCZE

Celem przeprowadzonych przez nas analiz była weryfikacja pełnej równoważności pomiarowej (na poziomie konfiguralnym, metrycznym, skalarnym i rezydualnym) PCL-5 w ośmiu krajach (kulturach). Do realizacji tego celu wykorzystaliśmy dane zebrane w międzynarodowym projekcie badawczym COVID-TEMPS. Chcieliśmy także sprawdzić trafność zbieżną pomiaru PCL-5 w poszczególnych kulturach. W związku z tym, że do rozwoju objawów tego zaburzenia predestynuje reaktywność emocjonalna (Cyniak-Cieciura i Zawadzki, 2021), częstość występowania różni się w zależności od zmiennych socjodemograficznych (Brewin i in., 2000), zaś same objawy są istotnie skorelowane z objawami depresyjnymi, lękowymi i poziomem doświadczanych stresorów (Forkus i in., 2023), oczekiwaliśmy, że w poszczególnych kulturach objawy PTSD mierzone za pomocą PCL-5 będą pozytywnie skorelowane z uwzględnionymi przez nas w badaniu zmiennymi: objawami depresyjnymi, objawami

lękowymi, objawami stresu, poziomem stresorów związanych z COVID-19 oraz z reaktywnością emocjonalną. Spodziewaliśmy się także, że dwie ostatnie zmienne, wraz z płcią i wiekiem, stanowiąc będą istotne predyktory rozwoju PTSD w analizach regresyjnych.

METODA

Procedura

Badanie przeprowadzone zostało w ramach międzynarodowego projektu COVID-TEMPS. Dane wykorzystane w opisywanym badaniu zebrane zostały w ośmiu państwach²: Polsce, USA, Japonii, Argentynie, Korei Południowej, Irlandii, Wielkiej Brytanii i w Kazachstanie. Dane zbierano w okresie sierpień 2020–październik 2021, podczas drugiej, trzeciej oraz czwartej fali pandemii COVID-19. Uczestnicy badania poproszeni zostali o wypełnienie szeregu samoopisowych kwestionariuszy (opisanych poniżej). Badanie w każdym kraju przeprowadzono w formie online, z wykorzystaniem profesjonalnej platformy badawczej (najczęściej Qualtrics). Uczestników rekrutowano w większości poprzez bezpośredni mailing w ramach uczelni badaczy, w tym poprzez uczelniane platformy badawcze, umożliwiając udział w badaniu nie tylko studentom (z wyjątkiem Wielkiej Brytanii, gdzie uczestnicy zostali zrekrutowani poprzez platformę MTurk). W zależności od polityki uczelni uczestnicy otrzymywali punkty w zamian za udział w badaniu. Jedynym kryterium wykluczającym z udziału w badaniu był wiek poniżej 18 lat. Wszyscy uczestnicy zaznaczyli świadomą zgodę na udział w badaniu. Badanie uzyskało pozytywne opinie lokalnych Komisji Etycznych.

Uczestnicy

Opis prób z poszczególnych państw zamieszczono w tabeli 2.

² W projekcie COVID-TEMPS dane zebrane zostały także przez badaczy z Włoch, jednak ze względu na brak adaptacji PCL-5 do populacji włoskiej do czasu uruchomienia projektu w 2021 roku, narzędzie to zostało w tym kraju pominięte, a próba włoska nie została uwzględniona w opisywanych w tym rozdziale analizach.

Tabela 2*Szczegółowe informacje o uczestnikach badania z poszczególnych państw*

Proba	N (K, M)	Przedział wiekowy (M; SD)	Wykształcenie	Miejsce zamieszkania	Stan cywilny	Diagnoza PTSD (N; %)	Okres badania
Polska	945 (810, 129)	18–60 (26,77; 8,67)	Brak danych	49,2% – > 500 000 mieszkańców 15,4% – 100 000–500 000 mieszkańców 15,1% – 20 000–100 000 mieszkańców 7,8% < 20 000 mieszkańców 12,4% – wieś	61,0% – związek małżeński/ stały związek 36,7% – brak stałej relacji 2,3% – po rozwodzie/ wdowa/wdowiec	300; 27,3%	sierpień 2020– luty 2021
USA	596 (407, 177)	18–52 (20,39; 4,65)	1,2% – podstawowe 1,2% – zawodowe 54,5% – średnie 41,9% – wyższe	5,7% – > 500 000 mieszkańców 8,2% – 100 000–500 000 mieszkańców 30,7% – 20 000–100 000 mieszkańców 32,9% < 20 000 mieszkańców 20,8% – wieś	28,5% – związek małżeński/ stały związek 69,1% – brak stałej relacji 1,2% – po rozwodzie/ wdowa/wdowiec	269; 33,1%	listopad 2020– kwiecień 2021
Japonia	349 (248, 69)	18–86 (22,22; 8,11)	0,6% – podstawowe 26,4% – średnie 66,8% – wyższe	61,6% – > 500 000 mieszkańców 24,9% – 100 000–500 000 mieszkańców 5,7% – 20 000–100 000 mieszkańców 1,1% < 20 000 mieszkańców 0,3% – wieś	5,4% – związek małżeński/ stały związek 87,1% – brak stałej relacji 0,9% – po rozwodzie/ wdowa/wdowiec	39; 7,8%	listopad 2020– październik 2021
Argentyna	634 (495, 138)	18–85 (39,34; 13,06)	1,3% – podstawowe 11,5% – zawodowe 12,1% – średnie 75,1% – wyższe	77,8% – > 500 000 mieszkańców 12,8% – 100 000–500 000 mieszkańców 6,3% – 20 000–100 000 mieszkańców 2,5% < 20 000 mieszkańców 0,6% – wieś	58,8% – związek małżeński/ stały związek 31,2% – brak stałej relacji 9,9% – po rozwodzie/ wdowa/wdowiec	65; 10,2%	maj 2020– lipiec 2020

Ciąg dalszy tabeli na następnej stronie

Ciąg dalszy tabeli z poprzedniej strony

Próba	N (K, M)	Przedział wiekowy (M; SD)	Wykształcenie	Miejsce zamieszkania	Stan cywilny	Diagnoza PTSD (N; %)	Okres badania
Korea Południowa	551 (280, 271)	20–59 (40, 18; 10, 66)	5,4% – zawodowe 7,1% – średnie 87,5% – wyższe	70,8% – > 500 000 mieszkańców 17,6% – 100 000–500 000 mieszkańców 6,9% – 20 000–100 000 mieszkańców 4,0% < 20 000 mieszkańców 0,7% – wieś	55,2% – związek małżeński/ stały związek 41,2% – brak stałej relacji 3,6% – po rozwodzie/ wdowa/wdowiec	59; 10,7%	listopad 2020– październik 2021
Irlandia	216 (144, 72)	18–69 (26, 17; 9, 83)	3,2% – podstawowe 3,2% – zawodowe 25,9% – średnie 67,6% – wyższe	30,6% – > 500 000 mieszkańców 6,0% – 100 000–500 000 mieszkańców 13,0% – 20 000–100 000 mieszkańców 22,7% < 20 000 mieszkańców 27,8% – wieś	42,6% – związek małżeński/ stały związek 54,6% – brak stałej relacji 2,8% – po rozwodzie/ wdowa/wdowiec	53; 24,5%	listopad 2020– październik 2021
Wielka Brytania	271 (105, 164)	18–71 (34, 25; 10, 78)	1,2% – podstawowe 1,2% – zawodowe 54,5% – średnie 41,9% – wyższe	33,9% – > 500 000 mieszkańców 25,1% – 100 000–500 000 mieszkańców 24,4% – 20 000–100 000 mieszkańców 10,0% < 20 000 mieszkańców 6,6% – wieś	61,6% – związek małżeński/ stały związek 35,4% – brak stałej relacji 3,0% – po rozwodzie/ wdowa/wdowiec	108; 39,9%	listopad 2020– październik 2021
Kazachstan	502 (392, 104)	18–59 (21, 81; 6, 14)	4,8% – podstawowe 4,8% – zawodowe 12,5% – średnie 77,9% – wyższe	Brak danych	17,5% – związek małżeński/ stały związek 79,9% – brak stałej relacji 2,4% – po rozwodzie/ wdowa/wdowiec	111; 22,1%	listopad 2020– październik 2021

Adnotacja. N – liczność próby, K – kobiety, M – mężczyźni, pozostałe osoby nie udzieliły informacji na temat swojej płci; M – średnia, SD – odchylenie standardowe; % diagnoz PTSD policzono na podstawie odpowiedzi udzielonych w kwestionariuszu PCL-5 i spełnienia kryteriów DSM-5. Próby różniły się w odniesieniu do płci: $\chi^2(7) = 403,48; p < 0,001$ i wieku: $\chi^2(7) = 2208,72; p < 0,001$ z wyjątkiem Polski i Irlandii, Japonii i Kazachstanu oraz Argentyny i Korei Południowej; próby były nierównoliczne: $\chi^2(7) = 987,69; p < 0,001$; poza Koreą Południową: $\chi^2(1) = 0,147; p = 0,701$ we wszystkich pozostałych próbach w badaniu uczestniczyło istotnie więcej kobiet niż mężczyzn.

Narzędzia

Dane demograficzne oraz inne podstawowe informacje dotyczące COVID-19 uzyskano za pomocą ankiety opracowanej specjalnie na potrzeby tego badania. Uwzględniała ona pytania o płeć, wiek, miejsce zamieszkania i stan cywilny.

Pomiar nasilenia objawów PTSD w badaniu COVID-TEMPS odbywał się za pomocą sześciu różnych wersji językowych PCL-5 – polskiej w tłumaczeniu Popiel i in. (2021), angielskiej (USA, Wielka Brytania, Irlandia) w adaptacji Ashbaugh i in. (2016), japońskiej w adaptacji Ito i in. (2019), hiszpańskiej w adaptacji Martinez-Levy i in. (2021), koreańskiej w adaptacji Lee i in. (2020) oraz rosyjskiej w adaptacji Cheung i in. (2019).

Pomiaru reaktywności emocjonalnej dokonano za pomocą *Formalnej charakterystyki zachowania – kwestionariusza markerów temperamentalnych* (*Formal Characteristic of Behaviour – Temperament Markers Inventory* [FCB-TMI]; Cyniak-Cieciura i in., 2024) – wspólnej kulturowo wersji kwestionariusza temperamentu służącego do badania markerów cech temperamentalnych zgodnie z Regulacyjną Teorią Temperamentu. W ramach projektu COVID-TEMPS, na dziewięciu próbach z różnych państw, wykazano równoważność metryczną pomiaru całym narzędziem. W analizie wykorzystano tylko skalę reaktywności emocjonalnej, której wynik obliczony został jako suma odpowiedzi na sześć stwierdzeń (przy czterostopniowej skali odpowiedzi: 1 – *zdecydowanie nie zgadzam się*, 4 – *zdecydowanie zgadzam się*).

Pomiaru stresorów związanych z pandemią COVID-19 dokonano za pomocą opracowanego na potrzeby tego badania kwestionariusza, w skład którego weszło 17 pytań o konieczność przebywania w przymusowej kwarantannie, nadmierne zagęszczenie domowników związane z izolacją lub kwarantanną, zmianę rytmu dnia oraz sposobu pracy, konieczność rezygnacji z ulubionych aktywności (sportowych, kulturalnych itp.), poczucie izolacji i ograniczenie kontaktów z ludźmi, samotność, pogorszenie sytuacji finansowej, obawę o pogorszenie sytuacji materialnej lub utratę pracy, obawę o możliwość zakażenia się wirusem i rozwiniecie choroby, zwiększenie obciążenia nauką i/lub pracą, stres związany ze zmianą sposobu nauki i/lub pracy, konieczność bezpośredniego kontaktu z ludźmi z zewnątrz, bezpośredni kontakt z ludźmi zakażonymi, utratę możliwości kontynuowania nauki i/lub wykonywania pracy zarobkowej, chorobę COVID-19 w gronie bliskich i dalszych znajomych oraz rodziny. Badani ustosunkowywali się do poszczególnych stwierdzeń na czterostopniowej skali odpowiedzi od 1 – *Zupełnie mnie nie dotyczyło do*

4 – *Dotyczyło i stanowiło duży problem*. W analizach wykorzystano wynik ogólny, będący sumą punktów z udzielonych odpowiedzi.

Pomiaru negatywnego stanu emocjonalnego dokonano za pomocą kwestionariusza DASS-21 (*Depression, Anxiety & Stress Scale*; Lovibond i Lovibond, 1995), który składa się z 21 pozycji i służy do pomiaru objawów depresyjnych, lękowych oraz odczuwanego stresu (trzy skale, każda po siedem pozycji). W badaniu wykorzystano polską wersję narzędzia, opracowaną przez Makarę-Studzińską i in. (2024) oraz wynik ogólny obliczany na podstawie wszystkich pozycji kwestionariusza.

Wskaźniki rzetelności skal (alfa Cronbacha) były zadowalające, zaprezentowano je w tabeli A1 w załączniku.

Procedura analizy danych

W pierwszym kroku, na podstawie wyników konfirmacyjnej analizy czynnikowej dla całej próby oraz w podziale na kraje (z założeniem jednakowej struktury objawów we wszystkich próbach), porównano szereg modeli struktury prostej symptomów PTSD, opisanych powyżej (zob. tabela 1). Modele z uwzględnieniem czynnika ogólnego – hierarchiczny lub dwuczynnikowy albo wychodzące poza strukturę prostą – nie były testowane. Modele analizowano z wykorzystaniem odpornej metody największej wiarygodności (*Maximum Likelihood Robust* [MLR]) jako metody estymacji (Rhemtulla i in., 2012). Dopasowanie modeli oceniane było na podstawie wartości statystyki chi-kwadrat (χ^2) oraz innych wskaźników: *robust* RMSEA, *robust* CFI, *robust* TLI, SRMR, AIC, BIC oraz SABIC. Dobroć dopasowania modeli do danych oceniono zgodnie z kryteriami zaproponowanymi przez Hu i Bentlera (1999), przytoczonymi powyżej. Wyboru modelu o najlepszym dopasowaniu dokonano na podstawie wartości BIC. W drugim kroku zweryfikowano równoważność pomiarową struktury objawów PTSD, która wynikała z modelu o najlepszym dopasowaniu (wyodrębnionego w kroku pierwszym) w poszczególnych kulturach (krajach). W tym celu przeprowadzono wielogrupową konfirmacyjną analizę czynnikową – analizę modeli zagnieżdżonych, zakładając równość ładunków czynnikowych oraz stałych regresji (równoważność struktury założono już na pierwszym etapie analiz konfirmacyjnych, opierając na jej wynikach wybór najlepszego modelu). W ocenie równoważności pomiarowej posłużono się kryteriami zaproponowanymi przez Chen (2007) oraz Rutkowskiego i Svetinę (2014), które przedstawiono powyżej w tekście.

Na koniec przeprowadzono analizy mające na celu sprawdzenie trafności teoretycznej narzędzia w aspekcie zbieżnym w poszczególnych kulturach. W tym celu obliczono współczynniki korelacji *r*-Pearsona pomiędzy objawami PTSD a objawami depresyjnymi, lękowymi i stresowymi oraz stresorami związanymi z COVID-19 i reaktywnością emocjonalną. Przeprowadzono także hierarchiczną analizę regresji, w której predyktorami PTSD były stresory związane z COVID-19, reaktywność emocjonalna oraz dwie zmienne demograficzne – płeć i wiek.

Wszystkie analizy przeprowadzono w programie R (pakiet *lavaan*, wersja 0.6-13; R Core Team, 2018; Rosseel, 2012) oraz SPSS, wersja 25.

WYNIKI

W tabeli 3 zamieszczono wskaźniki dopasowania dla testowanych modeli struktury objawów PTSD. Zarówno w przypadku analiz przeprowadzonych w całej próbie, jak i z założeniem równoważnej struktury w poszczególnych próbach najsłabsze dopasowanie do danych wykazał model zakładający ułożenie objawów w strukturę jednoczynnikową; model bazujący na aktualnej klasyfikacji DSM-5 uzyskał zadowalające wskaźniki dopasowania; zaś najlepsze wskaźniki uzyskano dla siedmioczynnikowego modelu hybrydowego (Armour i in., 2015): $\chi^2(149) = 1325,451$; $p < 0,001$; *robust* CFI = 0,970; *robust* RMSEA = 0,057 (90% CI [0,054; 0,060]); *robust* TLI = 0,962; SRMR = 0,026; AIC = 190985,220; BIC = 191365,593; SABIC = 191171,764 w całej próbie i $\chi^2(1192) = 2486,437$; $p < 0,001$; *robust* CFI = 0,966; *robust* RMSEA = 0,058 (90% CI [0,055; 0,061]); *robust* TLI = 0,957; SRMR = 0,032; AIC = 186777,678; BIC = 190818,363; SABIC = 188759,328 przy założeniu równoważnej struktury objawów we wszystkich próbach. Standardowe wartości ładunków dla modelu siedmioczynnikowego w całej próbie przedstawiono na rysunku 1.

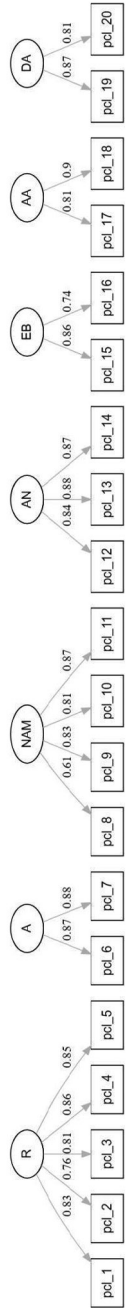
Tabela 3*Wskaźniki dobroci dopasowania poszczególnych modeli w całej próbie*

Model	χ^2 (df)	p	AIC	BIC	SABIC	robust CFI	robust RMSEA [90% CI]	robust TLI	SRMR
Analizy w całej próbie									
1-czynnikowy ogólny	5471,502 (170)	< 0,001	197576,766	197826,191	197699,090	0,862	0,115 [0,112; 0,117]	0,845	0,050
4-czynnikowy model DSM-5	2683,809 (164)	< 0,001	193118,643	193405,482	193259,316	0,935	0,080 [0,077; 0,083]	0,923	0,035
4-czynnikowy model dysforyczny	2915,560 (164)	< 0,001	193485,612	193772,451	193626,285	0,929	0,084 [0,081; 0,086]	0,918	0,037
5-czynnikowy model dysforyczno-pobudzeniowy	2468,003 (160)	< 0,001	192780,775	193092,557	192933,680	0,941	0,077 [0,075; 0,080]	0,930	0,034
6-czynnikowy model z negatywnymi przekonaniami i odrętwieniem	1623,685 (155)	< 0,001	191447,471	191790,431	191615,667	0,963	0,062 [0,060; 0,065]	0,954	0,033
6-czynnikowy model anhedoniczny	1524,981 (155)	< 0,001	191295,381	191638,340	191463,576	0,965	0,060 [0,058; 0,063]	0,957	0,028
7-czynnikowy model hybrydowy	1325,451 (149)	< 0,001	190985,220	191365,593	191171,764	0,970	0,057 [0,054; 0,060]	0,962	0,026
Analizy z założeniem równoważności struktury we wszystkich próbach									
1-czynnikowy ogólny	7156,693 (1360)	< 0,001	193519,675	196512,776	194987,564	0,843	0,116 [0,113; 0,119]	0,825	0,057
4-czynnikowy model DSM-5	4059,152 (1312)	< 0,001	188919,702	192212,113	190534,380	0,927	0,081 [0,078; 0,084]	0,915	0,042
4-czynnikowy model dysforyczny	4321,869 (1312)	< 0,001	189318,688	192611,098	190933,366	0,919	0,085 [0,082; 0,088]	0,907	0,044
5-czynnikowy model dysforyczno-pobudzeniowy	3801,321 (1280)	< 0,001	188591,222	192083,173	190303,759	0,933	0,078 [0,076; 0,081]	0,920	0,041
6-czynnikowy model z negatywnymi przekonaniami i odrętwieniem	2880,675 (1240)	< 0,001	187278,315	191019,691	189113,176	0,957	0,064 [0,061; 0,067]	0,947	0,039
6-czynnikowy model anhedoniczny	2716,849 (1240)	< 0,001	187041,020	190782,395	188875,881	0,961	0,061 [0,058; 0,064]	0,952	0,034
7-czynnikowy model hybrydowy	2486,437 (1192)	< 0,001	186777,678	190818,363	188759,328	0,966	0,058 [0,055; 0,061]	0,957	0,032

Adnotacja. AIC – kryterium informacyjne Akaikego, BIC – kryterium informacyjne Bayesa z poprawką na wielkość próby, robust CFI – odporny porównawczy indeks dopasowania, robust RMSEA – odporny pierwiastek średniokwadratowego błędu aproksymacji, 90% CI – przedział ufności, robust TLI – odporny wskaźnik Tuckera-Lewisa, SRMR – średni standardzowany kwadrat reszt. Pogrubiono rozwiązanie o najlepszych wskaźnikach dopasowania.

Rysunek 1

Standardowe ładunki pozycji w całej próbie (siedmioczynnikowy model hybrydowy)



Adnotacja: R – ponowne przeżywanie, A – unikanie, NAM – negatywne zmiany w zakresie przekonań i nastroju, AN – anhedonia, EB – zachowania eksternalizacyjne, AA – pobudzenie lękowe, DA – pobudzenie dysforyczne, PCL_1–PCL_20 – numery pozycji kwestionariuszowych *PTSD Checklist dla DSM-5*. Dla klarowności pominięto korelacje pomiędzy latentnymi.

W kolejnym kroku sprawdzono równoważność pomiarową na poziomach metrycznym i skalarnym 7-czynnikowej hybrydowej struktury objawów PTSD w poszczególnych kulturach (krajach).

Wyniki przedstawiono w tabeli 4. Wskaźniki dobroci dopasowania oraz porównanie poszczególnych modeli pozwoliły na potwierdzenie równoważności metrycznej. Nie potwierdzono natomiast równoważności skalarnej, która uzasadnia dokonywanie porównań międzygrupowych. Z tego powodu w dalszej części pracy postanowiliśmy zaprezentować dane oddzielnie dla poszczególnych krajów (kultur). Zrezygnowaliśmy także z porównania różnic w średnim poziomie objawów PTSD pomiędzy krajami (kulturami).

Tabela 4

Równoważność pomiarowa PCL-5 w ośmiu kulturach

Model	χ^2 (df)	χ^2 diff	$\Delta\chi^2$	$p\Delta\chi^2$ diff	robust CFI	Δ CFI	robust RMSEA [90% CI]	Δ RMSEA	SRMR	Δ SRMR
Równoważność metryczna	2961,495 (1283)	< 0,001	475,058	< 0,001	0,956	−0,010	0,063 [0,060; 0,066]	0,005	0,062	0,030
Równoważność skalarna	3748,303 (1374)	< 0,001	786,808	< 0,001	0,940	−0,016	0,072 [0,069; 0,074]	0,009	0,071	0,009

Adnotacja. χ^2 diff – istotność wartości statystyki chi-kwadrat, $p\Delta\chi^2$ diff – istotność różnicy pomiędzy wartościami statystyki chi-kwadrat, robust CFI – odporny porównawczy indeks dopasowania, robust RMSEA – odporny pierwiastek średniokwadratowego błędu aproksymacji, 90% CI – przedział ufności, SRMR – średni standaryzowany kwadrat reszt. Pogrubiono rozwiązanie o najlepszych wskaźnikach dopasowania.

Analiza korelacji (tabela A1 w załączniku) wykazała podobne zależności w poszczególnych kulturach. Odnotowaliśmy silne lub umiarkowane pozytywne korelacje pomiędzy sumą wszystkich objawów PTSD, jak również poszczególnymi grupami objawów (ponowne przeżywanie, unikanie, negatywne zmiany w zakresie przekonań i nastroju, anhedonia, zachowania eksternalizacyjne, pobudzenie lękowe i pobudzenie dysforyczne), a objawami depresyjnymi, lękowymi oraz stresowymi. Te same zmienne okazały się także umiarkowanie lub słabo pozytywnie skorelowane z poziomem stresorów związanych z COVID-19 oraz z reaktywnością emocjonalną. Zakresy korelacji i średnie

korelacje ze wszystkich kultur pomiędzy objawami PTSD a pozostałymi zmiennymi zamieściliśmy w tabeli 5. Uwzględnienie w kolejnych krokach w modelu regresyjnym płci i wieku, stresorów związanych z COVID-19, reaktywności emocjonalnej oraz interakcji dwóch ostatnich zmiennych pozwoliło wyjaśnić łącznie 9,5–27,1% wariancji w zależności od kraju. Stresory związane z COVID-19 oraz reaktywność emocjonalna okazały się istotnymi predyktorami objawów PTSD we wszystkich kulturach. Interakcja pomiędzy tymi zmiennymi była istotna tylko w Stanach Zjednoczonych, Argentynie i Korei Południowej (oraz na poziomie tendencji w Polsce i Irlandii). Wyniki te przedstawiono w tabeli 6.

Tabela 5

Zakresy rzetelności, korelacji oraz średnie korelacje pomiędzy objawami PTSD a stresorami związanymi z COVID-19, poziomem objawów depresji, lęku i stresu oraz reaktywnością emocjonalną

Zmienne	PTSD	R	A	NAM	AN	EB	AA	DA
C-19	0,19–0,37 0,31	0,17–0,37 0,29	0,20–0,29 0,25	0,14–0,30 0,25	0,17–0,49 0,25	0,10–0,29 0,20	0,10–0,32 0,25	0,18–0,33 0,26
DEPR	0,56–0,71 0,66	0,47–0,75 0,51	0,34–0,62 0,45	0,47–0,75 0,57	0,51–0,73 0,57	0,49–0,74 0,56	0,48–0,76 0,60	0,49–0,71 0,60
LĘK	0,54–0,84 0,65	0,42–0,75 0,48	0,30–0,62 0,38	0,41–0,76 0,57	0,53–0,77 0,61	0,47–0,71 0,55	0,43–0,76 0,57	0,48–0,73 0,60
STR	0,54–0,85 0,67	0,44–0,77 0,56	0,30–0,62 0,45	0,46–0,76 0,58	0,54–0,76 0,62	0,46–0,74 0,56	0,44–0,75 0,57	0,47–0,72 0,60
RE	0,23–0,40 0,29	0,14–0,40 0,24	0,16–0,34 0,22	0,19–0,38 0,27	0,18–0,34 0,25	0,11–0,38 0,19	0,13–0,38 0,23	0,23–0,36 0,27
α	0,94–0,97	0,86–0,92	0,80–0,89	0,77–0,88	0,80–0,92	0,49–0,81	0,75–0,90	0,69–0,87

Adnotacja. PTSD – poziom objawów PTSD, R – ponowne przeżywanie, A – unikanie, NAM – negatywne zmiany w zakresie przekonań i nastroju, AN – anhedonia, EB – zachowania eks-
ternalizacyjne, AA – pobudzenie lękowe, DA – pobudzenie lękowe, C-19 – stresory związane z COVID-19, DEPR – poziom objawów depresyjnych, LĘK – nasilenie objawów lękowych,
STR – nasilenie objawów stresu, RE – reaktywność emocjonalna, α – alfa Cronbacha. Różnice w poziomie rzetelności są istotne statystycznie: PTSD: $\chi^2(7, N = 4064) = 90,83; p < 0,001$;
R: $\chi^2(7, N = 4064) = 52,64; p < 0,001$; A: $\chi^2(7, N = 4064) = 57,29; p < 0,001$; NAM: $\chi^2(7, N = 4064) = 67,70; p < 0,001$; AN: $\chi^2(7, N = 4064) = 111,88; p < 0,001$; EB: $\chi^2(7, N = 4064) = 62,48$;
 $p < 0,001$; AA: $\chi^2(7, N = 4064) = 19,30; p < 0,01$; DA: $\chi^2(7, N = 4064) = 85,38; p < 0,001$.

Tabela 6*Predyktory rozwoju objawów PTSD w ośmiu badanych próbach*

Model	Zmienne	Polska	USA	Japonia	Argentyna	Korea Południowa	Irlandia	Wielka Brytania	Kazachstan
1	Płeć	0,14**	0,18**	-0,03	0,06	0,02	0,07	0,05	0,04
	Wiek	-0,17**	0,01	-0,06	-0,21**	-0,07	0,04	-0,01	-0,03
2	Płeć	0,10**	0,13**	-0,05	0,04	0,01	-0,00	0,04	0,01
	Wiek	-0,14**	0,06	-0,07	-0,19**	-0,06	0,06	-0,00	-0,04
	C-19	0,30**	0,30**	0,21**	0,34**	0,32**	0,36**	0,32**	0,22**
3	Płeć	0,06	0,06	-0,10	0,00	-0,10	-0,06	-0,05	-0,04
	Wiek	-0,11**	0,07	-0,04	-0,17**	-0,02	0,07	-0,02	-0,03
	C-19	0,29**	0,27**	0,17**	0,31**	0,25**	0,34**	0,26**	0,18**
	RE	0,21**	0,21**	0,24**	0,18**	0,37**	0,19**	0,33**	0,22**
4	Płeć	0,06*	0,07	-0,10	0,00	-0,10*	-0,05	-0,04	-0,03
	Wiek	-0,11**	0,06	-0,04	-0,17**	-0,01	0,06	-0,03	-0,03
	C-19	0,28**	0,27**	0,17**	0,34**	0,26**	0,36**	0,15**	0,18**
	RE	0,20**	0,21**	0,24**	0,14*	0,37**	0,19*	0,34**	0,21**
	C-19 x RE	0,06#	0,07*	0,01	0,28**	0,09*	0,12#	0,07	0,04

Adnotacja. C-19 – poziom stresorów związanych z COVID-19, RE – reaktywność emocjonalna, C-19 x RE – interakcja poziomu stresorów związanych z COVID-19 z reaktywnością emocjonalną. Procent wyjaśnionej wariancji (kolejno modele 1–4): Polska: 4,8%, 13,7%, 17,8% i 18,1%; USA: 3,3%, 12%, 15,6% i 16,1%; Japonia: 0,04%, 4,7%, 10% i 10%; Argentyna: 5,2%, 16,4%, 19,3% i 27,1%; Korea Południowa: 0,6%, 11%, 22,6% i 23,5%; Irlandia: 0,6%, 13%, 16,3% i 17,7%; Wielka Brytania: 0,3%, 10,6%, 20,4% i 20,9%; Kazachstan: 0,2%, 5,1%, 9,4% i 9,5%.

$p < 0,10$; * $p < 0,05$; ** $p < 0,01$.

DYSKUSJA

Podstawowym celem przedstawionych przez nas analiz było zweryfikowanie równoważności pomiarowej narzędzia PCL-5, przy uwzględnieniu danych z ośmiu kultur: polskiej, amerykańskiej, japońskiej, argentyńskiej, koreańskiej, irlandzkiej, brytyjskiej oraz kazachskiej. Uwzględniliśmy zatem dane z czterech różnych kontynentów: Europy, Azji, Ameryki Północnej i Ameryki Południowej. Wedle naszej najlepszej wiedzy jest to jedyne dotychczas przeprowadzone badanie, w którym porównano dane dotyczące pomiaru PCL-5 z tak wielu i tak zróżnicowanych kultur.

W pierwszym kroku dokonaliśmy porównania dopasowania poszczególnych modeli opisujących strukturę objawów PTSD w całej próbie oraz przy założeniu jednakowej struktury we wszystkich próbach. Podobnie jak w poprzednich badaniach (Blevins in., 2015; van Praag i in., 2020), model oparty na DSM-5 okazał się dobrze dopasowany do danych, jednak opisywał je słabiej niż modele wyodrębniające więcej niż cztery czynniki. Najlepsze dopasowanie do danych wykazał hybrydowy model 7-czynnikowy zaproponowany przez Armour i in. (2015), co jest spójne z wynikami uzyskanymi przez Ashbaugh i in. (2016), Blevins i in. (2015), Caldas i in. (2020), Contractor i in. (2018), Forkus i in. (2023), van Praag i in. (2020). Weryfikację równoważności pomiarowej na kolejnych poziomach (metrycznym i skalarnym) przeprowadziliśmy względem tego modelu.

Uzyskaliśmy potwierdzenie dla równoważności pomiarowej w zakresie jednakowej struktury oraz jednakowych ładunków czynnikowych, a zatem na poziomie konfiguralnym oraz metrycznym. Nie potwierdziliśmy natomiast równoważności w zakresie stałych regresji (co stanowiło brak uzasadnienia dla liczenia równoważności rezydualnej – w zakresie błędów pomiarów).

Potwierdzenie równoważności konfiguralnej oraz metrycznej stanowi metodologiczny warunek uprawniający do badania relacji pomiędzy zmiennymi w różnych grupach (Lubiewska i Głogowska, 2018; Putnick i Bornstein, 2016; Schmitt i Kuljanin, 2008). Na podstawie uzyskanych wyników możemy zatem stwierdzić, że różne kulturowe wersje PCL-5 są porównywalne w obrębie danej kultury i w każdej z nich badają ten sam konstrukt, rozumiany w ten sam sposób. Brak potwierdzenia równoważności skalarnej nie pozwala jednak na dokonywanie porównań międzykulturowych, w tym na porównania średnich.

Uzyskany wynik nie jest zaskakujący. Choć badań nad równoważnością pomiarową PCL-5 oraz innych narzędzi do pomiaru PTSD nie ma na tę chwilę

wiele, na podstawie tych dostępnych można pokusić się o wstępne wnioski. Badania, w których dokonano porównań pomiędzy dwiema grupami dość jednorodnymi pod względem kulturowym, pozwoliły na potwierdzenie równoważności na poziomie skalarnym, a nawet rezydualnym (wspomniane powyżej badania Caldas i in., 2020; Contractor i in., 2018, jak również Tay i in., 2017, w odniesieniu do grup płciowych i dwóch grup etnicznych mieszkańców Sri-Lanki i *Harwardzkiego kwestionariusza do pomiaru traumy* [HTG]; *Harvard Trauma Questionnaire*). Dla porównania Rasmussen i in. (2015) potwierdzili jedynie równoważność konfiguralną HTG w wielokulturowej i wielojęzycznej próbie uchodźców. Z kolei w pięciu różnych grupach językowych/kulturowych uchodźców Wind i in. (2017) potwierdzili tylko równoważność konfiguralną i częściową równoważność metryczną (w odniesieniu do niektórych pozycji) tego samego narzędzia. A zatem wydaje się, że ze względu na zróżnicowanie prób pod względem czynników biologicznych i środowiskowych, jak również w odniesieniu do poziomu traumatyzacji, oczekiwanie równości w zakresie średnich latentnych pomiędzy kulturowo różnorodnymi próbami może okazać się nie do zrealizowania. Odrębnym argumentem pozostają także ograniczenia metodologiczne. Jak wykazali Rutkowski i Svetina (2014), potwierdzenie równoważności pomiarowej w większej liczbie porównywanych grup jest zadaniem trudniejszym i wymaga przyjęcia bardziej liberalnych kryteriów. W naszym badaniu uwzględniliśmy dane aż z ośmiu państw ułożonych na czterech różnych kontynentach, przy czym próby te były nieporównywalne pod względem liczebności i danych demograficznych. Nie były to też próby jednorodne pod względem poziomu traumatyzacji czy rodzaju doświadczonej traumy. Mimo tak dużego zróżnicowania, pomiar PCL-5 okazał się porównywalny pod względem struktury, jak i ładunków czynnikowych pozycji. W celu dalszej weryfikacji równoważności na poziomie skalarnym przyszłe badania warto byłoby przeprowadzić w próbach jak najbardziej zróżnicowanych kulturowo, a jednocześnie podobnych pod względem demograficznym, przy zachowaniu kontroli doświadczonych wydarzeń traumatycznych.

Relacje objawów PTSD z innymi uwzględnionymi zmiennymi, tj. poziomem stresorów związanych z COVID-19, reaktywnością emocjonalną, poziomem objawów depresyjnych, lękowych i stresu były w badanych próbach analogiczne, różniąc się jedynie siłą. Zgodnie z oczekiwaniami były to korelacje dodatnie, silniejsze w odniesieniu do objawów depresyjnych, lękowych i stresu (które częściowo pokrywają się z objawami PTSD), a słabsze w odniesieniu do reaktywności emocjonalnej. Wyniki te są spójne z tymi uzyskanymi we wcześniej-

szych badaniach (por. Cyniak-Cieciura i Zawadzki, 2021; Forkus i in., 2023) i potwierdzają trafność zbieżną pomiaru objawów PTSD za pomocą PCL-5.

Ograniczenia odnośnie do wniosków wyciąganych z naszego badania zostały już częściowo przedstawione. Dotyczą one przede wszystkim nierównoliczności i niejednorodności zbadanych prób, szczególnie w odniesieniu do płci, wieku i wykształcenia osób badanych (które nie zostało skontrolowane w próbie polskiej).

Podsumowując, nasze analizy przeprowadzone na danych z ośmiu różnych kultur pozwalają stwierdzić, że PCL-5, zgodnie z założeniami klasyfikacji DSM-5, mierzy uniwersalne kulturowo wskaźniki PTSD. Jednocześnie, niezgodnie z założeniami te same klasyfikacji, dane lepiej grupują się w inną strukturę niż ta zaproponowana w DSM-5, co po raz kolejny stawia pod znakiem zapytania adekwatność zaproponowanych kryteriów diagnostycznych. Badaczom, którzy chcieliby wykorzystywać PCL-5 do badań międzykulturowych, proponujemy każdorazowe zweryfikowanie równoważności pomiaru w uwzględnionych grupach w celu potwierdzenia ekwiwalentności na poziomie skalarnym.

BIBLIOGRAFIA

- American Psychiatric Association (1980). *Diagnostic and statistical manual of mental disorders* (wyd. 3).
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (wyd. 4).
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (wyd. 5). <https://doi.org/10.1176/appi.books.9780890425596>
- Arcaya, M. C., Lowe, S. R., Asad, A. L., Subramanian, S. V., Waters, M. C., Rhodes, J. (2017). Association of posttraumatic stress disorder symptoms with migraine and headache after the natural disaster. *Health Psychology*, 36(5), 411–418. <https://doi.org/10.1037/hea0000433>
- Armour, C., Müllerová, J., Elhai, J. D. (2016). A systematic literature review of PTSD's latent structure in the Diagnostic and Statistical Manual of Mental Disorders: DSM-IV to DSM-5. *Clinical Psychology Review*, 44, 60–74. <https://doi.org/10.1016/j.cpr.2015.12.003>
- Armour, C., Tsai, J., Durham, T. A., Charak, R., Biehn, T. L., Elhai, J. D., Pietrzak, R. H. (2015). Dimensional structure of DSM-5 posttraumatic stress symptoms: Support for a hybrid Anhedonia and Externalizing Behaviors model. *Journal of Psychiatric Research*, 61, 106–113. <https://doi.org/10.1016/j.jpsychires.2014.10.012>

- Ashbaugh, A. R., Houle-Johnson, S., Herbert, C., El-Hage, W. (2016). Psychometric validation of the English and French versions of the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5). *PLoS ONE*, 11(10), artykuł e0161645. <https://doi.org/10.1371/journal.pone.0161645>
- Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress*, 28(6), 489–498. <https://doi.org/10.1002/jts.22059>
- Bockhop, F., Zeldovich, M., Cunitz, K., Van Praag, D., van der Vlegel, M., Beissbarth, T., Hagmayer, Y., von Steinbuechel, N., CENTER-TBI participants and investigators. (2022). Measurement invariance of six language versions of the Post-Traumatic Stress Disorder Checklist for DSM-5 in civilians after traumatic brain injury. *Scientific Reports*, 12(1), artykuł 16571. <https://doi.org/10.1038/s41598-022-20170-2>
- Bramsen, I., van der Ploeg, H. M. (1999). Fifty years later: The long-term psychological adjustment of aging World War II survivors. *Acta Psychiatrica Scandinavica*, 100(5), 350–358. <https://doi.org/10.1111/j.1600-0447.1999.tb10878.x>
- Brewin, C. R., Andrews, B., Valentine, J. D. (2000). Meta-analysis of risk factors for posttraumatic stress disorder in trauma-exposed adults. *Journal of Consulting and Clinical Psychology*, 68(5), 748–766. <https://doi.org/10.1037//0022-006x.68.5.748>
- Caldas, S. V., Contractor, A. A., Koh, S., Wang, L. (2020). Factor structure and multi-group measurement invariance of posttraumatic stress disorder symptoms assessed by the PCL-5. *Journal of Psychopathology and Behavioral Assessment*, 42(2), 364–376. <https://doi.org/10.1007/s10862-020-09800-z>
- Carvalho, T., da Motta, C., Pinto-Gouveia, J. (2020). Portuguese version of the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Comparison of latent models and other psychometric analyses. *Journal of Clinical Psychology*, 76(7), 1267–1282. <https://doi.org/10.1002/jclp.22930>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Cheung, A., Makhachvili, N., Javakhishvili, J., Karachevsky, A., Kharchenko, N., Shpiker, M., Roberts, B. (2019). Patterns of somatic distress among internally displaced persons in Ukraine: Analysis of a cross-sectional survey. *Social Psychiatry and Psychiatric Epidemiology*, 54(10), 1265–1274. <https://doi.org/10.1007/s00127-019-01652-7>
- Contractor, A. A., Caldas, S. V., Dolan, M., Lagdon, S., Armour, C. (2018). PTSD's factor structure and measurement invariance across subgroups with differing count of trauma types. *Psychiatry Research*, 264, 76–84. <https://doi.org/10.1016/j.psychres.2018.03.065>

- Cyniak-Cieciura, M., Popiel, A., Zawadzki, B., Cremeans-Smith, J. K., Alessandri, G., Bielak, P., Camino, V., Cha, E. J., Cho, Y., Dobrowolski, P., Fajkowska, M., Filosa, L., Fruehstorfer, D. B., Galarregui, M., Goldfarb, R., Hyun, M.-H., Kalinina, Z., Keegan, E., Mambetalina, A., [...] Topanova, G. T. (2024). Development of a Culture-Common Formal Characteristics of Behavior – Temperament Markers Inventory (FCB-TMI-CC). *Journal of Personality Assessment*. <https://doi.org/10.1080/00223891.2024.2363967>
- Cyniak-Cieciura, M., Zawadzki, B. (2021). The relationship between temperament traits and post-traumatic stress disorder symptoms and its moderators: Meta-analysis and meta-regression. *Trauma, Violence, & Abuse*, 22(4), 702–716. <https://doi.org/10.1177/1524838019876702>
- Davidson, J. R. T. (2000). Trauma: The impact of post-traumatic stress disorder. *Journal of Psychopharmacology*, 14(2), S5–S12. <https://doi.org/10.1177/026988110001400203>
- Di Tella, M., Romeo, A., Zara, G., Castelli, L., Settanni, M. (2022). The Post-Traumatic Stress Disorder Checklist for DSM-5: Psychometric properties of the Italian version. *International Journal of Environmental Research and Public Health*, 19(9), artykuł 5282. <https://doi.org/10.3390/ijerph19095282>
- Elhai, J. D., Biehn, T. L., Armour, C., Klopfer, J. J., Frueh, B. C., Palmieri, P. A. (2011). Evidence for a unique PTSD construct represented by PTSD's D1-D3 symptoms. *Journal of Anxiety Disorders*, 25(3), 340–345. <https://doi.org/10.1016/j.janxdis.2010.10.007>
- Forkus, S. R., Raudales, A. M., Rafiuddin, H. S., Weiss, N. H., Messman, B. A., Contractor, A. A. (2023). The Posttraumatic Stress Disorder (PTSD) Checklist for DSM-5: A systematic review of existing psychometric evidence. *Clinical Psychology: Science and Practice*, 30(1), 110–121. <https://doi.org/10.1037/cps0000111>
- Goenjian, A. K., Steinberg, A. M., Najarian, L. M., Fairbanks, L. A., Tashjian, M., Pynoos, R. S. (2000). Prospective study of posttraumatic stress, anxiety, and depressive reactions after earthquake and political violence. *American Journal of Psychiatry*, 157(6), 911–916. <https://doi.org/10.1176/appi.ajp.157.6.911>
- Hu, L.-t., Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Ibrahim, H., Ertl, V., Catani, C., Ismail, A. A., Neuner, F. (2018). The validity of Post-traumatic Stress Disorder Checklist for DSM-5 (PCL-5) as screening instrument with Kurdish and Arab displaced populations living in the Kurdistan region of Iraq. *BMC Psychiatry*, 18(1), artykuł 259. <https://doi.org/10.1186/s12888-018-1839-z>
- Ito, M., Takebayashi, Y., Suzuki, Y., Horikoshi, M. (2019). Posttraumatic Stress Disorder Checklist for DSM-5: Psychometric properties in a Japanese population. *Journal of Affective Disorders*, 247, 11–19. <https://doi.org/10.1016/j.jad.2018.12.086>

- Krüger-Gottschalk, A., Knaevelsrud, C., Rau, H., Dyer, A., Schäfer, I., Schellong, J., Ehring, T. (2017). The German version of the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Psychometric properties and diagnostic utility. *BMC Psychiatry*, 17(1), artykuł 379. <https://doi.org/10.1186/s12888-017-1541-6>
- Lee, D., Ku, M., Kwon, W., Kim, S. (2020). A study on reliability and validity of the Korean version of PCL-5 (Posttraumatic Stress Disorder Checklist for DSM-5) for adults. *Korean Journal of Counseling and Psychotherapy*, 32(2), 559–582.
- Liu, P., Wang, L., Cao, C., Wang, R., Zhang, J., Zhang, B., Wu, Q., Zhang, H., Zhao, Z., Fan, G., Elhai, J. D. (2014). The underlying dimensions of DSM-5 posttraumatic stress disorder symptoms in an epidemiological sample of Chinese earthquake survivors. *Journal of Anxiety Disorders*, 28(4), 345–351. <https://doi.org/10.1016/j.janxdis.2014.03.008>
- Lovibond, P. F., Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behavior Research and Therapy*, 33(3), 335–343. [http://doi.org/10.1016/0005-7967\(94\)00075-U](http://doi.org/10.1016/0005-7967(94)00075-U)
- Lubiewska, K., Głogowska, K. (2018). Zastosowanie równoważności pomiarowej w badaniach psychologicznych. *Polskie Forum Psychologiczne*, 23(2), 330–356. <https://doi.org/10.14656/PFP20180207>
- Makara-Studzińska, M., Załuski, M., Adamczyk, K., Tyburski, E. (2022). Polish version of the Depression Anxiety Stress Scale (DASS-42) – adaptation and normalization. *Psychiatria Polska*, 58(1), 63–78. <https://doi.org/10.12740/PP/OnlineFirst/153064>
- Martínez-Levy, G. A., Bermúdez-Gómez, J., Merlín-García, I., Flores-Torres, R. P., Nani, A., Cruz-Fuentes, C. S., Briones-Velasco, M., Ortiz-León, S., Mendoza-Velásquez, J. (2021). After a disaster: Validation of PTSD Checklist for DSM-5 and the four- and eight-item abbreviated versions in mental health service users. *Psychiatry Research*, 305, artykuł 114197. <https://doi.org/10.1016/j.psychres.2021.114197>
- McCormack, L. (2009). Civilian women at war: Psychological impact decades after the Vietnam war. *Journal of Loss and Trauma*, 14(6), 447–458. <https://doi.org/10.1080/15325020902925209>
- Ogińska-Bulik, N., Juczyński, Z., Lis-Turlejska, M., Merecz-Kot, D. (2018). Polska adaptacja PTSD Checklist for DSM-5 – PCL-5. Doniesienie wstępne. *Przegląd Psychologiczny*, 61(2), 281–285.
- Ozer, E. J., Best, S. R., Lipsey, T. L., Weiss, D. S. (2003). Predictors of posttraumatic stress disorder and symptoms in adults: A meta-analysis. *Psychological Bulletin*, 129(1), 52–73. <https://doi.org/10.1037/0033-2909.129.1.52>
- Popiel, A., Zawadzki, B., Bielecki, M., Mroziński, B., Pragłowska, E. (2021). Czy doświadczenie pandemii może prowadzić do pourazowego zaburzenia stresowego (PTSD)? Wyniki badania COVID-STRES. W: W. J. Paluchowski i L. Bakiera

- (red.), *Psychospołeczny obraz pierwszej fali pandemii COVID-19 w Polsce* (s. 45–58). Wydawnictwo Naukowe UAM.
- Putnick, D. L., Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org>
- Rao, A. S., Sher, A. I., Viera, R. V., Merikangas, A. L., Peterlin, B. L. (2015). The impact of post-traumatic stress disorder on the burden of migraine: Results from the National Comorbidity Survey-Replication. *Headache*, 55(10), 1323–1341. <https://doi.org/10.1111/head.12698>
- Rasmussen, A., Verkuilen, J., Ho, E., Fan, Y. (2015). Posttraumatic stress disorder among refugees: Measurement invariance of Harvard Trauma Questionnaire scores across global regions and response patterns. *Psychological Assessment*, 27(4), 1160–1170. <https://doi.org/10.1037/pas0000115>
- Rhemtulla, M., Brosseau-Liard, P., Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373.
- Rosseel, Y. (2012). *lavaan*: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rutkowski, L., Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74(1), 31–57. <https://doi.org/10.1177/0013164413498257>
- Schmitt, N., Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210–222. <https://doi.org/10.1016/j.hrmr.2008.03.003>
- Simms, L. J., Watson, D., Doebbeling, B. N. (2002). Confirmatory factor analyses of posttraumatic stress symptoms in deployed and nondeployed veterans of the Gulf War. *Journal of Abnormal Psychology*, 111(4), 637–647. <https://doi.org/10.1037/0021-843X.111.4.637>
- Tay, A. K., Jayasuriya, R., Jayasuriya, D., Silove, D. (2017). Assessing the factorial structure and measurement invariance of PTSD by gender and ethnic groups in Sri Lanka: An analysis of the modified Harvard Trauma Questionnaire (HTQ). *Journal of Anxiety Disorders*, 47, 45–53. <https://doi.org/10.1016/j.janxdis.2017.02.001>
- Tortella-Feliu, M., Fullana, M. A., Pérez-Vigil, A., Torres, X., Chamorro, J., Littarelli, S. A., Solanes, A., Ramella-Cravaro, V., Vilar, A., González-Parra, J. A., Andero, R., Reichenberg, A., Mataix-Cols, D., Vieta, E., Fusar-Poli, P., Ioannidis, J. P. A.,

- Stein, M. B., Radua, J., Fernández de la Cruz, L. (2019). Risk factors for posttraumatic stress disorder: An umbrella review of systematic reviews and meta-analyses. *Neuroscience & Biobehavioral Reviews*, 107, 154–165. <https://doi.org/10.1016/j.neubiorev.2019.09.013>
- Tsai, J., Harpaz-Rotem, I., Armour, C., Southwick, S. M., Krystal, J. H., Pietrzak, R. H. (2014). Dimensional structure of DSM-5 posttraumatic stress disorder symptoms: Results from the National Health and Resilience in Veterans Study. *The Journal of Clinical Psychiatry*, 76(5), 546–553. <https://doi.org/10.4088/JCP.14m09091>
- van Praag, D. L. G., Fardzadeh, H. E., Covic, A., Maas, A. I. R., von Steinbüchel, N. (2020). Preliminary validation of the Dutch version of the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5) after traumatic brain injury in a civilian population. *PLoS ONE*, 15(4), artykuł e0231857. <https://doi.org/10.1371/journal.pone.0231857>
- Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., Schnurr, P. P. (2013). *The PTSD Checklist for DSM-5 (PCL-5)*. <https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp>
- Wind, T. R., van der Aa, N., Knipscheer, J., de la Rie, S. (2017). “The assessment of psychopathology among traumatized refugees: Measurement invariance of the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist-25 across five linguistic groups”: Corrigendum. *European Journal of Psychotraumatology*, 8(S2), artykuł 1321357. <https://doi.org/10.1080/20008198.2017.1321357>

Załączniki

Tabela A1
Statystyki opisowe, zgodność wewnętrzną (alfa Cronbacha) oraz współczynniki korelacji (r-Pearsona)

Próba	Zmienne	M	SD	Skasność (SE)	Kurtoza (SE)	α	PTSD	R	A	NAM	AN	EB	AA	DA	C-19	DEPR	LĘK	STR
PL	PTSD	44,29	18,13	0,50 (0,08)	−0,66 (0,16)	0,95												
	R	10,93	5,14	0,65 (0,08)	−0,50 (0,16)	0,89	0,871**											
	A	4,91	2,45	0,41 (0,08)	−0,91 (0,16)	0,81	0,753**	0,687**										
	NAM	8,88	4,09	0,56 (0,08)	−0,62 (0,16)	0,80	0,872**	0,702**	0,629**									
	AN	6,15	3,34	0,83 (0,08)	−0,36 (0,16)	0,85	0,810**	0,582**	0,481**	0,671**								
	EB	3,87	2,03	1,01 (0,08)	0,24 (0,16)	0,71	0,755**	0,546**	0,443**	0,605**	0,633**							
	AA	4,88	2,44	0,44 (0,08)	−0,95 (0,16)	0,80	0,810**	0,627**	0,535**	0,628**	0,609**	0,649**						
	DA	4,67	2,48	0,57 (0,08)	−0,82 (0,16)	0,82	0,811**	0,603**	0,517**	0,630**	0,663**	0,638**	0,717**					
	C-19	36,50	7,42	0,13 (0,08)	−0,01 (0,16)	0,76	0,329**	0,307**	0,291**	0,302**	0,214**	0,211**	0,265**	0,257**				
	DEPR	26,44	9,81	0,73 (0,08)	−0,20 (0,16)	0,93	0,650**	0,531**	0,443**	0,538**	0,525**	0,519**	0,605**	0,595**	0,339**			
USA	LĘK	27,69	10,27	0,59 (0,08)	−0,47 (0,16)	0,94	0,617**	0,462**	0,404**	0,528**	0,564**	0,502**	0,550**	0,571**	0,304**	0,917**		
	STR	24,85	9,39	0,93 (0,08)	0,19 (0,16)	0,92	0,646**	0,512**	0,434**	0,554**	0,574**	0,499**	0,556**	0,590**	0,318**	0,932**	0,933**	
	RE	15,51	3,78	−0,15 (0,08)	−0,37 (0,16)	0,77	0,276**	0,253**	0,245**	0,251**	0,182**	0,133**	0,253**	0,230**	0,136**	0,319**	0,311**	0,331**
	PTSD	44,56	20,31	0,55 (0,09)	−0,80 (0,18)	0,96												
	R	11,44	5,69	0,65 (0,09)	−0,66 (0,18)	0,91	0,903**											
	A	5,16	2,72	0,34 (0,09)	−1,18 (0,18)	0,89	0,805**	0,742**										
	NAM	8,75	4,45	0,66 (0,09)	−0,67 (0,18)	0,85	0,892**	0,753**	0,689**									
	AN																	
	EB																	
	AA																	

ciąg dalszy tabeli na następnej stronie

Ciąg dalszy tabeli z poprzedniej strony

Próba	Zmienne	M	SD	Skłoność (SE)	Kurtzoza (SE)	α	PTSD	R	A	NAM	AN	EB	AA	DA	C-19	DEPR	LĘK	STR
USA	AN	6,37	3,74	0,80	(0,09)	-0,65	(0,18)	0,92	0,868"	0,696"	0,610"	0,729"						
	EB	3,55	1,99	1,19	(0,09)	0,49	(0,18)	0,74	0,779"	0,624"	0,521"	0,666"	0,700"					
	AA	4,60	2,52	0,60	(0,09)	-0,81	(0,18)	0,83	0,783"	0,649"	0,580"	0,641"	0,610"	0,618"				
	DA	4,78	2,71	0,55	(0,09)	-1,04	(0,18)	0,87	0,834"	0,665"	0,580"	0,681"	0,775"	0,648"				
	C-19	41,14	9,78	-0,02	(0,09)	-0,14	(0,18)	0,88	0,321"	0,312"	0,236"	0,271"	0,306"	0,288"	0,216"	0,249"		
	DEPR	28,90	8,73	0,49	(0,09)	-0,29	(0,18)	0,89	0,559"	0,468"	0,400"	0,511"	0,507"	0,491"	0,476"	0,487"	0,353"	
	LĘK	29,37	9,46	0,50	(0,09)	-0,39	(0,18)	0,92	0,535"	0,416"	0,379"	0,497"	0,533"	0,469"	0,427"	0,476"	0,313"	0,907"
	STR	27,14	9,23	0,67	(0,09)	-0,28	(0,18)	0,91	0,542"	0,442"	0,376"	0,490"	0,535"	0,464"	0,443"	0,470"	0,321"	0,930"
	RE	16,16	3,63	-0,30	(0,09)	-0,14	(0,18)	0,74	0,285"	0,268"	0,248"	0,256"	0,240"	0,146"	0,205"	0,267"	0,234"	0,430"
	PTSD	36,08	15,21	0,90	(0,14)	0,13	(0,29)	0,94										0,410"
JAP	R	8,95	4,38	1,07	(0,14)	0,19	(0,29)	0,86	0,881"									
	A	4,30	2,49	0,80	(0,14)	-0,55	(0,29)	0,87	0,712"	0,624"								
	NAM	7,54	3,70	0,99	(0,14)	0,06	(0,29)	0,77	0,878"	0,723"	0,548"							
	AN	5,22	2,78	1,31	(0,14)	1,01	(0,29)	0,80	0,820"	0,619"	0,430"	0,686"						
	EB	2,65	1,14	2,23	(0,14)	5,51	(0,29)	0,49	0,657"	0,474"	0,276"	0,563"	0,640"					
	AA	3,87	2,09	0,98	(0,14)	-0,08	(0,29)	0,75	0,778"	0,594"	0,512"	0,607"	0,590"	0,502"				
	DA	3,67	2,02	1,27	(0,14)	0,94	(0,29)	0,69	0,771"	0,582"	0,448"	0,589"	0,644"	0,540"	0,648"			
	C-19	31,61	5,81	0,20	(0,12)	-0,37	(0,24)	0,74	0,189"	0,170"	0,200"	0,142"	0,165"	0,095	0,104	0,176"		
	DEPR	21,24	6,43	1,06	(0,14)	0,59	(0,28)	0,87	0,660"	0,547"	0,340"	0,604"	0,581"	0,514"	0,567"	0,565"	0,179"	
	LĘK	23,14	8,21	0,99	(0,14)	0,21	(0,28)	0,92	0,702"	0,563"	0,330"	0,656"	0,676"	0,559"	0,564"	0,609"	0,205"	0,888"
	STR	21,50	6,79	1,02	(0,14)	0,35	(0,28)	0,88	0,672"	0,539"	0,300"	0,634"	0,660"	0,550"	0,527"	0,571"	0,168"	0,865"
	RE	17,10	3,52	-0,42	(0,13)	-0,09	(0,26)	0,73	0,254"	0,226"	0,173"	0,218"	0,224"	0,111	0,193"	0,225"	0,124"	0,339"
																	0,310"	0,311"

Ciąg dalszy tabeli na następnej stronie

Ciąg dalszy tabeli z poprzedniej strony

Proba	Zmienne	M	SD	Skłonność (SE)	Kurtosis (SE)	α	PTSD	R	A	NAM	AN	EB	AA	DA	C-19	DEPR	LĘK	STR
ARG	PTSD	37,43	16,05	1,28 (0,13)	1,28 (0,26)	0,95	0,850 ^{***}											
	R	9,77	4,68	1,08 (0,13)	0,31 (0,26)	0,88	0,811 ^{***}											
	A	4,18	2,26	0,86 (0,13)	-0,26 (0,26)	0,80	0,861 ^{***}	0,680 ^{***}										
	NAM	7,55	3,84	1,30 (0,13)	1,21 (0,26)	0,84	0,842 ^{***}	0,626 ^{***}	0,714 ^{***}									
	AN	5,02	2,85	1,76 (0,13)	2,72 (0,26)	0,90	0,721 ^{***}	0,578 ^{***}	0,580 ^{***}	0,690 ^{***}								
	EB	3,01	1,55	1,91 (0,13)	3,63 (0,26)	0,63	0,851 ^{***}	0,476 ^{***}	0,455 ^{***}	0,720 ^{***}								
	AA	4,15	2,12	1,03 (0,13)	0,24 (0,26)	0,82	0,801 ^{***}	0,670 ^{***}	0,622 ^{***}	0,664 ^{***}	0,689 ^{***}	0,664 ^{***}						
	DA	3,75	2,00	1,15 (0,13)	0,68 (0,26)	0,70	0,350 ^{***}	0,600 ^{***}	0,573 ^{***}	0,598 ^{***}	0,709 ^{***}	0,582 ^{***}	0,723 ^{***}					
	C-19	41,66	9,06	0,32 (0,10)	-0,39 (0,19)	0,88	0,689 ^{***}	0,367 ^{***}	0,237 ^{***}	0,266 ^{***}	0,268 ^{***}	0,232 ^{***}	0,304 ^{***}	0,288 ^{***}				
	DEPR	22,53	7,72	1,40 (0,11)	2,17 (0,21)	0,92	0,706 ^{***}	0,599 ^{***}	0,520 ^{***}	0,523 ^{***}	0,580 ^{***}	0,507 ^{***}	0,651 ^{***}	0,622 ^{***}	0,382 ^{***}			
KPD	LĘK	23,33	8,50	1,42 (0,11)	2,02 (0,21)	0,93	0,732 ^{***}	0,572 ^{***}	0,546 ^{***}	0,556 ^{***}	0,670 ^{***}	0,514 ^{***}	0,623 ^{***}	0,618 ^{***}	0,345 ^{***}	0,909 ^{***}		
	STR	20,65	7,53	1,91 (0,11)	4,09 (0,21)	0,92	0,850 ^{***}	0,606 ^{***}	0,574 ^{***}	0,575 ^{***}	0,686 ^{***}	0,528 ^{***}	0,639 ^{***}	0,630 ^{***}	0,359 ^{***}	0,898 ^{***}	0,938 ^{***}	
	RE	14,25	3,14	0,19 (0,10)	-0,13 (0,19)	0,73	0,257 ^{***}	0,214 ^{***}	0,206 ^{***}	0,254 ^{***}	0,231 ^{***}	0,108 ^{***}	0,174 ^{***}	0,239 ^{***}	0,151 ^{***}	0,383 ^{***}	0,376 ^{***}	0,407 ^{***}
	PTSD	21,41	17,73	0,63 (0,10)	-0,49 (0,21)	0,97												
	R	5,45	4,91	0,68 (0,10)	-0,47 (0,21)	0,92	0,915 ^{***}											
	A	2,68	2,21	0,44 (0,10)	-0,76 (0,21)	0,89	0,807 ^{***}	0,728 ^{***}										
	NAM	4,22	3,78	0,62 (0,10)	-0,50 (0,21)	0,88	0,917 ^{***}	0,784 ^{***}	0,723 ^{***}									
	AN	3,43	3,23	0,68 (0,10)	-0,52 (0,21)	0,92	0,890 ^{***}	0,758 ^{***}	0,668 ^{***}	0,776 ^{***}								
	EB	1,44	1,80	1,15 (0,10)	0,41 (0,21)	0,78	0,827 ^{***}	0,706 ^{***}	0,574 ^{***}	0,755 ^{***}	0,714 ^{***}							
	AA	1,91	2,11	0,92 (0,10)	-0,08 (0,21)	0,85	0,870 ^{***}	0,728 ^{***}	0,622 ^{***}	0,775 ^{***}	0,752 ^{***}	0,745 ^{***}						
C-19	DA	2,28	2,16	0,66 (0,10)	-0,53 (0,21)	0,79	0,832 ^{***}	0,693 ^{***}	0,605 ^{***}	0,716 ^{***}	0,725 ^{***}	0,650 ^{***}	0,773 ^{***}					
	C-19	37,84	8,71	0,68 (0,10)	0,21 (0,21)	0,91	0,326 ^{***}	0,312 ^{***}	0,277 ^{***}	0,295 ^{***}	0,281 ^{***}	0,283 ^{***}	0,278 ^{***}	0,243 ^{***}				

Ciąg dalszy tabeli na następnej stronie

Ciąg dalszy tabeli z poprzedniej strony

Próba	Zmienne	M	SD	Skóśność (SE)	Kurtzoza (SE)	α	PTSD	R	A	NAM	AN	EB	AA	DA	C-19	DEPR	LĘK	STR
KPD	DEPR	24,46	8,29	0,96	(0,10)	0,73	(0,21)	0,94	0,829 ^{***}	0,751 ^{***}	0,621 ^{***}	0,745 ^{***}	0,730 ^{***}	0,736 ^{***}	0,712 ^{***}	0,331 ^{***}		
	LĘK	25,53	8,81	0,88	(0,10)	0,53	(0,21)	0,95	0,837 ^{***}	0,748 ^{***}	0,620 ^{***}	0,756 ^{***}	0,768 ^{***}	0,712 ^{***}	0,728 ^{***}	0,310 ^{***}	0,953 ^{***}	
	STR	23,79	8,42	1,04	(0,10)	0,87	(0,21)	0,94	0,840 ^{***}	0,765 ^{***}	0,616 ^{***}	0,757 ^{***}	0,763 ^{***}	0,736 ^{***}	0,749 ^{***}	0,314 ^{***}	0,961 ^{***}	
	RE	15,81	2,66	0,10	(0,10)	0,40	(0,21)	0,70	0,395 ^{***}	0,336 ^{***}	0,343 ^{***}	0,378 ^{***}	0,343 ^{***}	0,265 ^{***}	0,380 ^{***}	0,213 ^{***}	0,401 ^{***}	0,384 ^{***}
IRL	PTSD	40,72	17,91	0,78	(0,17)	-0,22	(0,33)	0,96										
	R	10,26	4,93	0,83	(0,17)	-0,21	(0,33)	0,88	0,902 ^{***}									
	A	4,69	2,47	0,53	(0,17)	-0,92	(0,33)	0,85	0,752 ^{***}	0,716 ^{***}								
	NAM	8,44	4,12	0,76	(0,17)	-0,36	(0,33)	0,84	0,883 ^{***}	0,736 ^{***}	0,612 ^{***}							
	AN	5,67	3,21	1,16	(0,17)	0,30	(0,33)	0,88	0,860 ^{***}	0,681 ^{***}	0,475 ^{***}	0,767 ^{***}						
	EB	3,34	1,89	1,53	(0,17)	1,72	(0,33)	0,77	0,803 ^{***}	0,643 ^{***}	0,494 ^{***}	0,661 ^{***}	0,767 ^{***}					
	AA	4,22	2,35	0,81	(0,17)	-0,47	(0,33)	0,84	0,777 ^{***}	0,653 ^{***}	0,558 ^{***}	0,568 ^{***}	0,600 ^{***}	0,620 ^{***}				
	DA	4,10	2,29	0,92	(0,17)	-0,19	(0,33)	0,79	0,811 ^{***}	0,650 ^{***}	0,512 ^{***}	0,657 ^{***}	0,718 ^{***}	0,635 ^{***}	0,663 ^{***}			
	C-19	38,38	8,76	0,18	(0,17)	-0,06	(0,33)	0,85	0,355 ^{***}	0,356 ^{***}	0,276 ^{***}	0,271 ^{***}	0,266 ^{***}	0,287 ^{***}	0,324 ^{***}	0,283 ^{***}		
	DEPR	28,58	8,24	0,45	(0,17)	-0,22	(0,33)	0,91	0,582 ^{***}	0,492 ^{***}	0,343 ^{***}	0,473 ^{***}	0,515 ^{***}	0,505 ^{***}	0,549 ^{***}	0,570 ^{***}	0,309 ^{***}	
UK	LĘK	29,58	8,67	0,40	(0,17)	-0,26	(0,33)	0,93	0,542 ^{***}	0,422 ^{***}	0,300 ^{***}	0,412 ^{***}	0,538 ^{***}	0,516 ^{***}	0,499 ^{***}	0,576 ^{***}	0,296 ^{***}	0,886 ^{***}
	STR	27,84	8,72	0,44	(0,17)	-0,37	(0,33)	0,92	0,575 ^{***}	0,482 ^{***}	0,321 ^{***}	0,463 ^{***}	0,551 ^{***}	0,494 ^{***}	0,506 ^{***}	0,584 ^{***}	0,313 ^{***}	0,926 ^{***}
	RE	16,22	3,44	-0,17	(0,17)	-0,37	(0,33)	0,72	0,232 ^{***}	0,140 ^{***}	0,164 ^{***}	0,282 ^{***}	0,228 ^{***}	0,166 ^{***}	0,133	0,233 ^{***}	0,180 ^{***}	0,416 ^{***}
	PTSD	46,87	20,87	0,33	(0,15)	-0,94	(0,29)	0,97										0,436 ^{***}
AN	R	12,13	5,65	0,24	(0,15)	-1,14	(0,29)	0,92	0,910 ^{***}									
	A	5,05	2,44	0,30	(0,15)	-1,02	(0,29)	0,83	0,832 ^{***}	0,763 ^{***}								
	NAM	9,39	4,46	0,37	(0,15)	-1,08	(0,29)	0,85	0,920 ^{***}	0,783 ^{***}	0,709 ^{***}							
	AN	6,79	3,51	0,44	(0,15)	-1,05	(0,29)	0,88	0,897 ^{***}	0,735 ^{***}	0,722 ^{***}	0,803 ^{***}						
	PTSD																	

Ciąg dalszy tabeli na następnej stronie

Ciąg dalszy tabelizacji poprzedniej strony

Proba	Zmienne	M	SD	Skłonność (SE)	Kurtozja (SE)	α	PTSD	R	A	NAM	AN	EB	AA	DA	C-19	DEPR	LĘK	STR
UK	EB	4,15	2,37	0,84	(0,15)	-0,49	(0,29)	0,81	0,881 ^{***}	0,743 ^{***}	0,683 ^{***}	0,790 ^{***}	0,792 ^{***}					
	AA	4,65	2,52	0,51	(0,15)	-0,99	(0,29)	0,84	0,873 ^{***}	0,738 ^{***}	0,641 ^{***}	0,803 ^{***}	0,749 ^{***}	0,770 ^{***}				
	DA	4,70	2,50	0,50	(0,15)	-0,97	(0,29)	0,86	0,865 ^{***}	0,719 ^{***}	0,676 ^{***}	0,747 ^{***}	0,784 ^{***}	0,765 ^{***}	0,774 ^{***}			
	C-19	38,80	9,20	0,05	(0,15)	-0,07	(0,29)	0,87	0,325 ^{***}	0,267 ^{***}	0,259 ^{***}	0,280 ^{***}	0,308 ^{***}	0,292 ^{***}	0,319 ^{***}	0,329 ^{***}		
	DEPR	28,24	9,23	0,32	(0,15)	-0,72	(0,29)	0,93	0,714 ^{***}	0,637 ^{***}	0,553 ^{***}	0,627 ^{***}	0,608 ^{***}	0,693 ^{***}	0,660 ^{***}	0,692 ^{***}	0,297 ^{***}	
	LĘK	29,35	9,59	0,25	(0,15)	-0,65	(0,29)	0,94	0,661 ^{***}	0,572 ^{***}	0,530 ^{***}	0,580 ^{***}	0,613 ^{***}	0,611 ^{***}	0,581 ^{***}	0,649 ^{***}	0,223 ^{***}	0,922 ^{***}
	STR	27,93	9,68	0,27	(0,15)	-0,95	(0,29)	0,93	0,716 ^{***}	0,619 ^{***}	0,548 ^{***}	0,642 ^{***}	0,663 ^{***}	0,674 ^{***}	0,632 ^{***}	0,693 ^{***}	0,265 ^{***}	0,940 ^{***}
	RE	15,83	3,37	-0,24	(0,15)	-0,17	(0,29)	0,72	0,366 ^{***}	0,315 ^{***}	0,229 ^{***}	0,356 ^{***}	0,327 ^{***}	0,383 ^{***}	0,315 ^{***}	0,344 ^{***}	0,188 ^{***}	0,405 ^{***}
	PTSD	37,27	16,92	1,09	(0,11)	0,73	(0,22)	0,96										0,360 ^{***}
	R	9,22	4,51	1,15	(0,11)	0,89	(0,22)	0,88	0,895 ^{***}									
KAZ	A	4,04	2,31	0,93	(0,11)	-0,17	(0,22)	0,86	0,825 ^{***}	0,755 ^{***}								
	NAM	7,61	4,06	1,18	(0,11)	0,68	(0,22)	0,88	0,930 ^{***}	0,806 ^{***}	0,790 ^{***}							
	AN	5,51	3,00	1,24	(0,11)	0,86	(0,22)	0,87	0,849 ^{***}	0,654 ^{***}	0,613 ^{***}	0,738 ^{***}						
	EB	3,40	1,78	1,42	(0,11)	1,61	(0,22)	0,68	0,791 ^{***}	0,606 ^{***}	0,538 ^{***}	0,664 ^{***}	0,708 ^{***}					
	AA	3,71	1,96	1,15	(0,11)	0,63	(0,22)	0,80	0,850 ^{***}	0,686 ^{***}	0,624 ^{***}	0,767 ^{***}	0,696 ^{***}	0,720 ^{***}				
	DA	3,78	2,01	1,11	(0,11)	0,47	(0,22)	0,76	0,781 ^{***}	0,614 ^{***}	0,503 ^{***}	0,648 ^{***}	0,680 ^{***}	0,692 ^{***}	0,693 ^{***}			
	C-19	32,93	8,44	0,42	(0,11)	0,36	(0,22)	0,86	0,244 ^{***}	0,237 ^{***}	0,224 ^{***}	0,204 ^{***}	0,204 ^{***}	0,153 ^{***}	0,205 ^{***}	0,212 ^{***}		
	DEPR	27,10	9,32	0,75	(0,11)	0,29	(0,22)	0,92	0,565 ^{***}	0,467 ^{***}	0,412 ^{***}	0,512 ^{***}	0,491 ^{***}	0,498 ^{***}	0,505 ^{***}	0,535 ^{***}	0,214 ^{***}	
	LĘK	28,04	9,78	0,74	(0,11)	0,18	(0,22)	0,93	0,602 ^{***}	0,488 ^{***}	0,459 ^{***}	0,547 ^{***}	0,552 ^{***}	0,513 ^{***}	0,516 ^{***}	0,553 ^{***}	0,194 ^{***}	0,930 ^{***}
	STR	25,91	9,61	0,90	(0,11)	0,28	(0,22)	0,93	0,598 ^{***}	0,494 ^{***}	0,454 ^{***}	0,538 ^{***}	0,550 ^{***}	0,503 ^{***}	0,510 ^{***}	0,554 ^{***}	0,165 ^{***}	0,922 ^{***}
KAZ	RE	15,63	3,28	-0,01	(0,11)	0,37	(0,22)	0,75	0,234 ^{***}	0,181 ^{***}	0,164 ^{***}	0,192 ^{***}	0,240 ^{***}	0,197 ^{***}	0,262 ^{***}	0,223 ^{***}	0,299 ^{***}	0,305 ^{***}
	PTSD	37,27	16,92	1,09	(0,11)	0,73	(0,22)	0,96										0,360 ^{***}

Adnotacja. M – średnia, SD – odchylenie standardowe, SE – błąd standardowy, α – alfa Cronbacha, PTSD – poziom objawów PTSD, R – ponowne przeżywanie, A – unikanie, NAM – negatywne zmiany w zakresie przekonań i nastrojów, AN – anhedonia, EB – zachowania eksternalizacyjne, AA – pobudzenie lękowe, DA – pobudzenie lękowe, C-19 – stresory związane z COVID-19, DEPR – poziom objawów depresyjnych, LĘK – nasilenie objawów lękowych, STR – nasilenie objawów stresu, RE – reaktywność emocjonalna, PL – Polska, USA – Stany Zjednoczone Ameryki Północnej, JAP – Japonia, ARG – Argentyna, KPD – Korea Południowa, IRL – Irlandia, UK – Wielka Brytania, KAZ – Kazachstan.

* $p < 0,05$, ** $p < 0,01$.

BOGDAN ZAWADZKI  <https://orcid.org/0000-0002-6578-8412>

Wydział Psychologii, Uniwersytet Warszawski

SZYMON BARCIŚ  <https://orcid.org/0009-0000-6942-992X>

Wydział Psychologii, Uniwersytet Warszawski

MARIA CYNIAK-CIECIURA  <https://orcid.org/0000-0002-6293-8729>

Centrum Badań Klinicznych i Doskonalenia Psychoterapii, Instytut Psychologii, Uniwersytet SWPS

Struktura objawów PTSD – analiza równoważności wersji inwentarza PCL-5 o stałej i losowej kolejności pozycji

The structure of PTSD symptoms – analysis of the measurement equivalence of the PCL-5 versions with fixed and random item order

Autor korespondencyjny: Bogdan Zawadzki
e-mail: bogdan@psych.uw.edu.pl

Sposób cytowania: Zawadzki, B., Barciś, S., Cyniak-Cieciura, M. (2024). Struktura objawów PTSD – analiza równoważności wersji inwentarza PCL-5 o stałej i losowej kolejności pozycji. W: B. Zawadzki, K. Fronczyk i K. Bargiel-Matusiewicz (red.), *Wybrane problemy współczesnej diagnostyki klinicznej w psychologii* (s. 217–256). Wydawnictwo Liberi Libri. <https://doi.org/10.47943/lib.9788363487676.rozdzial08>

Streszczenie

Celem pracy była analiza struktury objawów pourazowego zaburzenia stresowego, badanych za pomocą wersji o losowej i stałej kolejności pozycji *PTSD Checklist dla DSM-5*. Oczekiwano potwierdzenia hybrydowego modelu 7-czynnikowego, obejmującego objawy intruzji, unikania, negatywnego afektu, anhedonii, zachowań eksternalizacyjnych, pobudzenia lękowego i dysforycznego, niezależnie od porządku prezentacji pozycji. Badania zrealizowano w grupie 740 osób, które doznały różnego rodzaju traumatycznych wydarzeń i wypełniły dwukrotnie (w odstępie dwóch tygodni) *PTSD Checklist dla DSM-5*: w planie wersja stała-losowa albo wersja losowa-stała. Analizy wykazały pełną równoważność obu pomiarów, ale jedynie równoważność konfiguralną dla wersji o odmiennej kolejności pozycji. Stwierdzono, że efekty te są pochodną lepszego różnicowania treści pozycji, prowadzącego do nieco większej spójności odpowiadania na pozycje czynnikowo jednorodne niż na pozycje czynnikowo niejednorodne w wersji stałej niż losowej (dla której stwierdzono odwrotny wzór spójności odpowiedzi). W dyskusji wskazano implikacje tych wyników dla diagnozowania pourazowego zaburzenia stresowego, w tym zwłaszcza dla stosowania stałej oraz losowej kolejności pozycji.

Słowa kluczowe: pourazowe zaburzenie stresowe, struktura objawów PTSD, *PTSD Checklist dla DSM-5*, stała i losowa kolejność pozycji

Abstract

The study aimed at analyzing the structure of post-traumatic stress disorder symptoms, assessed by the randomized and fixed-order versions of the PTSD Checklist for DSM-5. It was expected that data support the hybrid 7-factor model, comprising the symptoms of intrusion, avoidance, negative affect, anhedonia, externalizing behaviors, anxiety and dysphoric arousal, regardless of the order of the items. The research was conducted in the group of 740 subjects who experienced various sorts of

traumatic events and completed the PTSD Checklist for DSM-5 twice (with interval of two weeks): either the fixed-random or the random-fixed version. The analyzes showed full equivalence of both assessments, but only configural invariance was found for different item order. These effects are due to the better differentiation of the item content, leading to a slightly greater consistency of responses to factor-homogenous items than to factor-heterogeneous items in the fixed version in comparison to random version (for which the opposite pattern of response consistency was found). The discussion underlines the implications of these results for diagnosing post-traumatic stress disorder, especially for the use of a fixed and random order of items.

Keywords: posttraumatic stress disorder, structure of PTSD symptoms, PTSD Checklist for DSM-5, fixed and random items order

Metody kwestionariuszowe bazujące na relacji o zachowaniu są powszechnie stosowane przez psychologów, pomimo świadomości obciążenia samoopisu wieloma czynnikami zniekształcającymi, wynikającymi przykładowo z aprobaty społecznej, stylów odpowiadania itp. (Zawadzki, 2006). Jednym z podnoszonych problemów jest zależność udzielanych na pozycje kwestionariuszowe odpowiedzi od kontekstu, w tym od treści/zabarwienia emocjonalnego innych pozycji, ich kolejności lub porządku ułożenia w kwestionariuszu (Brzeziński, 1980; Knowles, 1988; Zawadzki, 2006). Badania w tym nurcie dotyczyły znaczenia grupowania vs losowania (a wcześniej przypadkowego rozmieszczenia, które w wersjach „papier-ołówek” odpowiadały randomizacji) pozycji. Schriesheim (1981) sądził, że pogrupowanie pozycji w skale (w wielowymiarowych inwentarzach) prowadzi do uzyskania ich większej trafności zbieżnej i wyższej rzetelności pomiaru, kosztem obniżonej trafności różnicowej w stosunku do narzędzi o losowej kolejności pozycji. Można sądzić, że efekt ten wynika z nadspójności odpowiedzi (w istocie będącej artefaktem badawczym), która jest spowodowana czynnikami zniekształcającymi w postaci pozytywnej autoprezentacji. Schriesheim zdecydowanie opowiadał się (zwłaszcza w kontekście zmiennych odzwierciedlających poziom funkcjonowania zawodowego)

za randomizacją pozycji (Schriesheim, Kopelman i Solomon, 1989). Inne badania nie wykazały jednak systematycznych efektów grupowania-losowania i dały stosunkowo sprzeczne wyniki zarówno w odniesieniu do rzetelności pomiaru, jak też siły interkorelacji skal (Franke, 1997; Solomon i Kopelman, 1984), jednak spójnie wskazywały na odmienności charakterystyki psychometrycznej badań z zastosowaniem obu wersji kwestionariuszy. Kolejne badania Schriesheima i in. (Schriesheim, Solomon i Kopelman, 1989) wykazały jednak, że różnice te są bardzo nieznaczne w zakresie rzetelności i trafności zbieżnej pomiaru, zaś efekt obniżenia trafności różnicowej pomiaru nie zawsze jest rejestrowany. Wynikiem uzupełniającym było stwierdzenie, że dla silnych pomiarowo skal różnice dla obu wersji były minimalne, zaś zalety grupowania uwidaczniały się dla skal, których charakterystyka psychometryczna była wyjściowo słaba. W efekcie sugestia końcowa dotyczyła możliwości grupowania pozycji raczej w przypadku słabych skal, w celu wzmocnienia charakterystyki psychometrycznej pomiaru (rzetelności, trafności zbieżnej, a nawet różnicowej; por. Schell i Oswald, 2013). Kolejne, bardziej współczesne badania wskazują, iż czynnik ten nie ma istotnego znaczenia dla przypadku kwestionariuszowego badania cech osobowości (Schell i Oswald, 2013). Kwestia wpływu grupowania-losowania pozycji na pomiar nie została jednak jednoznacznie rozstrzygnięta, a jej istotność wzrasta wraz z popularyzacją badań w formie online poprzez platformy badawcze umożliwiające zarówno pełne randomizowanie kolejności pozycji kwestionariuszowych (wersja losowa dla każdego badania), jak też zamieszczanie pozycji w porządku stałym – jak w wersjach typu „papier-ołówek” – poprzez zgrupowanie pozycji w skale albo w kolejności przypadkowej, ale stałej dla każdej z osób badanych (prze mieszan e dla różnych skal). Na podstawie wyników relacjonowanych badań można sądzić, że pełna losowość kolejności pozycji prowadzi do wyrównania dystansu fizycznego między pozycjami (bardziej zróżnicowanego, a jednocześnie stałego dla wersji o pozycjach pogrupowanych i eksponowanych w stałej kolejności), a w konsekwencji – do zmniejszenia zgodności odpowiedzi na pozycje o zbliżonej treści oraz zwiększenia zgodności odpowiedzi na pozycje treściowo odmienne. W odniesieniu do wersji o stałej kolejności można zatem oczekiwać większej zgodności odpowiedzi na pozycje o treści zbliżonej (pozycje pogrupowane) oraz zmniejszenia zgodności odpowiedzi na pozycje treściowo odmienne (przy założeniu, że zgodność odpowiedzi nie jest systematycznie niższa dla wersji o kolejności losowej). Odpowiadałoby to oczekiwaniu większej trafności zbieżnej pomiaru dla wersji pogrupowanej oraz lepszej – różnicowej

(niższe interkorelacje skal). Takie efekty nie wydają się wynikać z obciążenia odpowiedzi, a jedynie z lepszego różnicowania treści pozycji przez osoby badane. Rodzi to jednak następujący problem: jeśli procedura grupowania ułatwia osobie badanej identyfikację specyficznych treści i zwiększa spójność odpowiedzi, jednocześnie sprzyjając różnicowaniu treści odrębnych i zmniejszaniu spójności odpowiedzi na pozycje treściowo odmienne, to czy nie jest odpowiedzialna za możliwość wyróżniania czynnikowo jednorodnych wiązek pozycji. Taką możliwość sugeruje stan praktyki diagnozowania objawów porażkowego zaburzenia stresowego (*posttraumatic stress disorder* [PTSD]), jak też tocząca się od wielu lat dyskusja dotycząca struktury objawów zaburzenia i kryteriów jego rozpoznania, ujmowanych w systemach diagnostycznych DSM (*Diagnostic and Statistical Manual of Mental Disorders*).

Wspomniana dyskusja dotycząca empirycznej struktury objawów PTSD w relacji do kryteriów rozpoznania tego zaburzenia istniała już od pierwszych jego konceptualizacji i nabrała na sile po opublikowaniu DSM-IV, a także DSM-5 (American Psychiatric Association, 2000, 2013). Stosunkowo szybko udało się ustalić, że zarówno model trzech kategorii 17 objawów wg DSM-IV (intruzji, unikania/odrętwienia, nasilonego pobudzenia i reaktywności), jak też model czterech kategorii 20 objawów wg DSM-5 (intruzji, unikania, negatywnych zmian poznawczych i emocjonalnych oraz nasilonego pobudzenia i reaktywności; zob. tabela 1), nie odpowiadają empirycznej strukturze objawów (pomimo tego, iż większość narzędzi diagnostycznych wykorzystywała procedury grupowania pozycji wg kolejności ich opisu w DSM; Brewin, 2005).

Tabela 1

Objawy PTSD wg DSM-5 (American Psychiatric Association, 2013)

Kryteria (kategorie) i objawy PTSD według DSM-5

B: Obecność przynajmniej jednego objawu intruzji związanego z wydarzeniem traumatycznym, które pojawiło się po wydarzeniu traumatycznym:

B1. Powracające niechciane, intruzywne i stresujące wspomnienia traumatycznego zdarzenia.

B2. Powracające, dręczące sny, których treść i/lub nacechowanie emocjonalne jest związane z traumatycznym/i zdarzeniem/ami.

B3. Reakcje dysocjacyjne (np. przebliski wspomnień), w czasie których osoba czuje się lub zachowuje jakby traumatyczne zdarzenie/a znów miało/y miejsce. (Reakcje te można rozpatrywać na kontinuum z najbardziej ekstremalnym nasileniem polegającym na utracie orientacji w otaczającej rzeczywistości).

Ciąg dalszy tabeli na następnej stronie

Ciąg dalszy tabeli z poprzedniej strony

Kryteria (kategorie) i objawy PTSD według DSM-5

B4. Silny lub przedłużający się stres psychologiczny w przypadku ekspozycji na wewnętrzne lub zewnętrzne sygnały symbolizujące bądź przypominające jakiś aspekt traumatycznego zdarzenia.

B5. Silne reakcje fizjologiczne na wewnętrzne lub zewnętrzne bodźce symbolizujące lub przypominające jakiś aspekt traumatycznego zdarzenia.

C: Uporczywe unikanie bodźców związanych z traumą, które pojawiło się po wydarzeniu traumatycznym, przejawiające się jednym lub obydwoma poniższymi objawami:

C1. Unikanie lub usiłowanie uniknięcia przykrych wspomnień, myśli lub uczuć dotyczących albo ściśle związanych z traumatycznym/i wydarzeniem/ami.

C2. Unikanie lub usiłowanie uniknięcia zewnętrznych bodźców przypominających wydarzenie (ludzi, miejsc, rozmów, czynności, przedmiotów, sytuacji), które wzbudzają przykre wspomnienia, myśli lub uczucia o traumatycznym/ch wydarzeniu/ach lub są ściśle z nim/i związane.

D: Negatywne zmiany w sferze poznawczej i emocjonalnej, które pojawiły się lub nasiliły po wydarzeniu/ach traumatycznym/ch, przejawiające się co najmniej dwoma z poniższych objawów:

D1. Niemożność przypomnienia sobie ważnych aspektów traumatycznego/ych wydarzenia/ń (zazwyczaj wiążące się z amnezją dysocjacyjną, a nie z urazem głowy, spożyciem alkoholu lub użyciem narkotyków).

D2. Utrwalone i przesadne negatywne przekonania dotyczące siebie, innych ludzi lub świata (np. „jestem złym człowiekiem”, „nikomu nie można ufać”, „świat jest absolutnie niebezpieczny” czy „mój układ nerwowy jest kompletnie zrujnowany”).

D3. Utrwalony i zniekształcony sposób myślenia o przyczynach lub konsekwencjach traumatycznego/ych wydarzenia/ń, który prowadzi do obwiniania siebie lub innych osób.

D4. Utrwalony, negatywny stan emocjonalny (np. strach, przerażenie, złość, poczucie winy lub wstydu).

D5. Wyraźne ograniczone zainteresowanie lub udział w ważnych aktywnościach.

D6. Poczucie oddalenia od innych ludzi lub wyobcowania.

D7. Uporczywa niezdolność do przeżywania pozytywnych uczuć (np. niezdolność do odczuwania szczęścia, zadowolenia lub miłości).

E: Wyraźne zmiany we wzбудzeniu i reaktywności związane z wydarzeniem/ami traumatycznym/i, które zaczęły się lub nasiliły po tym wydarzeniu, w postaci co najmniej dwóch z poniżej wymienionych:

E1. Drażliwość lub wybuchy złości.

E2. Zachowania ryzykanckie lub autodestrukcyjne.

E3. Wzmożona czujność.

E4. Nadmiernie silna reakcja na niespodziewane bodźce.

E5. Trudności z koncentracją.

E6. Trudności z zasypianiem lub snem.

Adnotacja. Kursywą zaznaczono trzy objawy, które zostały dodane (oraz jedno – przeformułowane) do DSM-5 w stosunku do DSM-IV. Tłumaczenie kryteriów DSM-5 za: Zawadzki i Popiel, 2014.

W istocie konfirmacyjne analizy czynnikowe wskazały na zasadność klasyfikowania objawów PTSD wg DSM-IV w pięć czynników (Elhai i in., 2011): objawy intruzji, unikania, odrętwienia, pobudzenia lękowego i dysforycznego oraz siedem czynników (Armour i in., 2015; zob. tabela 2): intruzji, unikania, negatywnego afektu, anhedonii, zachowań eksternalizacyjnych, pobudzenia lękowego i dysforycznego.

Tabela 2

Modele struktury objawów PTSD ujmowanych w DSM-5

Objawy PTSD wg DSM-5	Model DSM-5 (4-czynnikowy)	Model dysforyczny (4-czynnikowy)	Model dysforyczno-pobudzeniowy (5-czynnikowy)	Model negatywnych zmian poznawczych i odrętwienia (6-czynnikowy)	Model anhedoniczny (6-czynnikowy)	Model hybrydowy (7-czynnikowy)
B1	I	I	I	I	I	I
B2	I	I	I	I	I	I
B3	I	I	I	I	I	I
B4	I	I	I	I	I	I
B5	I	I	I	I	I	I
C1	UN	UN	UN	UN	UN	UN
C2	UN	UN	UN	UN	UN	UN
D1	NZPE	D	OD	OD	NA	NA
D2	NZPE	D	OD	ZP	NA	NA
D3	NZPE	D	OD	ZP	NA	NA
D4	NZPE	D	OD	ZP	NA	NA
D5	NZPE	D	OD	OD	AN	AN
D6	NZPE	D	OD	OD	AN	AN
D7	NZPE	D	OD	OD	AN	AN
E1	P	D	PD	PD	PD	ZE
E2	P	D	PD	PD	PD	ZE
E3	P	PL	PL	PL	PL	PL
E4	P	PL	PL	PL	PL	PL
E5	P	D	PD	PD	PD	PD
E6	P	D	PD	PD	PD	PD

Adnotacja. Specyfikacja modelu jednoczynnikowego została pominięta. I – intruzja, UN – unikanie, NZPE – negatywne zmiany poznawcze i emocjonalne, P – zmiany pobudzenia i reaktywności, D – dysforia, PL – pobudzenie lękowe, OD – odrętwienie, ZP – zmiany poznawcze, NA – negatywny afekt, AN – anhedonia, PD – pobudzenie dysforyczne, ZE – zachowania eksternalizacyjne.

Istotną kwestią był fakt, iż czynniki te klasyfikowały pozycje dokładnie wg kolejności odpowiadającej kryteriom opisywanym w DSM, a więc ułożone w bliskim dystansie fizycznym (wyodrębniając podkategorie w ramach poszczególnych ogólniejszych wiązek objawów). Już w odniesieniu do DSM-IV podniesiono kwestię, czy wynik taki nie jest artefaktem wynikającym ze stałej kolejności prezentacji pozycji (Marshall i in., 2013), a wątpliwości objęły następnie także objawy ujęte w DSM-5 (Witte i in., 2015). Obecnie uważa się, że kolejność pozycji nie ma istotnego znaczenia (decyduje faktycznie specyfika treściowa, skoro także kryteria w DSM są w istocie pogrupowane). Jednakże problem pozostaje aktualny ze względu na fakt, iż zdecydowana większość narzędzi badawczych nadal wykorzystuje procedurę stałej kolejności pozycji wg kryteriów DSM, zaś narzędzia o przemieszanej kolejności pozycji są znacznie rzadziej stosowane ze względu na ich mniejszą ekonomiczność i słabszą charakterystykę psychometryczną (Brewin, 2005). Analizie tego problemu poświęcona jest poniższa praca, która dotyczy empirycznej struktury objawów PTSD wg DSM-5.

PROBLEM BADAŃ, PYTANIA I HIPOTEZY BADAWCZE

Problem podjęty w tej pracy dotyczył struktury objawów PTSD oraz równoważności pomiarowej dwóch wersji *PTSD Checklist dla DSM-5* (PCL-5). Oczekiwano potwierdzenia modelu 7-czynnikowego niezależnie od wersji kolejności pozycji (1): stałej według porządku wyodrębniania objawów w DSM-5 oraz losowej. Jednocześnie, sformułowano pytanie (2) o równoważność pomiarową obu wersji (przyjmując, że zidentyfikowana struktura będzie zbieżna, czyli będzie wykazywać tzw. równoważność konfiguralną) – pytanie zatem dotyczyło równoważności metrycznej, skalarnej itd. (zob. Cyniak-Cieciura i in., w tym tomie). Sformułowano także oczekiwanie (3), że ewentualne rozbieżności obu wersji mogą być efektem spójności odpowiadania – większej w obrębie pozycji tworzących dane czynniki i mniejszej pomiędzy pozycjami czynników odrębnych (co odpowiadałoby możliwej wyższej trafności zbieżnej i różnicowej dla wersji stałej niż losowej, dla której oczekiwano mniejszej spójności odpowiadania na pozycje tworzące dane czynniki i większej na pozycje czynników odrębnych). Zadano też pytanie badawcze (4) o związki objawów PTSD badanych oboma wersjami z aprobatą społeczną, traktowaną jako czynnik obciążający odpowiedzi kwestionariuszowe oraz z neurotycznością

(zasadniczym osobowościowym korelatem objawów zaburzenia; Kotov i in., 2010) – jako zmiennej mogącej ujawnić obniżoną trafność zbieżną (efektu zasadniczo możliwego do oczekiwania dla wersji losowej).

METODA

Zastosowane narzędzia badawcze

Inwentarz PCL-5

We wszystkich relacjonowanych w tej pracy badaniach zastosowano PCL-5, inwentarz przeznaczony do diagnozy objawów PTSD (wersja w tłumaczeniu Popiel i in., 2021, pozyskana ze strony: <https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp>). Jest on wersją PCL (*PTSD Checklist*; Weathers i Ford, 1996; Weathers i in., 1991), dostosowaną do DSM-5 (zob. tabela 1), zawierającą 20 pozycji badających objawy PTSD doświadczane przez okres ostatniego miesiąca (Blevins i in., 2015; Weathers, Litz i in., 2013). Są one oceniane na 5-stopniowej skali, gdzie 0 oznacza, że objaw *wcale* nie był dokuczliwy, 1 – *nieznacznie*, 2 – *umiarkowanie*, 3 – *znacznie*, zaś 4 – *ogromnie* dokuczliwy, przy czym w standardowej wersji objawy są zamieszczone w kolejności stałej, odpowiadającej kryteriom DSM-5. Rozpoznanie PTSD jest możliwe na podstawie klasycznej procedury konfiguracyjnej: dychotomizacja odpowiedzi (0–1 punktów = brak objawu; 2–4 punkty = objaw) oraz stwierdzenie przynajmniej po jednym objawie kategorii B (intruzja) i kategorii C (unikanie) oraz przynajmniej po dwa objawy kategorii D (negatywne zmiany poznawcze i emocjonalne) i E (zmiany pobudzenia i reaktywności). Minimalnie zatem do diagnozy PTSD niezbędne jest stwierdzenie sześciu objawów, a w dodatku rozpoznanie wszystkich czterech kategorii. Alternatywna procedura zakłada obliczenie wyników sumarycznych 20 pozycji i zdychotomizowanie ich wokół 31–33 punktów (<https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp>). Skala doczekała się szerokiej walidacji i jest traktowana jako jedno z podstawowych samoopisowych narzędzi do badania PTSD (Bovin i in., 2016; Ogińska-Bulik i in., 2018; Popiel i in., 2021; Witte i in., 2015). Warto przy tym zwrócić uwagę, że pomimo postulowanej w DSM-5 struktury objawów PTSD, modele wieloczynnikowe, w tym model 7-czynnikowy, opisują ją lepiej niż model 4-czynnikowy, także w przypadku PCL-5 (zob. Bovin i in., 2016).

Objawy są relacjonowane w odniesieniu do wskazanego przez osobę badaną wydarzenia, przy czym lista traumatycznych zdarzeń jest przedstawiona w liście

LEC-5 (*Life Events Checklist for DSM-5*; Popiel i in., 2021; Rzeszutek i in., 2018; Weathers, Blake i in., 2013). Lista ta obejmuje kilkanaście zdarzeń, przy czym osoba wskazuje sposób konfrontacji (zgodnie z kryterium A w DSM-5; American Psychiatric Association, 2013), tzn. czy doświadczyła tego zdarzenia sama, była świadkiem, dowiedziała się albo została skonfrontowana w trakcie pracy/służby zawodowej. Umożliwia to kontrolę całkowitej traumatyzacji, a jednocześnie ułatwia osobie badanej wskazanie zdarzenia najbardziej traumatycznego. Na potrzeby tej pracy wskazane zdarzenia były dodatkowo pogrupowane w szersze kategorie, niż jest to ujęte w LEC-5 (zob. tabela 3).

Tabela 3

Wskazane kategorie traumatycznych wydarzeń w całej badanej próbie

Kategoria zdarzenia	Pomiar pierwszy – liczebność	Pomiar pierwszy – procent	Pomiar drugi – liczebność	Pomiar drugi – procent
Choroba/zgon/ samobójstwo/uraz	371	50,1	291	39,3
Przemoc	57	7,7	59	8,0
Wypadki komunikacyjne	112	15,1	113	15,3
Katastrofy przemysłowe/pożar	34	4,6	37	5,0
Gwałt/molestowanie	34	4,6	33	4,5
Przemoc domowa/ przemoc – partner	24	3,2	19	2,6
Katastrofy naturalne	34	4,6	45	6,1
Wiele zdarzeń/wojenne	45	6,1	31	4,2
Odmowa wskazania konkretnego zdarzenia/ brak opisu	29	3,9	112	15,1
Grupa stała-losowa	Pomiar pierwszy – liczebność	Pomiar pierwszy – procent	Pomiar drugi – liczebność	Pomiar drugi – procent
Choroba/zgon/ samobójstwo/uraz	181	50,0	135	37,3
Przemoc	31	8,6	30	8,3

Ciąg dalszy tabeli na następnej stronie

Ciąg dalszy tabeli z poprzedniej strony

Kategoria zdarzenia	Pomiar pierwszy – liczebność	Pomiar pierwszy – procent	Pomiar drugi – liczebność	Pomiar drugi – procent
Wypadki komunikacyjne	53	14,6	66	18,2
Katastrofy przemysłowe/pożar	16	4,4	17	4,7
Gwałt/molestowanie	16	4,4	15	4,1
Przemoc domowa/przemoc – partner	7	1,9	9	2,5
Katastrofy naturalne	16	4,4	18	5,0
Wiele zdarzeń/wojenne	24	6,6	19	5,2
Odmowa wskazania konkretnego zdarzenia/ brak opisu	18	5,0	53	14,6
Grupa losowa-stała	Pomiar pierwszy – liczebność	Pomiar pierwszy – procent	Pomiar drugi – liczebność	Pomiar drugi – procent
Choroba/zgon/ samobójstwo/uraz	190	50,3	156	41,3
Przemoc	26	6,9	29	7,7
Wypadki komunikacyjne	59	15,6	47	12,4
Katastrofy przemysłowe/pożar	18	4,8	20	5,3
Gwałt/molestowanie	18	4,8	18	4,8
Przemoc domowa/ przemoc – partner	17	4,5	10	2,6
Katastrofy naturalne	18	4,8	27	7,1
Wiele zdarzeń/wojenne	21	5,6	12	3,2
Odmowa wskazania konkretnego zdarzenia/ brak opisu	11	2,9	59	15,6

Adnotacja. Nasilenie objawów: kategoria zdarzenia: $F(8, 722) = 3,377$; $p < 0,001$; wersja PCL-5: $F(1, 722) = 0,641$; $p = 0,424$; interakcja obu czynników: $F(8, 722) = 0,557$; $p = 0,813$. Nasilenie objawów i diagnoza PTSD – najwyższe wartości/odsetek dla przemocy domowej, gwałtu/molestowania, wielu zdarzeń/traum wojennych i odmowy wskazania.

Zastosowane zostały dwie wersje PCL-5 – z pozycjami w kolejności stałej (jak w wersji drukowanej), odpowiadającej kolejności wyodrębniania w kryteriach DSM-5 oraz kolejności całkowicie losowej (zmienna kolejność dla każdego zastosowania inwentarza). Wersje stałe, ale o przypadkowej kolejności (nieodpowiadającej kryteriom DSM-5) nie były brane pod uwagę. W efekcie poprzez wersję stałą należy rozumieć ekspozowanie w badaniach pozycji w kolejności jednakowej dla wszystkich osób badanych i w dodatku odpowiadającej kryteriom DSM-5, zaś w poprzez wersję losową – pełną losowość kolejności pozycji, w efekcie czego każda osoba była badana wersją o odmiennej kolejności. Pozycje w LEC-5 nie były randomizowane.

Dodatkowe narzędzia badawcze

W badaniu właściwym zastosowano obok skali PCL-5 także skalę neurotyczności, wybraną z inwentarza polskiej adaptacji IPIP-NEO-FFI-50 Goldberga (*International Personality Item Pool NEO-Five Factor Inventory-50*; <https://www.ipip.uksw.edu.pl/test.skale.php?id=35#0>), opracowanej przez Strusa i in. (2014). Zdecydowano się na pomiar tej zmiennej, gdyż jest ona głównym osobowościowym korelatem objawów PTSD (Kotov i in., 2010). Do kontroli możliwych zniekształceń samoopisu użyto tłumaczenia oryginalnej wersji *Skali aprobaty społecznej Marlowe’a-Crowne’a* (Crowne i Marlowe, 1960), dokonanej przez Barcisia (2021). Zawiera ona 33 pozycje z dwukategorialnymi odpowiedziami *fałsz-prawda*. Pozycje w tych narzędziach miały stałą kolejność, a zastosowano je w porządku: LEC-5, PCL-5, skalę neurotyczności i aprobaty społecznej.

Analizy wstępne

We wstępnej analizie wykorzystano dane zebrane we wcześniejszych badaniach w grupie 5008 osób (Borek, 2020; Cyniak-Cieciura i in., w tym tomie [wyłącznie dane polskie]; Feldman, 2019; Popiel i in., 2021; Witowska, 2020; Wojciechowska, 2021). Wszystkie zostały zrealizowane w formie badań online na platformie Qualtrics, z zastosowaniem wersji stałej pozycji PCL-5 (pozycje w kolejności odpowiadającej objawom PTSD wg DSM-5; American Psychiatric Association, 2013) albo wersji zrandomizowanej (kolejność pozycji generowana losowo dla każdej osoby badanej). Można zatem przyjąć, że w obu wersjach dystanse fizyczne dzielące poszczególne pozycje PCL-5 były w średniej zbliżone (choć bardzo zróżnicowane dla wersji stałej, zaś wyrównane średnio dla każdej pozycji w wersji losowej). Grupy wypełniające obie wersje PCL-5 wykazywały

jednak istotne różnice ze względu na zmienne demograficzne, jak płeć czy wiek oraz zmienne odnoszące się do traumatyzacji, jak liczba doświadczonych traumatycznych doświadczeń czy wydarzenia wskazane przez osoby badane jako najbardziej urazowe (badane za pomocą LEC-5; Rzesutek i in., 2018; Weathers, Blake i in., 2013). Zmienne te wykazują powiązanie z objawami PTSD, toteż odmienność obu grup mogła rzutować na ocenę równoważności pomiarowej PCL-5. Po usunięciu z całej próby danych tych osób, które nie wskazały doznania żadnego traumatycznego zdarzenia, zastosowano dobór kwotowy do obu grup ze względu na płeć, wiek, liczbę i rodzaj traumatycznych wydarzeń. Niestety dobór ten nie doprowadził do wyrównania charakterystyki obu grup ze względu na nasilenie objawów PTSD – nadal rejestrowane były różnice dla całych grup, jak w obrębie poszczególnych kategorii traumatycznych zdarzeń w kierunku wyższego nasilenia objawów w próbie o stałej kolejności pozycji PCL-5 (co mogło oznaczać, że dla obu wersji „doświadczenie przemoc domowej” nie musiało oznaczać tej samej ciężkości tego doświadczenia). W efekcie równoważność skalarna z definicji nie byłaby możliwa do wykazania (Lubiewska i Głogowska, 2018). W ostatecznym kroku wyrównano zatem grupy poprzez proporcjonalny dobór osób dodatkowo na podstawie diagnozy PTSD. Decyzja ta jest mocno problematyczna, bowiem faktycznie oznacza ona ingerencję w analizowaną zmienną zależną, toteż dane te potraktowano jedynie jako wstępne i przybliżone. Do ostatecznej analizy zakwalifikowano zatem dane zebrane w próbie 3126 osób (wersja stała: $n_s = 1908$; 61,0%; wersja losowa: $n_l = 1218$; 39,0%), 2402 kobiety (76,8%) i 724 mężczyzn (23,2%) w wieku 18–83 lata ($M = 32,27$; $SD = 11,27$). Obie grupy nie różniły się płcią: $\chi^2(1, N = 3126) = 0,26$; $p = 0,61$, wiekiem: $F(1, 3124) = 0,23$; $p = 0,63$, liczbą traumatycznych wydarzeń: $F(1, 3124) = 1,47$; $p = 0,19$, wskazanym wydarzeniem urazowym: $\chi^2(6, N = 3126) = 3,63$; $p = 0,73$ ani diagnozą: $\chi^2(1, N = 3126) = 0,01$; $p = 0,97$ czy nasileniem objawów PTSD: $F(1; 2732,47) = 2,73$; $p = 0,10$; przy wyższej wariancji w grupie o stałej kolejności: $F(1, 3124) = 11,28$; $p < 0,001$. Do analizy zastosowano model odporny największej wiarygodności (MLR; *Maximum Likelihood Robust*): 5-stopniowa skala pomiarowa (Rhemtulla i in., 2012), ale rozkłady odpowiedzi o dodatniej asymetryczności (Konarski, 2009) w środowisku R (R Core Team, 2021) z zastosowaniem pakietu *laavan* (Rosseel, 2015). Porównań dokonano dla siedmiu modeli struktury objawów PTSD (Cyniak-Cieciura i in., 2017; zob. tabela 2), w tym czynnika ogólnego, 4-czynnikowej struktury postulowanej przez DSM-5 (American Psychiatric Association, 2013), poszerzonego modelu dysforycznego (Simms i in., 2002),

poszerzonego modelu 5-czynnikowego dysforyczno-pobudzeniowego (Elhai i in., 2011)¹, a także modeli 6-czynnikowych: negatywnych zmian poznawczych i odrętwienia (Zawadzki i in., 2015) oraz modelu anhedonicznego (Armour i in., 2015), a także 7-czynnikowego modelu hybrydowego (Armour i in., 2015; opis modeli: zob. Cyniak-Cieciura i in., 2017). Dopasowanie modeli oceniane było na podstawie wartości robustowej statystyki χ^2 oraz wskaźników *robust* RMSEA, *robust* CFI, *robust* TLI, SRMR, AIC, BIC i SABIC. Zgodnie z kryteriami zaproponowanymi przez Hu i Bentler (1999) wartości CFI $\geq 0,90$; TLI $\geq 0,90$; RMSEA i SRMR $\leq 0,08$ uznane zostały za wskazujące na dobre dopasowanie modelu do danych. Wyboru modelu o najlepszym dopasowaniu dokonano na podstawie wartości uwzględniających dobroć dopasowania oraz złożoność modelu, tj. AIC, SABIC i najbardziej restrykcyjny BIC. W drugim kroku zweryfikowano równoważność pomiarową struktury objawów PTSD, która wynikała z modelu o najlepszym dopasowaniu (wyodrębnionym w kroku pierwszym) w poszczególnych próbach. W tym celu przeprowadzono analizę modeli zagnieżdżonych, zakładając równoważność struktury (tzw. równoważność konstruktów; *configural invariance*), a następnie zakładając dodatkowo równość ładunków czynnikowych (tzw. równoważność metryczna; *metric invariance*) oraz stałych regresji (tzw. równoważność skalarna; *scalar invariance*) i błędów pomiaru (*residual invariance*), a dodatkowo równość średnich, wariancji i kowariancji zmiennych latentnych (te trzy ostatnie modele nie muszą być testowane w ścisłej kolejności; Milfont i Fisher, 2010; w prezentowanych analizach przyjęto jak w porządku wymienienia). Sugeruje się (Lubiewska i Głogowska, 2018; Zercher i in., 2015), że przy testowaniu równoważności metrycznej w dużych próbach $\Delta CFI \leq -0,010$ oraz $\Delta RMSEA \leq 0,015$ lub $\Delta SRMR \leq 0,030$ pomiędzy zagnieżdżonymi modelami wskazują na równoważność pomiarową. Z kolei przy testowaniu stałych regresji krytyczne różnice wartości wskaźników wynoszą: $\Delta CFI \leq -0,010$ oraz $\Delta RMSEA \leq 0,015$ lub $\Delta SRMR \leq 0,010$. Niekiedy brana jest też pod uwagę zmiana parametrów Akaike (ΔAIC) oraz BIC (ΔBIC). Wartości kryterialne dla modeli zakładających jeszcze silniejszą równoważność niż skalarna nie są jednoznacznie określone. Oprócz tych wskaźników testowano też zmiany wartości robustowej statystyki χ^2 .

Analizy zastosowane w planie dwugrupowym jednoznacznie wykazały, że najlepsze dopasowanie wykazał 7-czynnikowy model hybrydowy – jednoznacznie

¹ Modele opracowane dla DSM-IV (American Psychiatric Association, 2000) – rozszerzone na objawy PTSD wg DSM-5 (zob. Zawadzki i in., 2015).

najniższe wskaźniki AIC, BIC i SABIC oraz akceptowalne wskaźniki dobroci dopasowania: *robust* CFI = 0,968; *robust* TLI = 0,969; *robust* RMSEA = 0,053 (90% CI [0,050; 0,056]); SRMR = 0,030; przy istotnej wartości statystyki χ^2 dla wszystkich modeli. Analiza równoważności wykazała jednak, że nie zostały spełnione aspekty równoważności, poczynając od równości ładunków czynnikowych (poprzez stałe regresji, błędy pomiaru, równość średnich, wariancji i kowariancji zmiennych latentnych). Zmiany wskaźników dobroci dopasowania były co prawda nieznaczne (znacznie poniżej sugerowanych krytycznych wartości przyrostowych), co można byłoby uznać za spełniające kryteria dopasowania, ale za rozstrzygające uznano istotne przyrosty wartości *robust* statystyki Satorra-Bentler χ^2 . W efekcie stwierdzono, że obie wersje PCL-5: stała i losowa, choć mierzą zbliżone konstrukty (i potwierdzają hipotezę o 7-czynnikowej strukturze objawów PTSD), to jednak pomiarowo nie są równoważne. Zmienność ta jednak mogła wynikać z odmienności obu wersji PCL-5, jak też odmienności charakterystyki obu prób, nie mówiąc już o obciążeniach analiz wynikających z doboru osób badanych do obu grup na podstawie testowanej zmiennej, tj. diagnozy PTSD. Z tego względu zaprojektowane zostało badanie obejmujące obie wersje PCL-5 w tej samej grupie osób badanych.

Badania właściwe

Badania zrealizowano przez internetowy panel badawczy Ariadna w planie dwóch pomiarów skrzyżowanych: wersja stała-losowa oraz wersja losowa-stała z drugim pomiarem zrealizowanym po upływie dwóch tygodni. Ze względu na kryterium F diagnozy PTSD (American Psychiatric Association, 2013) przyjęto, że upływ czasu między oboma badaniami powinien być jak najkrótszy i nie przekraczać jednego miesiąca. Ze względów organizacyjnych okres dwóch tygodni był najkrótszym możliwym czasem realizacji badań test-retest. W pomiarze pierwszym udział wzięło 1214 osób, w drugim – 742 osoby. Osoby badane były użytkownikami panelu badawczego, których identyfikacja odbywała się na etapie rejestracji w panelu i odbiorze nagród. Każdy respondent został wynagrodzony za udział w badaniu punktami do wymiany wybranej nagrody rzeczowej z katalogu nagród panelu badawczego. Pomiarzy były anonimowe i dobrowolne. Użytkownik miał możliwość odstąpić od badania w każdym momencie i był poinformowany o jego ogólnej tematyce. W bazie danych

użytkownicy są rozróżnieni anonimowym numerem ID niepozwalającym na zewnętrzną identyfikację ich tożsamości. Proces rejestracji i wynagrodzeń panelu opiera się na nagrodach rzeczowych wymagających dostawy za pomocą usług kurierskich, co jest dodatkowym etapem weryfikacji prawdziwości osoby fizycznej stojącej za zarejestrowanym kontem na panelu. System ankietowy automatycznie wykluczał osoby, które wypełniały ankietę w nierealistycznym tempie, pozbywając się tym samym potencjalnych botów i osób wypełniających nieuważnie. Między pomiarami została wykonana kontrola jakości zebranych danych (zob. Webb i Tangney, 2022). Osoby odpowiadające na pytania otwarte losowymi ciągami znaków, zamiast prawdziwymi słowami, zostały usunięte z bazy i wykluczone z kolejnego pomiaru. Dodatkowo w pomiarze pierwszym zastosowane zostało pytanie kontrolujące uwagę respondenta, a osoby, które odpowiedziały błędnie nie zostały zaproszone do wzięcia udziału w drugim pomiarze. W efekcie tych działań do drugiego pomiaru zaproszone zostały wyłącznie osoby wykazujące się rzetelnym odpowiadaniem. Ponadto ze zbioru danych z pierwszego pomiaru usunięto dane tych osób, które nie wskazały doświadczenia przynajmniej jednego wydarzenia traumatycznego wymienionego w LEC-5 (choć uwzględniono osoby, które odmówiły opisania najbardziej urazowego doświadczenia). Proces kontroli jakości danych został powtórzony po drugim pomiarze, skutkując dodatkowo usunięciem danych dwóch osób, które wykazały się odpowiadaniem nieuważnym lub nieszczerym. Badanie zostało zaakceptowane przez Komisję Etyczną Wydziału Psychologii Uniwersytetu Warszawskiego.

Ostatecznie w analizach uwzględnione zostały wyniki zebrane w grupie 740 osób badanych ($n_s = 362$; 48,9% – wersja stała-losowa; $n_l = 378$; 51,1% – wersja losowa-stała, czyli w sumie po 740 pomiarów dla każdej wersji), 447 kobiet (60,4%) i 293 mężczyzn (39,6%) w wieku 18–79 lat ($M = 45,28$; $SD = 14,61$). Obie badane podgrupy (klasyfikacja na podstawie konfiguracji pomiarów) nie różniły się:

- płcią – $\chi^2(1, N = 740) = 0,30$; $p = 0,58$ (grupa stała-losowa: 215 kobiet; 59,3%; grupa losowa-stała: 232 kobiety; 61,4%);
- wiekiem – $F(1, 738) = 0,96$; $p = 0,33$ (grupa stała-losowa: 18–79 lat; $M = 44,74$; $SD = 14,83$; grupa losowa-stała: 18–79 lat; $M = 45,79$; $SD = 14,39$);
- wykształceniem – $\chi^2(5, N = 740) = 4,24$; $p = 0,52$ (po zrekodowaniu wykształcenie wyższe – grupa stała-losowa: 193 osoby; 53,3%; grupa losowa-stała: 189 osób; 50,0%);

- miejscem zamieszkania – $\chi^2(4, N = 740) = 3,23; p = 0,52$ (po zrekodowaniu = duże miasto, tj. pow. 100 tys. mieszkańców; grupa stała-losowa: 168 osób; 46,4%; grupa losowa-stała: 155 osób; 41,0%);
- liczbą traumatycznych wydarzeń – $F(1, 738) = 0,27; p = 0,61$ (grupa stała-losowa: $M = 5,50; SD = 3,25$; zakres: 1–13; grupa losowa-stała: $M = 5,62; SD = 3,29$; zakres: 1–13; dla pomiaru drugiego: $F(1, 738) = 0,35; p = 0,56$ (grupa stała-losowa: $M = 5,47; SD = 3,49$; zakres: 1–13; grupa losowa-stała: $M = 5,62; SD = 3,55$; zakres: 1–13);
- wskazanym wydarzeniem urazowym – $\chi^2(8, N = 740) = 7,05; p = 0,53$; pomiar drugi: $\chi^2(8, N = 740) = 8,66; p = 0,37$ (zob. tabela 3);
- diagnozą – $\chi^2(1, N = 740) = 0,28; p = 0,60$ (grupa stała-losowa: 152 osoby; 42,0%; grupa losowa-stała: 166 osób; 43,8%; pomiar drugi – grupa stała-losowa: 164 osoby; 45,3%; grupa losowa-stała: 143 osoby; 37,8%);
- nasileniem objawów PTSD – $F(1, 738) = 0,13; p = 0,72$ (grupa stała-losowa: $M = 27,90; SD = 18,06$; grupa losowa-stała: $M = 27,43; SD = 17,24$); jak też dla pomiaru drugiego: $F(1, 738) = 2,55; p = 0,11$ (grupa stała-losowa: $M = 26,96; SD = 19,21$; grupa losowa-stała: $M = 24,75; SD = 18,37$).

Dla zastosowanej tylko w pomiarze pierwszym skali neurotyczności nie stwierdzono także istotnych różnic międzygrupowych: $F(1, 738) = 0,67; p = 0,42$ (grupa stała-losowa: $M = 28,12; SD = 7,76$; grupa losowa-stała: $M = 27,65; SD = 7,76$; 10 pozycji, w McDonalda = 0,900), ale uzyskano je dla aprobaty społecznej: $F(1, 738) = 3,97; p = 0,047; \eta^2 = 0,005$ (grupa stała-losowa: $M = 17,69; SD = 5,84$; grupa losowa-stała: $M = 18,55; SD = 5,86$; 33 pozycje, w McDonalda = 0,818²).

Dla obu momentów badania nie stwierdzono różnic w diagnozie PTSD: test McNemara $\chi^2(1, N = 740) = 0,58; p = 0,45$; zgodność diagnoz: $\chi^2(1, N = 740) = 201,03; p < 0,001$; współczynnik kappa $\kappa = 0,52$ (zgodnych diagnoz: 567/740; 76,6%; pomiar pierwszy: 318 diagnoz; 43,0%; pomiar drugi: 307 diagnoz; 41,5%) czy liczbie traumatycznych wydarzeń: $t(739) = 0,12; p = 0,91$; korelacja wewnątrzklasowa ICC = 0,56 ($M = 5,56; SD = 3,27$ oraz $M = 5,55; SD = 3,52$), a tylko w nasileniu objawów PTSD: $t(739) = 3,56; p < 0,001$ – niższe nasilenie objawów w badaniu drugim; ICC = 0,70 ($M = 27,66; SD = 17,63$ oraz $M = 25,83; SD = 18,81$); ANOVA z powtórzonym pomiarem: $F(1, 379) = 12,63$;

2 Rozkłady nasilenia objawów PTSD oraz liczby traumatycznych wydarzeń dla wszystkich 1480 pomiarów, jak też nasilenia neurotyczności i aprobaty społecznej w pomiarze pierwszym (740 osób) wykazywały wskaźniki kurtozy i skośności nieodbiegające znacząco od charakterystyki rozkładu normalnego (w przedziale od -1 do +1 SD).

$p < 0,01$; $\eta^2 = 0,014$)³. Zastosowano ten sam model CFA (*Confirmatory Factor Analysis*) i te same modele struktury objawów PTSD oraz wskaźniki dopasowania modeli teoretycznych do empirycznej struktury objawów PTSD, jak w analizach wstępnych opisanych powyżej. W pierwszym kroku testowano równoważność obu pomiarów PCL-5, co było jednym z warunków metodologicznych możliwości połączenia danych uzyskanych dla danej wersji inwentarza z obu momentów badania. Analizy wyraźnie wskazały na najlepsze dopasowanie modelu 7-czynnikowego hybrydowego (model konfiguralny) dla obu momentów badania (zob. tabela 4).

³ Diagnoza PTSD wg DSM-5 wymaga zrekodowania odpowiedzi na skalę dychotomiczną, rejestrującą występowanie (lub brak) objawu i dalej ma charakter konfiguracyjny wymagający spełnienia danego kryterium objawów oraz wszystkich czterech kryteriów. Z tego względu nie w pełni odpowiada intensywności objawów.

Tabela 4*Wskaźniki dobrego dopasowania modeli struktury objawów PTSD dla obu pomiarów PCL-5*

Model	robust χ^2 (df)	p	AIC	BIC	SABIC	robust RMSEA [90% CI]	robust CFI	robust TLI	SRMR
1-czynnikowy ogólny	1448,923 (340)	< 0,001	71903,840	72539,815	72158,611	0,081 [0,081; 0,085]	0,929	0,921	0,035
4-czynnikowy model DSM-5	907,720 (328)	< 0,001	71113,965	71813,539	71394,214	0,059 [0,054; 0,064]	0,964	0,928	0,028
4-czynnikowy model dysforyczny	993,634 (328)	< 0,001	71246,692	71946,265	71526,941	0,063 [0,059; 0,068]	0,958	0,951	0,028
5-czynnikowy model dysforyczno-pobudzeniowy	892,690 (320)	< 0,001	71106,940	71848,911	71404,173	0,059 [0,055; 0,064]	0,964	0,957	0,028
6-czynnikowy model poznawczo-odrętwieniowy	780,901 (310)	< 0,001	70959,462	71754,432	71277,926	0,055 [0,050; 0,059]	0,971	0,964	0,026
6-czynnikowy model anhedoniczny	788,258 (310)	< 0,001	70969,751	71764,721	71288,215	0,055 [0,050; 0,060]	0,970	0,963	0,027
7-czynnikowy model hybrydowy	723,409 (298)	< 0,001	70896,550	71755,118	71240,492	0,053 [0,048; 0,058]	0,973	0,966	0,026

Adnotacja. AIC – kryterium informacyjne Akaikego, BIC – kryterium informacyjne Bayesa, SABIC – kryterium informacyjne Bayesa z poprawką na wielkość próby, robust CFI – odporny porównawczy indeks dopasowania, robust RMSEA – odporny pierwiastek średniokwadratowego błędu aproksymacji, 90% CI – przedział ufności, robust TLI – wskaźnik Tuckera-Lewisa, SRMR – średni standaryzowany kwadrat reszt. Czcionką pogrubioną oznaczono najniższe wskaźniki AIC, BIC i SABIC.

Model 7-czynnikowy wykazywał najniższe wskaźniki AIC i SABIC, a BIC był nieznacznie wyższy w stosunku do modelu 6-czynnikowego⁴. Dla całości danych (czyli 1480 pomiarów) model ten wykazywał też jednoznacznie najlepsze dopasowanie do danych, podobnie jak dla obu podgrup w pomiarze pierwszym. Analiza równoważności wykazała, że wszystkie wartości wskaźników przyrostowych są bardzo niewielkie (znacznie poniżej sugerowanych krytycznych wartości przyrostowych) i wskazują na pełną równoważność pomiarową PCL-5 dla obu momentów badania (zob. tabela 5).

⁴ Wskaźnik BIC jest mocno restrykcyjny i zakłada najsilniejszą korektę ze względu na złożoność modelu, choć jest też bardzo użyteczny ze względu na wskazanie krytycznych wartości przyrostowych (Kass i Raftery, 1995; Raftery, 1995).

Tabela 5

Równoważność pomiarowa modelu 7-czynnikowego obu pomiarów PCL-5

Model	robust χ^2 (df)	p	$\Delta \chi^2$	$p\Delta \chi^2$ diff	robust CFI	ΔCFI	robust RMSEA [90% CI]	$\Delta RMSEA$	SRMR	$\Delta SRMR$	AIC	BIC	SABIC
Równoważność struktury	723,409 (298)	< 0,001	–	–	0,973	–	0,053 [0,048; 0,058]	–	0,026	–	70896,550	71755,118	71240,492
Równoważność ładunków	741,172 (311)	< 0,001	10,154 (13)	0,681	0,974	0,001	0,052 [0,047; 0,056]	–0,001	0,028	0,002	70879,852	71669,522	71196,193
Równoważność stałych regresji (intercepts)	763,740 (324)	< 0,001	19,155 (13)	0,118	0,973	–0,001	0,051 [0,046; 0,055]	–0,001	0,028	0,000	70873,083	71593,856	71161,824
Równoważność błędów pomiaru (residuals)	804,849 (344)	< 0,001	41,654 (20)	0,003	0,972	–0,001	0,051 [0,046; 0,055]	0,000	0,029	0,001	70897,952	71512,728	71144,231
Równoważność średnich zmiennych latentnych	820,835 (351)	< 0,001	15,833 (7)	0,027	0,971	–0,001	0,050 [0,046; 0,055]	–0,001	0,031	0,002	70899,991	71477,669	71131,408
Równoważność wariancji zmiennych latentnych	832,038 (358)	< 0,001	5,275 (7)	0,626	0,972	0,001	0,050 [0,046; 0,054]	0,000	0,039	0,008	70889,415	71429,994	71105,971
Równoważność kowariancji zmiennych latentnych	881,121 (379)	< 0,001	49,039 (21)	0,001	0,969	–0,003	0,050 [0,046; 0,055]	0,000	0,049	0,010	70929,641	71358,924	71101,611

Adnotacja. $p\Delta \chi^2$ diff – istotność różnicy pomiędzy wartościami statystyki chi-kwadrat, *robust CFI* – odporny porównawczy indeks dopasowania, *robust RMSEA* – odporny pierwiastek średniokwadratowego błędu aproksymacji, 90% CI – przedział ufności, SRMR – średni standardyzowany kwadrat reszt, AIC – kryterium informacyjne Akaikego, BIC – kryterium informacyjne Bayesa, SABIC – kryterium informacyjne Bayesa z poprawką na wielkość próby. Wartości ΔCFI , $\Delta RMSEA$ i $\Delta SRMR$ poniżej wartości traktowanych jako krytyczne. Wartości SABIC i BIC (poza AIC do równości reszt) obniżają się wraz z bardziej restryktywnymi wymaganiami równoważności.

Z tego względu zastosowano mocne testy różnic, bazujące na wartościach przyrostowych statystyki χ^2 – dokładniej robustowej statystyki Satorra-Bentler. Analiza wykazała porównywalność ładunków oraz stałych regresji (interceptów) dla obu badań już przy odmienności błędów pomiaru. W praktyce zatem oba pomiary objawów PTSD z zastosowaniem PCL-5 można uznać za równoważne. Wyniki te uzasadniły zatem połączenie danych z obu badań dla wersji losowej i stałej oraz testowanie równoważności pomiarowej obu wersji inwentarza.

WYNIKI

Dla obu wersji inwentarza nie stwierdzono różnic w liczbie traumatycznych wydarzeń: $t(739) = 0,12$; $p = 0,91$ (grupa stała: $M = 5,56$; $SD = 3,40$; grupa losowa: $M = 5,55$; $SD = 3,39$) oraz nasileniu objawów PTSD: $t(739) = 1,75$; $p = 0,08$ (grupa stała: $M = 26,29$; $SD = 18,28$; grupa losowa: $M = 27,20$; $SD = 18,22$), choć uzyskano je dla częstości diagnozy PTSD – wyższej w wersji losowej: test McNemara $\chi^2(1, N = 740) = 6,68$; $p = 0,01$ (wersja stała: 295 diagnoz; 39,9%; wersja losowa: 330; 44,6%); zgodność diagnoz: $\chi^2(1, N = 740) = 203,51$; $p < 0,001$; współczynnik kappa $\kappa = 0,52$ (zgodnych diagnoz: 567/740; 76,6%)⁵. W analizach stwierdzono, że model 7-czynnikowy wykazuje lepsze dopasowanie niż inne modele (najniższe wartości AIC i SABIC, choć BIC był wyższy w stosunku do modeli 6-czynnikowych; wyniki takie rejestrowano także w porównaniach modelu 7-czynnikowego z modelami 6-czynnikowymi; Wortmann i in., 2016). Wynik ten wskazuje ponownie na zasadność rozpatrywania 7-czynnikowego modelu struktury objawów PTSD także dla obu wersji: stałej i losowej pozycji PCL-5 (zob. tabela 6).

5 Warto podkreślić, że alternatywna procedura rozpoznania PTSD, która wg autorów PCL-5 zakłada dychotomizację wyników liczbowych wg wyników 31–33 (na użytek tej pracy przyjęto próg odsiewowy 30/31, bowiem górna wartość, czyli 33 punkty, prowadziła do znacznego niedoszacowania częstości diagnoz), nie wykazała różnic w częstości rozpoznania PTSD dla obu procedur: test McNemara $\chi^2(1, N = 1480) = 3,53$; $p = 0,06$; zgodność diagnoz: $\chi^2(1, N = 1480) = 970,44$; $p < 0,001$; współczynnik kappa $\kappa = 0,81$ (zgodnych diagnoz: 1343/1480; 90,7%; rozpoznanie PTSD: 602/1480; 40,7%; w stosunku do procedury konfiguracyjnej wg DSM-5: 625; 42,2%). Zastosowanie tej procedury nie ujawniło różnic w diagnozie PTSD między oboma pomiarami: test McNemara $\chi^2(1, N = 740) = 1,21$; $p = 0,27$; jak też między wersjami stałą i losową: test McNemara $\chi^2(1, N = 740) = 2,15$; $p = 0,08$ – odmiennie niż zostało to stwierdzone dla procedury diagnozy konfiguracyjnej. Dla diagnozy konfiguracyjnej nie stwierdzono różnic w częstości rozpoznania jednorodnych czynnikowo kryteriów B i C PTSD w DSM-5, a jedynie dla kategorii D, obejmującej dwa czynniki: test McNemara $\chi^2(1, N = 740) = 5,01$; $p = 0,025$ (wersja stała: 55,7% rozpoznania; wersja losowa: 60,0% rozpoznania) oraz kategorii E, obejmującej trzy czynniki: test McNemara $\chi^2(1, N = 740) = 24,34$; $p = 0,01$; wersja stała: 54,1% rozpoznania; wersja losowa: 63,1% rozpoznania).

Tabela 6

Wskaźniki dobroci dopasowania modeli struktury objawów PTSD dla wersji stałej i losowej PCL-5

Model	<i>robust</i> χ^2 (df)	<i>p</i>	AIC	BIC	SABIC	<i>robust</i> RMSEA [90% CI]	<i>robust</i> CFI	<i>robust</i> TLI	SRMR
1-czynnikowy ogólny	1559,909 (340)	< 0,001	71895,597	72531,573	72150,369	0,085 [0,081; 0,089]	0,923	0,914	0,035
4-czynnikowy model DSM-5	928,641 (328)	< 0,001	70972,247	71671,820	71252,495	0,060 [0,056; 0,065]	0,962	0,956	0,028
4-czynnikowy model dysforyczny	1046,685 (328)	< 0,001	71151,742	71851,315	71431,990	0,066 [0,062; 0,071]	0,955	0,947	0,029
5-czynnikowy model dysforyczno-pobudzeniowy	917,470 (320)	< 0,001	70970,847	71712,819	71268,080	0,061 [0,056; 0,066]	0,962	0,955	0,028
6-czynnikowy model poznawczo-odrętwieniowy	791,693 (310)	< 0,001	70800,749	71595,719	71119,213	0,055 [0,051; 0,060]	0,970	0,963	0,027
6-czynnikowy model anhedoniczny	795,994 (310)	< 0,001	70807,231	71602,201	71125,695	0,056 [0,051; 0,060]	0,970	0,963	0,027
7-czynnikowy model hybrydowy	742,550 (298)	< 0,001	70747,297	71605,864	71091,238	0,054 [0,049; 0,059]	0,972	0,965	0,027

Adnotacja. AIC – kryterium informacyjne Akaikego, BIC – kryterium informacyjne Bayesa, SABIC – kryterium informacyjne Bayesa z poprawką na wielkość próby, *robust* CFI – odporny porównawczy indeks dopasowania, *robust* RMSEA – odporny pierwiastek średniokwadratowego błędu aproksymacji, 90% CI – przedział ufności, *robust* TLI – wskaźnik Tuckera-Lewisa, SRMR – średni standaryzowany kwadrat reszt. Czcionką pogrubioną oznaczono najlepsze wskaźniki AIC, BIC i SABIC.

Analiza równoważności wykazała, że wszystkie wartości wskaźników przyrostowych są stosunkowo niewielkie i wskazują na pełną równoważność pomiarową PCL-5 dla obu wersji inwentarza (zob. tabela 7).

Tabela 7Równoważność pomiarowa modelu 7-czynnikowego dla obu wersji losowej i statej *PC1-5*

Model	<i>robust</i> χ^2 (df)	<i>p</i>	$\Delta \chi^2$	<i>p</i> $\Delta \chi^2$ diff	<i>robust</i> CFI	Δ CFI	<i>robust</i> RMSEA [90% CI]	SRMR	Δ SRMR	AIC	BIC	SABIC
Równoważność struktury	742,550 (298)	< 0,001	–	–	0,972	–	0,054 [0,049; 0,059]	0,027	–	70747,297	71605,864	71091,238
Równoważność ładunków	775,160 (311)	< 0,001	32,733 (13)	0,002	0,972	0,000	0,054 [0,049; 0,058]	0,033	0,006	70751,139	71540,809	71067,480
Równoważność stałych regresji (<i>intercepts</i>)	820,218 (324)	< 0,001	50,298 (13)	< 0,001	0,970	–0,002	0,054 [0,050; 0,059]	0,033	0,000	70775,437	71496,210	71064,178
Równoważność błędów pomiaru (<i>residuals</i>)	909,915 (344)	< 0,001	86,172 (20)	< 0,001	0,966	–0,004	0,056 [0,052; 0,061]	0,034	0,001	70870,199	71484,975	71116,478
Równoważność średnich zmiennych latentnych	929,842 (351)	< 0,001	20,506 (7)	0,005	0,965	–0,001	0,056 [0,052; 0,061]	0,035	0,001	70876,762	71454,440	71108,179
Równoważność wariancji zmiennych latentnych	934,810 (358)	< 0,001	12,912 (7)	0,070	0,965	0,000	0,056 [0,051; 0,060]	0,039	0,004	70871,024	71411,603	71087,580
Równoważność kowariancji zmiennych latentnych	1007,239 (379)	< 0,001	60,541 (21)	< 0,001	0,962	–0,003	0,056 [0,052; 0,061]	0,039	0,000	70929,641	71358,924	71101,611

Adnotacja: p $\Delta \chi^2$ *diff* – istotność różnicy pomiędzy wartościami statystyki chi-kwadrat, *robust* CFI – odporny porównawczy indeks dopasowania, *robust* RMSEA – odporny pierwiastek średniokwadratowego błędów aproksymacji, 90% CI – przedział ufności, SRMR – średni standaryzowany kwadrat reszt, AIC – kryterium informacyjne Akaikego, BIC – kryterium informacyjne Bayesa, SABIC – kryterium informacyjne Bayesa z poprawką na wielkość próby. Wartości Δ CFI, Δ RMSEA i Δ SRMR poniżej wartości traktowanych jako krytyczne. Wartości BIC obniżały się wraz z bardziej restryktywnymi wymaganiami równoważności.

Po zastosowaniu mocnego testu różnic, bazującego na wartościach przyrostowych robustowej statystyki Satorra-Bentler χ^2 stwierdzono jednak, że obie wersje nie wykazują porównywalności ładunków, stałych regresji, błędów pomiaru oraz średnich, wariancji i kowariancji zmiennych latentnych. W praktyce zatem obie wersje PCL-5 można uznać za równoważne jedynie w aspekcie konfiguracyjnym – ich charakterystyka psychometryczna jest odmienna. Różnice te nie są szczególnie duże, ale na tyle znaczące, aby możliwe było uznanie wersji stałej i losowej inwentarza za psychometrycznie nierównoważne. Obie wersje wydają się zatem mierzyć te same konstrukty, ale wykazują nieco odmienną charakterystykę pomiarową.

W tabeli 8 zaprezentowano ładunki czynnikowe dla pozycji PCL-5 dla wersji stałej i losowej oraz ładunki siedmiu podskal dla czynnika ogólnego.

Tabela 8

Ładunki czynnikowe pozycji PCL-5 dla siedmiu czynników wersji stałej i losowej oraz dla siedmiu podskal i czynnika ogólnego

Objaw wg DSM-5/pozycja PCL-5	Wersja stała	Wersja losowa	Wersja stała – współczynnik omega	Wersja losowa – współczynnik omega
B1– Intruzja (I)	0,797	0,851	Podwymiar I	Podwymiar I
B2	0,791	0,735	0,920	0,911
B3	0,836	0,837		
B4	0,869	0,860		
B5	0,867	0,811		
C1 – Unikanie (UN)	0,824	0,710	Podwymiar UN	Podwymiar UN
C2	0,894	0,809	0,850#	0,733
D1 – Negatywny afekt (NA)	0,621	0,640	Podwymiar NA	Podwymiar NA
D2	0,849	0,818	0,870	0,858
D3	0,815	0,783		
D4 -	0,865	0,842		
D5 – Anhedonia (AN)	0,835	0,798	Podwymiar AN	Podwymiar AN
D6	0,846	0,809	0,885#	0,842
D7	0,863	0,796		
E1 – Zachowania eksternalizacyjne (ZE)	0,852	0,817	Podwymiar ZE	Podwymiar ZE
E2	0,758	0,670	0,787#	0,715

Ciąg dalszy tabeli na następnej stronie

Ciąg dalszy tabeli z poprzedniej strony

Objaw wg DSM-5/pozycja PCL-5	Wersja stała	Wersja losowa	Wersja stała – współczynnik omega	Wersja losowa – współczynnik omega
E3 – Pobudzenie lękowe (PL)	0,727	0,715	Podwymiar PL	Podwymiar PL
E4	0,886	0,815	0,790#	0,741
E5 – Pobudzenie dysforyczne (PD)	0,901	0,846	Podwymiar PD	Podwymiar PD
E6	0,792	0,746	0,837#	0,776
Czynnik	Wersja stała	Wersja losowa	Cała skala (20 pozycji)	Cała skala (20 pozycji)
I	0,859 (0,897)	0,928 (0,968)	0,966	0,965
UN	0,778 (0,838)	0,813 (0,944)	Czynnik ogólny (siedem podwymiarów)	Czynnik ogólny (siedem podwymiarów)
NA	0,893 (0,947)	0,912 (0,980)	0,947	0,958#
AN	0,879 (0,932)	0,878 (0,960)		
ZE	0,839 (0,944)	0,804 (0,949)		
PL	0,851 (0,966)	0,868 (0,988)		
PD	0,835 (0,919)	0,814 (0,927)		

Adnotacja. # – istotne różnice współczynnika rzetelności pomiaru (Hakstian i Whalen, 1976): UN – $\chi^2(1, N = 1480) = 30,22$; $p = 0,001$; AN – $\chi^2 = 12,36$; $p = 0,002$; ZE – $\chi^2 = 7,80$; $p = 0,005$; PL – $\chi^2 = 4,05$; $p = 0,045$; PD – $\chi^2 = 9,29$; $p = 0,002$; czynnik ogólny (siedem podwymiarów) – $\chi^2 = 8,54$; $p = 0,004$. Dla podskal obejmujących pozycje wartość współczynnika omega obliczono na podstawie ładunków czynnikowych estymowanych przez *lavaan* w środowisku R (w pozostałych przypadkach poprzez SPSS, wersja 28.0.1.0 (142)).

Należy podkreślić, że dane nie umożliwiały estymacji modelu z czynnikiem ogólnym, nasycającym wszystkie pozycje PCL-5 oraz dowolną (także mniejszą niż siedem) liczbą czynników specyficznych. Zrezygnowano też z analizy równoważności modelu hierarchicznego: 20 pozycji → siedem czynników I rzędu → czynnik ogólny, który jedynie zastosowano do analizy ładunków czynnikowych dla siedmiu zmiennych latentnych (ładunki dla tej analizy podano w nawiasach w dolnej części tabeli 7). Podstawowe ładunki, jak też współczynniki rzetelności pomiaru (omega) uzyskano zatem w odrębnych

analizach. Były one inspirowane oczekiwaniem, że wersja stała o zgrupowanych pozycjach w stosunku do wersji losowej powinna wykazywać wyższą trafność zbieżną i różnicową, tj. wyższe korelacje pozycji wchodzących w obręb danego czynnika i niższe korelacje pozycji między różnymi czynnikami. Na poziomie analizy czynnikowej oznaczało to oczekiwanie wyższych ładunków zbieżnych dla wersji stałej oraz niższych ładunków dla skal badających dany podwymiar.

Wstępne analizy korelacji potwierdziły to przypuszczenie: przy zbliżonej średniej interkorelacji wszystkich 20 pozycji wynoszącej 0,589 (zakres: 0,418–0,773) dla wersji stałej i 0,580 (zakres: 0,421–0,747) dla wersji losowej (co oznacza zbliżoną spójność odpowiedzi na pozycje całego inwentarza w obu wersjach i zbliżoną rzetelność pomiaru); średnia korelacji w obrębie pozycji tworzących czynniki wynosiła 0,681 dla wersji stałej i 0,605 dla wersji losowej (zarejestrowano także istotne różnice współczynnika rzetelności pomiaru dla pięciu podskal). Z kolei dla czynników było odwrotnie – średnia korelacji wynosiła 0,718 dla wersji stałej i 0,738 dla losowej. Odpowiednie średnie ładunków czynnikowych zaś kolejno 0,824 i 0,785 dla pozycji oraz 0,846 i 0,860 dla podskal (0,920 i 0,954 dla siedmiu zmiennych latentnych). Różnice te nie są algebraicznie duże, ale wystarczyły do wystąpienia nierównoważności pomiarowej obu wersji PCL-5. W rezultacie można zatem sądzić, że struktura 7-czynnikowa nie jest efektem grupowania pozycji, ale przy ich zgrupowaniu odpowiedzi na pytania blisko ułożone są nieco bardziej spójne oraz mniej spójne, gdy dystans między nimi rośnie. Przy losowej kolejności pozycji kwestia dystansu wydaje się tracić na znaczeniu i zmniejsza się różnica w spójności odpowiadania na pozycje badające dane i odmienne konstrukty. W konsekwencji efekt ten wydaje się odpowiedzialny za nieco większą liczbę diagnoz PTSD bazujących na konfiguracji objawów dla wersji losowej niż stałej (pomijamy w tym miejscu problem bazowania diagnozy na czterech kryteriach objawów wg DSM-5, pomimo iż wyniki tej analizy wskazywałyby na konieczność rozpoznania aż siedmiu kategorii). Prowadzi on bowiem do nieco większej spójności pomiędzy kategoriami objawów PTSD. Wskazują na to współczynniki kappa zgodności rozpoznania czterech kryteriów PTSD dla wersji stałej i losowej: $M = 0,57$ i $M = 0,60$, odpowiadające większej liczbie zgodnych diagnoz pomiędzy kategoriami: średnio 584/740 (78,9%) dla wersji stałej (zakres: 564–584) oraz średnio 600/740 (81,1%) dla wersji losowej (zakres: 575–616), a także istotnie większa liczba rozpoznanych kryteriów PTSD w wersji losowej w porównaniu z wersją stałą ($M = 2,49$; $SD = 1,62$ i $M = 2,36$; $SD = 1,62$; test Wilcoxona $Z = 2,67$; $p = 0,008$).

Uzyskane odmienności dla obu wersji nie są przy tym efektem aprobaty społecznej – korelacje wyników PCL-5 ze skalą aprobaty społecznej wynosiły

–0,204* dla wersji stałej i –0,193* dla wersji losowej ($\eta^2 = -0,171^*$ i $-0,160^*$ dla diagnozy PTSD) oraz dla podgrupy o wersji stałej –0,219* i losowej –0,180* w pierwszym pomiarze (przy korelacji –0,445* ze skalą neurotyczności). Podobnie nieznaczne były różnice korelacji między wersjami PCL-5 a skalą neurotyczności: 0,490* dla wersji stałej i 0,524* (a więc były nawet nieznacznie wyższe) dla wersji losowej; nieco większe rozrzuty zarejestrowano dla obu pomiarów: 0,528* (pierwszy pomiar) oraz 0,489* (drugi pomiar) oraz podgrupy o wersji stałej 0,505* i losowej 0,551* w pierwszym pomiarze (* $p < 0,05$ dla wszystkich zaprezentowanych powyżej współczynników korelacji).

DYSKUSJA

Podstawowa konkluzja wynikająca ze zrealizowanych analiz dotyczy kryteriów diagnozy PTSD w systemie diagnostycznym DSM. Na wątpliwości dotyczące konceptualizacji PTSD wskazywano już wielokrotnie i dotyczyły one także poprzednich wersji, a nie tylko DSM-5 (Andreasen, 2004). Główne aspekty krytyki dotyczyły podstaw empirycznych rozpoznania zaburzenia. W DSM-5 w stosunku do DSM-IV uwzględniono jedynie wyniki analiz Kinga i in. (1998), wskazujące na konieczność wyodrębnienia objawów unikania jako osiowej kategorii PTSD. Nie uwzględniono zupełnie innych modeli, jak model 5-czynnikowy Elhai i in. (2011), przyjmujący za osiowe objawy intruzji, unikania, odrętwienia, pobudzenia lękowego i dysforycznego, który doczekał się szerokiego wsparcia empirycznego (zob. Dragan i in., 2012; Witte i in., 2015). Wprowadzone zmiany w DSM-5 miały raczej charakter korekcyjny niż rewolucyjny. Stosunkowo szybko uzyskano szerokie dane wspierające 7-czynnikowy model hybrydowy: intruzji, unikania, negatywnego afektu, anhedonii, zachowań eksternalizacyjnych, pobudzenia lękowego i dysforycznego dla różnych narzędzi diagnostycznych (np. Armour i in., 2015; Cyniak-Cieciura i in., 2017; Zawadzki i in., 2015). Uzyskane w tej pracy wyniki wpisują się w te ustalenia empiryczne, podobnie zresztą jak wyniki analiz równoważności międzykulturowej (Cyniak-Cieciura i in., w tym tomie) i wskazują, że struktura czterech osiowych kategorii, postulowana w DSM-5, odbiega od empirycznie stwierdzanej struktury objawów PTSD. Zdecydowanie zatem wymagana byłaby rekonceptualizacja kryteriów rozpoznania PTSD w DSM-5. Jedno z możliwych rozwiązań zakłada wielokategorialną klasyfikację objawów i zrezygnowanie ze struktury prostej (oraz – w efekcie – powiązania

z kolejnością ekspozycji objawów w narzędziach diagnostycznych). Modele te (zob. Elklit i Shevlin, 2007) mają niestety dość oczywiste wady diagnostyczne, bowiem jeden objaw zaburzenia może umożliwić jednocześnie rozpoznanie kilku kategorii objawowych i nadszacować znacznie częstość rozpoznania PTSD. Inna możliwość związana jest z wymaganiem spełnienia wszystkich siedmiu kategorii, stosownie do modelu 7-czynnikowego (wymagane byłoby zwiększenie liczby wymaganych objawów PTSD z sześciu do siedmiu), albo stworzenia subkategorii objawów poznawczych i anhedonii w obrębie kryterium D (wymagane rozpoznanie obu subkategorii dla kryterium D) oraz zachowań eksternalizacyjnych, pobudzenia lękowego i pobudzenia dysforycznego w obrębie kryterium E (wymagane rozpoznanie dwóch z trzech subkategorii dla kryterium E dla zachowania wymaganej liczby sześciu objawów), na wzór klasyfikacji zbliżonej do przyjętego wcześniej rozwiązania w ICD-10 (*International Classification of Diseases, 10th Revision*; Pużyński, 1998). Konsekwencje takich klasyfikacji (także w efekcie zwiększenia liczby wymaganych objawów z sześciu do siedmiu) dla częstości rozpoznania PTSD wymagałyby odpowiednich analiz symulacyjnych oraz weryfikacyjnych – można jednak sądzić, że mogłoby to sprzyjać ograniczeniu częstości rozpoznania PTSD oraz generowaniu zwiększonej liczby diagnoz fałszywie negatywnych, zwłaszcza wśród osób o łagodniejszej formie zaburzenia (co skutkowałoby problemami społecznymi, wynikającymi z ograniczenia możliwości uzyskania pomocy terapeutycznej). Inna możliwość wiąże się z ograniczeniem objawów PTSD do stosunkowo nielicznej grupy objawów podstawowych – osiowych, jak zaproponowano w systemie diagnostycznym ICD-11 (World Health Organization, 2018; por. Zawadzki i Popiel, 2014), tj. objawów intruzji (objawy B2 i B3 wg DSM-5), unikania (C1 i C2) oraz pobudzenia lękowego (E3 i E4). Rozpoznanie tych objawów umożliwiałoby diagnozę tzw. prostego PTSD, zaś pozostałe objawy wg DSM-5 zostałyby zaliczone do tzw. złożonego PTSD (będącego efektem wielokrotnej traumatyzacji, choć nadal pozostaje kwestia odróżnienia ciężkiej postaci PTSD po jednokrotnej traumatyzacji – zredukowanej w tym przypadku do ciężkości objawów osiowych, z pominięciem ich zakresu klinicznego). Problem ten wymaga szerszej dyskusji teoretycznej i danych uzasadniających wyodrębnienie objawów osiowych i dodatkowych. Być może pomocne w tym przypadku będą wyniki analiz sieciowych PTSD (Birkeland i in., 2020; Jiang i in., 2020), które umożliwiają wskazanie objawów kluczowych i pochodnych, choć czynnikiem ograniczającym – ale zrozumiałym w świetle danych uzyskanych w tej pracy – są dane wskazujące na znaczące rozbieżności między

analizami objawów w wersjach narzędzi diagnostycznych o stałej i losowej kolejności objawów (Trachik i in., 2020).

Druga kwestia dotyczy wątpliwości, czy stwierdzana struktura objawów PTSD jest pochodną od ich dystansu fizycznego, a więc jest swego rodzaju artefaktem diagnostycznym. Teza, że może ona wynikać z kolejności ekspozowania objawów w narzędziach badawczych, a taka była i jest główna praktyka badania objawów PTSD, czyli zamieszczanie ich w kolejności wskazywanej w systemie DSM, była formułowana już wcześniej w odniesieniu do DSM-5 i bazujących na nim narzędzi diagnostycznych (Marshall i in., 2013; Witte i in., 2015). W tym przypadku odpowiedź na podstawie wyników naszych analiz jest jednoznaczna – struktura 7-czynnikowa może być stwierdzona niezależnie od kolejności ekspozowania pozycji w inwentarzu diagnostycznym czy wersji o losowej oraz stałej (tj. odpowiadającej kryteriom DSM-5) kolejności objawów PTSD. Mniej jednoznaczna jest natomiast odpowiedź na pytanie o problem równoważności pomiarowej wersji stałej i losowej PCL-5, wychodzącej poza ekwiwalencję konfiguralną. Należy podkreślić, że wniosek o nierównoważności obu wersji jest oparty o dość rygorystyczne kryteria testowania. Wielkości zmiany wskaźników dobroci dopasowania – rutynowo stosowane obecnie w psychologii – były bowiem nieznaczne. Odmienności pojawiły się dopiero przy uwzględnieniu rygorystycznych kryteriów, ale przyjęliśmy je ze względu na wyniki analiz dla obu pomiarów, które wskazywały na ich równoważność pomiarową. W tym sensie na metodologicznie solidnych podstawach mogliśmy wykazać, że w tej samej grupie oba pomiary są równoważne, a nierównoważność ujawnia się dopiero w analizie dwóch wersji inwentarza. Trzeba też podkreślić, że o wadze metodologicznej analiz decyduje też uzyskanie danych w tej samej grupie osób badanych, co uwolniło uzyskane wyniki od obciążenia wynikającego z odrębności demograficznych i psychologicznych oraz porównywalności odmiennych prób. Oczywiście jest w tym przypadku kwestia zrealizowania analiz na poziomie wewnątrzsobowym – zastosowane zostały procedury jak dla danych międzyosobowych, przez co odmienności dla obu wersji PCL-5 mogły być niedoestymowane. W tym kontekście powstaje też kwestia zarejestrowanej zmienności odpowiedzi między dwoma pomiarami – nie tylko objawów zaburzenia, ale też doświadczonej traumatyzacji, zwłaszcza dla tak krótkiego czasu, jakim jest okres dwóch tygodni (ale mieszający się w kryterium F wg DSM-5 jednego miesiąca trwania objawów; American Psychiatric Association, 2013). Uzyskane dane wskazują na daleką od perfekcyjnej zgodności, ale jest to w większym stopniu pochodną słabości

pomiaru w psychologii (Zawadzki, 2006) niż diagnozy PTSD czy samego inwentarza PCL-5. Decydować o tych rozbieżnościach może odmienna aktualizacja negatywnych doświadczeń w różnym czasie, jak też realna zmiana, np. w postaci fluktuacji nasilenia objawów (Biggs i in., 2019) czy też występująca w następstwie kolejnej traumatyzacji (co wykazano w trakcie pandemii COVID-19 przy porównywalnej z prezentowanymi badaniami stałości diagnoz PTSD: 81,5%; korelacja nasilenia objawów: 0,754 w grupie 247 osób badanych po sześciu tygodniach; Zawadzki, 2020).

Dalsze analizy wskazały na możliwą przyczynę braku równoważności pomiarowej obu wersji inwentarza – jak zakładano, wynika ona z większych fluktuacji spójności odpowiedzi w wersji stałej i w efekcie wyższej dla tej wersji trafności zbieżnej i różnicowej podskal. Z kolei – jak sądzymy – większa spójność odpowiedzi na pozycje między kategoriami dla wersji losowej może skutkować nieco większą częstością rozpoznania PTSD. Trudno rozstrzygnąć, czy jest to nadszacowanie dla wersji losowej, czy niedoestymowanie dla wersji stałej – w tym przypadku konieczne byłyby odpowiednie badania weryfikacyjne z zastosowaniem referentnego kryterium. Dla badań diagnostycznych PTSD przyjmuje się, że „złotymi standardami” diagnostycznymi są wywiady, a więc SCID-5 (*Structured Clinical Interview for DSM-5 Personality Disorders*) lub CAPS-5 (*Clinician-Administered PTSD Scale for DSM-5*; First i in., 2014; Weathers i in., 2018). Kwestia ta dotyczy także alternatywnej procedury diagnozowania PTSD na podstawie PCL-5 (zob. przypis 6). Zastosowanie tych bardzo czaso- i pracochłonnych metod diagnostycznych w prezentowanych badaniach przekraczało jednak nasze możliwości, toteż problem ten nie może zostać rozstrzygnięty.

Obie wersje wydają się w bardzo zbliżonym stopniu podatne na obciążenie odpowiedzi aprobatą społeczną (jak też wykazują zbliżone skorelowanie z neurotycznością). W istocie zatem oczekiwanie, iż stała kolejność pozycji prowadzi do nadspójności odpowiedzi, skutkującej uwikłaniem w pomiar czynników zniekształcających, które nie są związane z badanymi treściami, nie znalazła potwierdzenia. Należy jednak podkreślić, że główny wątek badań dotyczących tej kwestii był prowadzony w obszarze ocen zawodowych (Schriesheim, 1981; Schriesheim, Kopelman i Solomon, 1989; Schriesheim, Solomon i Kopelman, 1989), gdzie możliwe obciążenie samoopisu może być wyższe niż w prezentowanych badaniach (zwłaszcza ze względu na zastosowaną procedurę anonimizacji danych osobowych). Inna kwestia dotyczy jednak także możliwego uzyskiwania uogólnionej nadspójności odpowiedzi

w wersji stałej w stosunku do wersji losowej, co sugerowali Schriesheim i in. (Schriesheim, Kopelman i Solomon, 1989; Schriesheim, Solomon i Kopelman, 1989). Efekt taki faktycznie stwierdzono we wstępnych analizach (choć nie w badaniach właściwych), ale w tym przypadku odmienności mogły wynikać z nierównoważności badanych grup.

Dotychczasowe rozważania prowadzą w końcu do kluczowego pytania, zadawanego przez widzów filmów kryminalnych, a mianowicie „kto zabił?”. W tym przypadku, czy stosować wersję losową, czy stałą? Problem jest o tyle istotny, że obecnie wersje losowe znacznie się spopularyzowały ze względu na dostępność procedur losowania na platformach badawczych, a jednocześnie następuje odejście od narzędzi drukowanych, typu „papier-ołówek”, w stronę badań online. Jak sądzymy, problem ten nie ma szczególnie dużego znaczenia – efekty zróżnicowania ekwiwalencji diagnostycznej były wyraźne, ale dość słabe. Prawdopodobnie zbliżone odrębności byłyby możliwe do zarejestrowania, gdyby analizie poddać odrębne narzędzia samoopisowe, np. PDS-5 (*Posttraumatic Diagnostic Scale-5*) i PCL-5 czy wersje drukowane-online inwentarzy (choć ta teza także wymaga weryfikacji empirycznej). Jeśli celem badania jest replikacja struktury objawów czy równoważności pomiarowej narzędzia w różnych grupach, to bardziej uzasadnione wydaje się zastosowanie wersji stałej – dla niej bowiem łatwiej będzie można wykazać dopasowanie danej struktury objawów do danych ze względu na bardziej wyrazistą trafność zbieżną i różnicową (zob. wyniki analiz międzykulturowych dla wersji stałej: Cyniak-Cieciura i in., w tym tomie). W przypadku wersji losowej być może trzeba się także liczyć z ewentualnością większej liczby rozpoznań PTSD oraz możliwe, że także z większą liczbą diagnoz fałszywie pozytywnych. Jednakże rozbieżności diagnostyczne dla obu wersji nie były porażające – różnice częstości diagnoz PTSD tylko nieco przekraczały symetryczne rozbieżności wynikające z przypadku, a te były stwierdzone w każdej ze zrealizowanych analiz, w tym także dla obu pomiarów w formie zgodności dalekiej od doskonałości. Inne pytanie dotyczy konstruktu, dla którego bardziej uzasadnionym wyborem jest wersja stała albo losowa. Schriesheim, Solomon i Kopelman (1989) sądzili, że dla mocnych pomiarowo skal wybór wersji nie ma żadnego znaczenia, ale dla skal słabych – pozycje raczej należy grupować. W przypadku PTSD – jak wykazaliśmy – wybór może zależeć od rygorystyczności kryteriów oceny dopasowania i z racji dużej wartości diagnostycznej PCL-5 pozostaje nadal margines na pogorszenie. A co dla innych konstruktów teoretycznych? Zwykle zmienne osobowościowe są badane poprzez losowo rozmieszczane wskaźni-

ki i jest to sugerowane teoretycznie ze względu na wymóg wyeliminowania nadspójności odpowiedzi oraz uniezależnienie odpowiedzi od kontekstu, czyli treści innych pozycji itd., co ma zapewnić też ich zróżnicowanie międzytreściowe (Zawadzki, 2006). Konstrukty kliniczne zwykle zaś są badane wersjami inwentarzy zakładającymi grupowanie, np. inwentarz PBQ (*Personality Beliefs Questionnaire*) do badania zaburzeń osobowości (Zawadzki i in., 2017), a niekiedy wręcz – etykietowanie np. w formie rozdzielania pul pozycji – inwentarz YSQ (*Young Schema Questionnaire*) do badania schematów (Staniaszek i Popiel, 2017). Podobnie dotyczy to wywiadów klinicznych, np. SCID lub CAPS (First i in., 2014; Weathers i in., 2018). Dane empiryczne dotyczące znaczenia dla tych konstruktów procedury losowania-grupowania wskaźników wcale jednak nie są jednoznaczne, poza tym, że u podstaw narzędzi klinicznych – żartobliwie mówiąc – chyba stoi przekonanie autorów o słabości tych konstruktów lub oczekiwanej słabości psychometrycznej skal. Na podstawie tej pracy nie możemy zatem wskazać jednoznacznej rekomendacji dla tych zmiennych. Ale w końcu jest to już nie tyle inny film kryminalny, ale wręcz odmienny gatunek filmowy!

BIBLIOGRAFIA

- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (wyd. 4 zmien.).
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (wyd. 5). <https://doi.org/10.1176/appi.books.9780890425596>
- Andreasen, N. C. (2004). Acute and delayed posttraumatic stress disorders: A history and some issues. *American Journal of Psychiatry*, 161(8), 1321–1323. <https://doi.org/10.1176/appi.ajp.161.8.1321>
- Armour, C., Tsai, J., Durham, T. A., Charak, R., Biehn, T. L., Elhai, J. D., Pietrzak, D. (2015). Dimensional structure of DSM-5 posttraumatic stress symptoms: Support for a hybrid anhedonia and externalizing behaviors model. *Journal of Psychiatric Research*, 61, 106–113. <https://doi.org/10.1016/j.jpsychires.2014.10.012>
- Barciś, S. (2021). *Kłamię więc jestem – eksploracja czynników wpływających na negowanie posiadania cech niepożądanych w testach osobowości* [niepublikowany maszynopis]. Wydział Psychologii, Uniwersytet Warszawski.
- Biggs, Q. M., Ursano, R. J., Wang, J., Krantz, D. S., Carr, R. B., Wynn, G. H., Adams, D. P., Dacuyan, N. M., Fullerton, C. S. (2019). Daily variation in post traumatic stress symptoms in individuals with and without probable post traumatic stress disorder. *BMC Psychiatry*, 19, artykuł 56. <https://doi.org/10.1186/s12888-019-2041-7>

- Birkeland, M. S., Green, T., Spiller, T. R. (2020). The network approach to posttraumatic stress disorder: A systematic review. *European Journal of Psychotraumatology*, 11, artykuł 1700614. <https://doi.org/10.1080/20008198.2019.1700614>
- Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., Domino, J. L. (2015). The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress*, 28, 489–498. <https://doi.org/10.1002/jts.22059>
- Borek, J. (2020). *Pamięć odczuwanego bólu a reaktywność emocjonalna i nasilenie objawów pourazowego zaburzenia stresowego (PTSD)* [niepublikowana praca magisterska]. Wydział Psychologii, Uniwersytet Warszawski.
- Bovin, M. J., Marx, B. P., Weathers, F. W., Gallagher, M. W., Rodriguez, P., Schnurr, P. P., Keane, T. M. (2016). Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. *Psychological Assessment*, 28(11), 1379–1391. <https://doi.org/10.1037/pas0000254>
- Brewin, C. R. (2005). Systematic review of screening instruments for adults at risk of PTSD. *Journal of Traumatic Stress*, 18(1), 53–62. <https://doi.org/10.1002/jts.20007>
- Brzeziński, J. (1980). *Elementy metodologii badań psychologicznych*. Wydawnictwo Naukowe PWN.
- Crowne, D. P., Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349–354. <https://doi.org/10.1037/h0047358>
- Cyniak-Cieciura, M., Staniaszek, K., Popiel, A., Pragłowska, E., Zawadzki, B. (2017). The structure of PTSD symptoms according to DSM-5 and ICD-11 proposal: A multi sample analysis. *European Psychiatry*, 44, 179–186. <https://doi.org/10.1016/j.eurpsy.2017.02.491>
- Dragan, M., Lis-Turlejska, M., Popiel, A., Szumiał, S., Dragan, W. (2012). The validation of the Polish version of the Posttraumatic Diagnostic Scale and its factor structure. *European Journal of Psychotraumatology*, 3, artykuł 18479. <https://doi.org/10.3402/ejpt.v3i0.18479>
- Elhai, J. D., Biehn, T. L., Armour, C., Klopfer, J. J., Frueh, B. C., Palmieri, P. A. (2011). Evidence for a unique PTSD construct represented by PTSD's D1–D3 symptoms. *Journal of Anxiety Disorders*, 25(3), 340–345. <https://doi.org/10.1016/j.janxdis.2010.10.007>
- Elklit, A., Shevlin, M. (2007). The structure of PTSD symptoms: A test of alternative models using confirmatory factor analysis. *British Journal of Clinical Psychology*, 46(3), 299–313. <https://doi.org/10.1348/014466506X171540>
- Feldman, J. (2019). *Związek między objawami pourazowego zaburzenia stresowego i objawami depresyjnymi a przedziałowością/punktowością stylu jedzenia* [niepublikowana praca magisterska]. Wydział Psychologii, Uniwersytet Warszawski.
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W. (2014). *SCID-I. Ustrukturalizowany wywiad kliniczny do badania zaburzeń z osi I DSM-IV-TR. Podręcznik*

- (red. naukowa wersji polskiej: A. Popiel, B. Zawadzki i E. Pragłowska). Pracownia Testów Psychologicznych Polskiego Towarzystwa Psychologicznego.
- Franke, G. H. (1997). "The whole is more than the sum of its parts": The effects of grouping and randomizing items on the reliability and validity of questionnaires. *European Journal of Psychological Assessment*, 13(2), 67–74. <https://doi.org/10.1027/1015-5759.13.2.67>
- Hakstian, A. R., Whalen, T. E. (1976). A k -sample significance test for independent alpha coefficients. *Psychometrika*, 41(2), 219–231. <https://doi.org/10.1007/BF02291840>
- Hu, L. T., Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jiang, W., Ren, Z., Yu, L., Tan, Y., Shi, C. (2020). A network analysis of Post-traumatic Stress Disorder symptoms and correlates during the COVID-19 pandemic. *Frontiers in Psychiatry*, 11, artykuł 568037. <https://doi.org/10.3389/fpsy.2020.568037>
- Kass, R. E., Raftery, A. E. (1995). Bayes factors. *Journal of American Statistical Association*, 90(430), 773–795.
- King, D. W., Leskin, G. A., King, L. A., Weathers, F. W. (1998). Confirmatory factor analysis of the Clinician-Administered PTSD Scale: Evidence for the dimensionality of posttraumatic stress disorder. *Psychological Assessment*, 10(2), 90–96. <https://doi.org/10.1037/1040-3590.10.2.90>
- Knowles, E. S. (1988). Item context effects on personality scales: Measuring changes the measure. *Journal of Personality and Social Psychology*, 55(2), 312–320. <https://doi.org/10.1037/0022-3514.55.2.312>
- Konarski, R. (2009). *Modele równań strukturalnych. Teoria i praktyka*. Wydawnictwo Naukowe PWN.
- Kotov, R., Gamez, W., Schmidt, F., Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, 136(5), 768–821. <https://doi.org/10.1037/a0020327>
- Lubiewska, K., Głogowska, K. (2018). Zastosowanie analizy równoważności pomiarowej w badaniach psychologicznych. *Polskie Forum Psychologiczne*, 23(2), 330–356. <https://doi.org/10.14656/PFP20180207>
- Marshall, G. N., Schell, T. J., Miles, J. N. V. (2013). A multi-sample confirmatory factor analysis of PTSD symptoms: What exactly is wrong with the DSM-IV structure? *Clinical Psychology Review*, 33(1), 54–66. <https://doi.org/10.1016/j.cpr.2012.10.004>
- Milfont, T. L., Fisher, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, 3(1), 111–130. <https://doi.org/10.21500/20112084.857>

- Ogińska-Bulik, N., Juczynski, Z., Lis-Turlejska, M., Merecz-Kot, D. (2018). Polska adaptacja PTSD Checklist PCL-5 for DSM-5. Doniesienie wstępne. *Przegląd Psychologiczny*, 61(2), 281–285.
- Popiel, A., Zawadzki, B., Bielecki, M., Mroziński, B., Pragłowska, E. (2021). Czy doświadczenie pandemii może prowadzić do pourazowego zaburzenia stresowego (PTSD)? Wyniki badania COVID-STRES. W: W. J. Paluchowski i L. Bakiera (red.), *Psychospołeczny obraz pierwszej fali pandemii COVID-19 w Polsce* (s. 45–58). Wydawnictwo Naukowe UAM.
- Pużyński, S. (red.). (1998). *Klasyfikacja zaburzeń psychicznych i zaburzeń zachowania w ICD-10. Badawcze kryteria diagnostyczne*. Uniwersyteckie Wydawnictwo Medyczne „Vesalius”, Instytut Psychiatrii i Neurologii.
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org>
- Raftery, A. E. (1995). Bayesian model selection in social research. *Social Methodology*, 25, 111–163. <https://doi.org/10.2307/271063>
- Rhemtulla, M., Brosseau-Liard, P., Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315>
- Rosseel, Y. (2015). *lavaan: An R package for structural equation modelling*. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rzeszutek, M., Lis-Turlejska, M., Palich, H., Szumiał, S. (2018). Polska adaptacja narzędzia pomiaru ekspozycji na traumatyczne zdarzenia według definicji DSM-5: Life Events Checklist for DSM-5 (LEC-5). *Psychiatria Polska*, 52(3), 499–510. <https://doi.org/10.12740/PP/OnlineFirst/69218>
- Schell, K. L., Oswald, F. S. (2013). Item grouping and item randomization in personality measurement. *Personality and Individual Differences*, 55(3), 317–321. <https://doi.org/10.1016/j.paid.2013.03.008>
- Schriesheim, C. A. (1981). The effect of grouping or randomizing items on leniency response bias. *Educational and Psychological Measurement*, 41(2), 401–411. <https://doi.org/10.1177/001316448104100219>
- Schriesheim, C. A., Kopelman, R. E., Solomon, E. (1989). The effect of grouped versus randomized questionnaire format on scale reliability and validity: A three-study investigation. *Educational and Psychological Measurement*, 49(3), 487–508. <https://doi.org/10.1177/001316448904900301>
- Schriesheim, C. A., Solomon, E., Kopelman, R. E. (1989). Grouped versus randomized format: An investigation of scale convergent and discriminant validity using

- LISREL confirmatory factor analysis. *Applied Psychological Measurement*, 13(1), 19–32. <https://doi.org/10.1177/014662168901300103>
- Simms, L. J., Watson, D., Doebbellin, B. N. (2002). Confirmatory factor analyses of posttraumatic stress symptoms in deployed and nondeployed veterans of the Gulf War. *Journal of Abnormal Psychology*, 111(4), 637–647. <https://doi.org/10.1037/0021-843X.111.4.637>
- Solomon, E., Kopelman, R. E. (1984). Questionnaire format and scale reliability: An examination of three modes of item presentation. *Psychological Reports*, 54(2), 447–452. <https://doi.org/10.2466/pr0.1984.54.2.447>
- Staniaszek, K., Popiel, A. (2017). Opracowanie i walidacja eksperymentalnej polskiej wersji skróconej Kwestionariusza schematów Younga (YSQ-S3-PL) do badania wczesnych schematów dezadaptacyjnych. *Roczniki Psychologiczne*, 20(2), 373–399. <https://doi.org/10.18290/rpsych.2017.20.2-5pl>
- Strus, W., Cieciuch, J., Rowiński, T. (2014). *Polska wersja IPIP-NEO-FFI-50 Goldberga* [niepublikowany maszynopis]. Instytut Psychologii, Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie.
- Trachik, B., Elliman, T. D., Ganulin, M. J., Dretsch, M. N., Riviere, L. A., Cabrera, O. A., Thomas, J. L., Hog, C. W. (2020). Order effects in PTSD network analysis: Important implications for diagnostic conceptualization, treatment refinement, and research. *Psychological Medicine*, 52(13), 2492–2499. <https://doi.org/10.1017/s0033291720004377>
- Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., Keane, T. M. (2013). *The Life Events Checklist for DSM-5 (LEC-5)*. https://www.ptsd.va.gov/professional/assessment/te-measures/life_events_checklist.asp
- Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., Keane, T. M., Marx, B. P. (2018). The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5): Development and initial psychometric evaluation in military veterans. *Psychological Assessment*, 30(3), 383–395. <https://doi.org/10.1037/pas0000486>
- Weathers, F. W., Ford, J. (1996). Psychometric review of the PTSD Checklist. W: B. H. Stamm (red.), *Measurement of stress, trauma, and adaptation* (s. 250–251). Sidran Press.
- Weathers, F. W., Huska, J. A., Keane, T. M. (1991). *The PTSD Checklist—Civilian Version (PCL-C)*. <https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp>
- Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., Schnurr, P. P. (2013). *The PTSD Checklist for DSM-5 (PCL-5)*. <https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp>
- Webb, M. A., Tangney, J. P. (2022). Too good to be true: Bots and bad data from Mechanical Turk. *Perspectives on Psychological Science*. <https://doi.org/10.1177/17456916221120027>
- World Health Organization (2018). *ICD-11. International statistical classification of diseases and related health problems* (wyd. 11 zmien.).

- Witowska, O. (2020). *Traumatyczne zdarzenia i inteligencja emocjonalna a objawy pourazowego zaburzenia stresowego i działania magiczne* [niepublikowana praca magisterska]. Wydział Psychologii, Uniwersytet Warszawski.
- Witte, T. K., Domino, J. L., Weathers, F. W. (2015). Item order effects in the evaluation of posttraumatic stress disorder symptom structure. *Psychological Assessment*, 27(3), 852–864. <https://doi.org/10.1037/pas0000089>
- Wojciechowska, K. (2021). *Objawy PTSD i depresji a zaburzenia snu po traumatycznych doświadczeniach* [niepublikowana praca magisterska]. Wydział Psychologii, Uniwersytet Warszawski.
- Wortmann, J. H., Jordan, A. H., Weathers, F. W., Resick, P. A., Dondanville, K. A., Hall-Clark, B., Foa, E. B., Young-McCaughan, S., Yarvis, J. S., Hembree, E. A., Mintz, J., Peterson, A. L., Litz, B. T. (2016). Psychometric analysis of the PTSD Checklist-5 (PCL-5) among treatment-seeking military service members. *Psychological Assessment*, 28(11), 1392–1403. <https://doi.org/10.1037/pas0000260>
- Zawadzki, B. (2006). *Kwestionariusze osobowości. Strategie i procedura konstruowania*. Wydawnictwo Naukowe Scholar.
- Zawadzki, B. (24–25 września 2020). *Zaburzenie stresowe pourazowe w dobie pandemii wirusa SARS-CoV-2?* [wykład]. Doktorancki Festiwal Otwartej Nauki Uniwersytetu Warszawskiego, Warszawa, Polska.
- Zawadzki, B., Popiel, A. (2014). Na rozstaju dróg: struktura objawów stresu pourazowego (PTSD) po DSM-5, a przed ICD-11. *Nauka*, 4, 69–86.
- Zawadzki, B., Popiel, A., Foa, E. B., Jakubowska, B., Cyniak-Cieciura, M., Pragłowska, E. (2015). The structure of symptoms of posttraumatic stress disorder (PTSD) according to DSM-5 and assessed by PDS-5. *Current Issues in Personality Psychology*, 3(1), 1–11. <https://doi.org/10.5114/cipp.2015.49662>
- Zawadzki, B., Popiel, A., Pragłowska, E., Newman, C. (2017). Specyfika dezadapacyjnych przekonań w zaburzeniach osobowości: charakterystyka psychometryczna polskiej translacji i tawestacji Kwestionariusza przekonań (Personality Beliefs Questionnaire, PBQ). *Roczniki Psychologiczne*, 20(2), 337–354. <https://doi.org/10.18290/rpsych.2017.20.2-4pl>
- Zercher, F., Schmidt, P., Cieciuch, J., Davidov, E. (2015). The comparability of the universalism value over time and across countries in the European Social Survey: Exact vs. approximate measurement invariance. *Frontiers in Psychology*, 6, artykuł 733. <https://doi.org/10.3389/fpsyg.2015.00733>

Książka ta jest adresowana do psychologów oraz studentów tego kierunku, do psychiatrów, neurologów, badaczy prowadzących dociekania naukowe w obszarze różnic indywidualnych, zdrowia i jakości życia [...]. Tom zawiera wartościowe z naukowego (a także dydaktycznego) punktu widzenia prace. Poziom trudności jest zróżnicowany, a zaprezentowany zbiór tekstów obejmuje teksty naukowe, których czytanie stawia czytelnikowi wysoce specjalistyczne wymagania w zakresie znajomości psychologicznej (klinicznej) i medycznej terminologii teoretycznej i diagnostycznej stosowanej w obszarze klinicznym.

prof. dr hab. Jerzy Brzeziński

Tom zredagowany przez Bogdana Zawadzkiego, Krzysztofa Fronczyka oraz Kamillę Bargiel-Matusiewicz dotyczy problematyki aktualnej, zawierając zarówno elementy koncepcyjno-teoretyczne, jak i przede wszystkim diagnostyczno-badawcze. [...] Zawiera on osiem rozdziałów poświęconych zagadnieniom z różnych obszarów diagnozy psychologicznej i przygotowanych przez znakomite grono specjalistów – badaczy i praktyków. Prace zamieszczone w tomie mają kliniczno-aplikacyjny charakter, choć niektóre silnie akcentują też wątki naukowo-badawcze. Tom zawiera prace poświęcone zróżnicowanej problematyce. Niektóre prezentują ogólniejsze analizy i refleksje – dotyczące nowego podejścia w badaniach naukowych służących wyjaśnianiu zaburzeń psychicznych, albo odnoszące się do aktualnego stanu i perspektyw rozwoju warsztatu diagnostycznego w danym obszarze praktyki klinicznej [...]. Wszystkie rozdziały tomu łączy bardzo wysoki poziom merytoryczny, aktualność podjętej tematyki oraz bezpośrednio i pośrednio implikacje dla współczesnej psychologicznej diagnostyki klinicznej. Dla osób zainteresowanych tą problematyką kolejny tom serii wydawniczej Wydziału Psychologii Uniwersytetu Warszawskiego pt. *Psychological Currents. Theory and Research* będzie niezwykle cenną i ważną lekturą.

dr hab. Włodzimierz Strus, prof. uczelni

www.LiberiLibri.pl

ISBN: 978-83-63487-67-6