

Psychiatr. Pol. 2015; 49(5): 1043–1070

PL ISSN 0033-2674 (PRINT), ISSN 2391-5854 (ONLINE)

www.psychiatriapolska.pl

DOI: <http://dx.doi.org/10.12740/PP/59064>

Dwupoziomowy ortogonalno-ukośny model Kwestionariusza Oceny Skuteczności Psychoterapii (Outcome Questionnaire, OQ-45.2) – polska adaptacja przy użyciu analizy czynnikowej

The orthogonal-oblique bi-level model of the Outcome Questionnaire (OQ-45.2) – The case of the Polish factorial normalization

Witold Simon¹, Piotr Śliwka², Jerzy A. Sobański³, Katarzyna Klasa⁴,
Paweł Sala⁵, Wojciech Żak⁶, Gregory Busath⁷, Michael J. Lambert⁷

¹ Konsonans. Ośrodek Psychoterapii w Warszawie

² Wydział Matematyczno-Przyrodniczy Szkoły Nauk Ścisłych UKSW w Warszawie

³ Katedra Psychoterapii UJ CM

⁴ Zakład Psychoterapii Szpitala Uniwersyteckiego w Krakowie

⁵ Praktyka prywatna w Warszawie

⁶ Klinika Psychiatrii Dorosłych UM w Poznaniu

⁷ Psychology Department, Brigham Young University, Provo, UT, USA

Summary

Aim. The aim of this project was to conduct the cross-cultural factorial validation of the Outcome Questionnaire (OQ-45.2) using the Polish population.

Material and method. Data were obtained from day-patients (n = 211), inpatients (n = 234), outpatients (n = 137) and non-patients (n = 426). Statistical analyses included:

Realizacja tego projektu była możliwa dzięki wparciu David M. Kennedy Center Research Grant. Autorzy dziękują osobom, które pomogły w zbieraniu danych: Sławomirze Bering, Piotrowi Domaradzkiemu, Małgorzacie Dudkiewicz, Pawłowi Dybce, Wioletcie Gajowej, Markowi Gajowemu, Pawłowi Jankowskiemu, Andrzejowi Karczewskiemu, Annie Kielbasińskiej, Janowi Kraszewskiemu, Aleksandrze Kulczyńskiej, Leszkowi Madeyskiemu, Andrzejowi Miziołkowi, Magdalenie Musioł, Agacie Rusak, Ewelinie Seroka, Piotrowi Szetelnickiemu, Januszowi Sztencelowi, Piotrowi Waszkiewiczowi, Joannie Wergiles, Andrzejowi Winklerowi, Andrzejowi Zembrzuskemu oraz Beacie Zimowskiej.

parallel analysis, exploratory factor analysis, confirmatory factor analysis, correlation analysis, criterion equivalence, clinical significance and reliable change index, and test-retest.

Results. Statistical analyses provided the strongest support for the bi-level model of the total score, five orthogonal (subscales specific for the Polish OQ, i.e. Social Conflicts and Addictive Behaviors; and original yet modified subscales, i.e. Symptom Distress, Interpersonal Relations, and Social Role), and two oblique factors (Somatization and Anxiety, Social Role 2). The psychometric properties of the Polish OQ were found to be adequate and similar to the original American OQ and its international adaptations. Specific for the Polish OQ cut-off scores for clinical significance were established. The role of cultural differences and the passage of time in the process of the cross-cultural validation were elaborated upon.

Conclusions. The Polish version of the OQ 45-2 has been recognized as an instrument adequately measuring general functioning as well as specific areas of functioning of the individual (i.e. interpersonal relations; social role performance; social conflicts; symptom distress; somatization and anxiety; addictive behaviors). Results of this factorial analysis seem to be valuable for both clinicians using the OQ-45.2 and for creators of any psychotherapy outcome measure.

Słowa kluczowe: skuteczność psychoterapii, kwestionariusz objawowy, kwestionariusz wyniku

Key words: psychotherapy effectiveness, symptom checklist, outcome questionnaire

Wstęp

Wysiłki zmierzające do coraz rzetelniejszej oceny skuteczności psychoterapii doprowadziły m.in. do skonstruowania Kwestionariusza Oceny Skuteczności Psychoterapii (Outcome Questionnaire – OQ-45.2) [1]. Narzędzie to uchodzi za czwartą najczęściej stosowaną na świecie miarę – wśród ankiet wypełnianych przez pacjentów – efektywności psychoterapii [2]. Zapewne decydują o tym jej następujące właściwości: (1) struktura podskal – Nasilenie Objawów (Symptom Distress – SD), Relacje Interpersonalne (Interpersonal Relations – IR), Rola Społeczna (Social Role – SR) – umożliwiającą wielowymiarowy pomiar poprawy lub pogorszenia stanu pacjenta, (2) trafność w odniesieniu do rozmaitych zaburzeń oraz problematycznych sytuacji życiowych, co pozwala na porównywanie różnorodnych grup respondentów, (3) czułość na zmianę i zdolność monitorowania postępów pacjenta dzięki dokonywaniu pomiarów z sesji na sesję, (4) zdolność predykcyjna służąca do identyfikacji osób, których stan może ulec pogorszeniu lub które mogą przedwcześnie przerwać terapię, oraz (5) system informacji zwrotnych w postaci graficznych sygnałów ostrzegawczych (signal-alarm feedback system) generowanych przez program OQ-Analyst, zapewniający terapeutom stały dostęp do informacji klinicznych i w ten sposób umożliwiający maksymalizację wyników terapii, np. interweniowanie zanim dojdzie do deterioracji pacjenta [3].

Badania walidacyjne oryginalnej wersji OQ przeprowadzono głównie w różnorodnych populacjach pacjentów ambulatoryjnych: (1) z kliniki uniwersyteckiej [4], (2) z oddziału psychiatrycznego szpitala ogólnego [5], (3) ze środowiskowego ośrodka leczenia uzależnień [6] oraz wśród pacjentów z całodobowego ośrodka leczenia zaburzeń odżywiania się [7]. Meta-/megaanaliza badań z zastosowaniem OQ [8] potwierdziła, że informacje zwrotne zawarte w OQ oraz w sygnałach ostrzegawczych mają istotny klinicznie i statystycznie wpływ na końcowy wynik terapii, w tym na zmniejszanie

odsetka pacjentów, u których prognozowano niepowodzenie terapii (rozumiane jako deterioracja ich stanu psychicznego), z wyjściowych 20% do około 5,5%. Adaptacji poddano również kilka wersji językowych OQ, np. holenderską [9], niemiecką [10], włoską [11] oraz szwedzką [12].

W Polsce do najczęściej stosowanych narzędzi oceny efektywności psychoterapii należy Kwestionariusz Objawowy (KO„O”) i jego różne wersje [13]. Przy ich użyciu przeprowadzono badania procesu psychoterapii [14], w tym ryzyka jej przerywania [15], a także badania osób cierpiących z powodu dysfunkcji seksualnych [16] i myśli samobójczych [17].

Cel

Celem projektu jest transkulturowa adaptacja polskiej wersji OQ, obejmująca zarówno ocenę równoważności polskiej struktury czynnikowej w porównaniu z oryginalną wersją amerykańską, jak i określenie psychometrycznych właściwości polskiego OQ. Spodziewamy się, że uzyskane wyniki poszerzą zrozumienie struktury czynnikowej OQ, co jest o tyle istotne, że wyniki dotychczasowych badań uzasadniających trójczynnikową strukturę OQ-45.2 są szeroko i krytycznie dyskutowane [9, 11, 12, 18–21]. Wartość niniejszego projektu wydaje się więc trudna do przecenienia.

Psychometryczne właściwości Kwestionariusza Oceny Skuteczności Psychoterapii

Kwestionariusz OQ-45.2 jest narzędziem składającym się z 45 pozycji (zakres wyników całkowitych (total score) od 0 do 180 punktów). W narzędziu tym zastosowano skalę o pięciu punktach odcięcia (0 = nigdy, 1 = rzadko, 2 = czasami, 3 = często, 4 = prawie zawsze),

W celu ograniczenia prawdopodobieństwa zakłóceń (tendencji) odpowiedzi 20% (9/45) itemów jest punktowana odwrotnie. Wyższe wyniki oznaczają większe nasilenie objawów oraz/lub gorsze funkcjonowanie.

Rzetelność OQ jest zadowalająca dla wszystkich podskal i wyniku całkowitego. Lambert i wsp. [22] określili (na podstawie współczynnika α Cronbacha) spójność wewnętrzną (internal consistency reliability) na poziomie 0,93 oraz rzetelność – dla wyniku całkowitego OQ-45 – według metody powtórnego pomiaru w odstępie trzech tygodni (test–retest reliability) na poziomie 0,84. Trafność diagnostyczna (concurrent validity) została ustalona poprzez skorelowanie wyniku całkowitego OQ-45 z wynikiem Kwestionariusza Objawowego Derogatisa (SCL-90) [23], Kwestionariusza Depresji Becka (BDI) [24] oraz Kwestionariusza Lęku i Lękliwości Spielbergera (STAI) [25]. Wszystkie wartości parametrów trafności krzyżowej OQ-45 z wymienionymi narzędziami były istotne na poziomie 0,01 z zakresem współczynników korelacji r od 0,50 do 0,85. Jako punkt odcięcia (cut-off point) różnicujący niekliniczną populację niepacjentów od klinicznej populacji pacjentów został zaproponowany wynik całkowity na poziomie 64 punktów, natomiast różnica 14 punktów została wskazana jako miara istotnej klinicznie zmiany (reliable change) [22].

Struktura czynnikowa

Na wynik całkowity OQ-45 składają się trzy ortogonalne podskale mające mierzyć oddzielne obszary funkcjonowania. Podskala Relacje Interpersonalne (11 pozycji) bada jakość relacji interpersonalnych m.in. w zakresie nasilenia konfliktów czy tendencji do izolacji. Podskala Rola Społeczna (9 pozycji) odnosi się do aktywności zawodowych i sposobu spędzania czasu wolnego. Podskala Nasilenie Objawów (25 pozycji) służy do oceny rozmaitych objawów, głównie depresyjnych i lękowych. Należy przy tym zaznaczyć, że struktura trójczynnika OQ nie wynika bynajmniej z eksploracyjnej analizy czynnikowej tego narzędzia. Przeciwnie, została ona ustalona raczej arbitralnie przez doświadczonych klinicystów. Po przeprowadzeniu starannego przeglądu piśmiennictwa wybrali oni pozycje do każdej podskali, a procedury statystyczne ograniczyli jedynie do analizy korelacji między nominowanymi pozycjami [1].

W efekcie nawet ci autorzy, którzy do pewnego stopnia potwierdzają adekwatność modelu trójczynnika, wskazują na to, że być może inne rozwiązania mogą pełniej odzwierciedlać wielowymiarowość OQ. Na przykład de Jong i wsp. [9], mimo uznania modelu trójczynnika za rozsądnie dopasowany do danych holenderskich, zaproponowali dodatkowe czynniki, takie jak np. podskala Lęku i Stresu Somatycznego (Anxiety and Somatic Distress – ASD), pozwalające oszacować wariancję nie wyjaśnianą przez trzy oryginalne czynniki. Podobnie autorzy szwedzkiej wersji OQ [12] zaproponowali model dziesięcioczynnikowy jako potencjalnie bardziej adekwatną jego strukturę.

W przypadkach gdy nie znajdowano wystarczającego uzasadnienia dla struktury trójczynnika, eksploracyjna analiza czynnikowa często wskazywała, oprócz podskal SR, SR i IR, na ukośny model czynnikowy (oblique factor model). Na przykład Bludworth i wsp. [18], choć stwierdzili, że model trójczynnika w pewnym stopniu pasuje do ich danych, to za strukturę bardziej adekwatną uznali czteroczynnikowy model dwupoziomowy (four factor bi-level model), z autonomicznym czynnikiem Stresu Ogólnego (General Distress). Analogicznie, Lo Coco i wsp. [11] zaproponowali dwupoziomowe rozwiązanie czteroczynnikowe z czynnikiem Całościowego Nieprzystosowania (Overall Maladjustment) reprezentującym funkcjonowanie badanego w wielu obszarach. Wyniki te odzwierciedlają wcześniejsze rezultaty uzyskane przez Muellera i wsp. [21], którzy wskazali na nakładanie się trzech oryginalnych podskal, co w ich opinii może istotnie zniekształcać wielowymiarową interpretację OQ-45.

W przypadku gdy dopasowanie trójczynnika było słabe, znacząco wzrastała liczba podskal ortogonalnych. Na przykład osiem niezależnych czynników zostało wskazanych przez Minami i wsp. [20]: Anhedonia (Anhedonia), Stres Psychologiczny (Psychological Distress), Stres Fizyczny (Physical Distress), Utrata Produktyności (Loss of Productivity), Brak Intymności (Lack of Intimacy), Problemacyjne Używanie Substancji (Problematic Substance Use), Konflikt Interpersonalny (Interpersonal Conflict – IC) oraz Stres (Stress).

Alternatywnie, gdy konfirmacyjna analiza czynnikowa nie potwierdzała, aby model trójczynnika w sposób zadowalający pasował do danych, badacze modyfikowali nie tylko strukturę OQ, ale także zmniejszali liczbę pozycji. Przykładowo – Kim i wsp.

[19] zmniejszyli liczbę pozycji do 26 i otrzymali model czteroczynnikowy zawierający skale: Lęk-Somatyzacja (Anxiety-Somatization), Negatywna Samoocena (Negative Self-Worth), Używanie Substancji (Substance Use) oraz Utrata Zainteresowań (Loss of Interest). Autorzy uznali ten model za lepszy od struktury trójczynnikowej, aczkolwiek nadal niewystarczająco dobrze dopasowany.

Podsumowując, wydaje się, że chociaż wynik całkowity OQ może być traktowany przez klinicystów jako wskaźnik służący identyfikacji pacjentów z ryzykiem pogorszenia, to niekonkluzywne wyniki analiz czynnikowych zmuszają terapeutów do ostrożnego stosowania podskal jako punktów odniesienia dla decyzji klinicznych. Prawdopodobnie dlatego właśnie graficzna prezentacja oprogramowania OQ-Analyst, choć uwzględnia wartości poszczególnych podskal OQ, opiera się wyłącznie na wyniku całkowitym OQ.

Metoda

Tłumaczenie

W pierwszej kolejności kwestionariusz OQ został przełożony z języka angielskiego na polski przez dwóch niezależnie od siebie pracujących tłumaczy. Rozbieżności między tymi tłumaczeniami były rozwiązywane w trakcie dyskusji. Następnie uzgodnioną wersję przetestowano pod kątem przejrzystości i adekwatności na grupie ośmiu certyfikowanych terapeutów oraz ich pacjentów. Potem, w celu utrzymania semantycznej równoważności, treść kwestionariusza została zwrrotnie przetłumaczona na język angielski przez trzeciego tłumacza. Dokładność tłumaczenia zwrótnego została potwierdzona przez autorów oryginalnego narzędzia. Na końcu kwestionariusz został ponownie przetestowany pod kątem zrozumiałości i rzetelności na kolejnej grupie siedmiu polskich klinicystów i ich pacjentów leczonych psychoterapią. Wszyscy trzej tłumacze byli doświadczonymi klinicystami i badaczami psychoterapii. Ich średni wiek wynosił 46,33 roku (odch. std. = 13,58, zakres 38–62), a średnia liczba lat ich pracy klinicznej od momentu uzyskania stosownych uprawnień wynosiła 20,33 roku (odch. std. = 8,26, zakres 14–32). Wszyscy tłumacze mieli tytuł doktora nauk, dwóch było specjalistami psychologii klinicznej, a jeden psychiatr. Wszyscy trzej praktykowali psychoterapię zarówno w warunkach ambulatoryjnych, jak i stacjonarnych. Jeden z nich opisał swoją orientację teoretyczną jako poznawczo-behawioralną, drugi jako psychodynamiczną, a trzeci jako integratywną.

Uczestnicy

Grupa badana składała się z trzech podgrup klinicznych i jednej nieklinicznej (tab. 1). Grupa kliniczna składała się z 211 pacjentów oddziału dziennego z ośrodka psychoterapeutycznego na południu Polski, 234 pacjentów całodobowego ośrodka psychoterapeutycznego opartego na lecznictwie szpitalnym z Polski centralnej oraz 137 pacjentów psychoterapii ambulatoryjnej z dziesięciu praktyk prywatnych z całego kraju. Wszystkie te podgrupy składały się z pacjentów kolejno zgłaszających się na leczenie.

Ze względu na kryteria wykluczenia (poważne myśli samobójcze, psychoza, ciężka depresja, aktywne nadużywanie substancji psychoaktywnych) niektórzy respondenci (37/619; 5,98%) zostali wyłączeni z badania, ponieważ ich leczenie było w znacznym stopniu oparte na farmakoterapii, a psychoterapia odgrywała rolę marginalną. Wszyscy badani wypełnili po raz pierwszy OQ przed rozpoczęciem pierwszej sesji psychoterapii w ramach procedury przyjęcia. Następne kwestionariusze wypełniano piętnaście minut przed każdą z kolejnych sesji przez cały okres trwania leczenia. We wszystkich podgrupach kwestionariusze były rozdawane w odstępach jednodniowych. Przeciętna liczba kwestionariuszy wypełnionych przez jedną osobę wynosiła 10,45 (odch. std. = 4,19) dla pacjentów oddziału dziennego, 11,31 (odch. std. = 2,04) dla pacjentów oddziału całodobowego oraz 13,84 (odch. std. = 7,16) dla pacjentów ambulatoryjnych.

Danych dostarczyło osiemnastu psychologów, dziewięciu psychiatrów oraz dziesięciu psychoterapeutów w trakcie szkolenia. Przeciętny wiek terapeutów we wszystkich trzech podgrupach wynosił 39,35 roku (odch. std. = 7,59, zakres 28–52), a średnia liczba lat praktykowania po uzyskaniu specjalizacji wynosiła 11,37 roku (odch. std. = 8,50, zakres 3–26). Pacjentów przypisywano do klinicystów, kierując się dostępnością terapeutów oraz czynnikami klinicznymi. Większość (21/37, 56,76%) z uczestniczących terapeutów opisała swoją orientację teoretyczną jako integratywną, podczas gdy pozostali klinicyści identyfikowali się z podejściem psychodynamicznym (ośmioro), poznawczo-behawioralnym (sześcioro) oraz systemowym (dwóch). Nie przeprowadzono badania spójności stosowanego leczenia. Terapeuci nie byli propagatorami stosowanych kwestionariuszy i nie byli wynagradzani za ich stosowanie.

Podgrupa niekliniczna składała się z 426 studentów z siedmiu polskich uniwersytetów (nie byli oni studentami psychologii, pedagogiki, nauk społecznych ani medycyny). Osoby z podgrupy klinicznej były uznawane za niepacjentów, o ile w czasie zbierania danych nie uczestniczyły w żadnej formie psychoterapii ani farmakoterapii. Kwestionariusze były rozdawane przez asystentów dydaktycznych co tydzień przez cztery kolejne tygodnie. Respondenci (57/483, 11,8%), którzy wypełnili OQ mniej niż trzy razy, nie zostali włączeni do analiz. Uczestnictwo było dobrowolne i anonimowe. Przeciętny niepacjent wypełnił 3,98 (odch. std. = 0,1) kwestionariusza. Osobom z podgrupy nieklinicznej nie oferowano żadnego wynagrodzenia ani rekompensaty.

Tabela 1. Grupa badana

Grupa	n	Płeć		Wiek	
		Kobiety n (%)	Mężczyźni n (%)	Zakres	Średnia (odch. std.)
Pacjenci oddziału dziennego	211	143 (67,8)	68 (32,2)	19–57	31 (8,49)
Pacjenci oddziału całodobowego	234	170 (72,6)	64 (27,4)	18–57	31 (8,71)
Pacjenci ambulatoryjni	137	86 (62,8)	51 (37,2)	20–61	32 (7,82)
Pacjenci – suma	582	399 (68,6)	183 (31,4)	18–61	31 (8,44)
Niepacjenci	426	231 (54,2)	195 (45,8)	18–59	23 (6,79)

dalszy ciąg tabeli na następnej stronie

Grupa	Płeć			Wiek	
	n	Kobiety n (%)	Mężczyźni n (%)	Zakres	Średnia (odch. std.)
Cała grupa (pacjenci i niepacjenci)	1 008	630 (62,5)	378 (37,5)	18–61	28 (8,66)

odch. std. – odchylenie standardowe

Analiza danych

W pierwszym podejściu przetestowano trafność oryginalnego trójczynnиковego modelu na polskich danych przy założeniu ortogonalności podskal. Ze względu na niski poziom wariancji wyjaśnianej przez model trójczynnиковy oraz słabe wartości indeksów confirmacyjnej analizy czynnikowej, w drugim podejściu zdecydowano się na poszukiwanie alternatywnych układów podskal struktury OQ-45.

Analizę danych oparto na: 1) metodzie równoległości (parallel analysis – PA) ustalającej liczbę czynników utożsamianych z podskalami OQ-45, 2) eksploracyjnej analizie czynnikowej (exploratory factor analysis – EFA) umożliwiającej określenie struktury każdej z podskal OQ, 3) confirmacyjnej analizie czynnikowej (confirmatory factor analysis – CFA) testującej, na podstawie wartości mierników, dopasowanie sugerowanej przez PA i EFA struktury dla polskich danych, 4) analizie współczynników korelacji (correlation analysis) określającej stopień współzależności pomiędzy podskalami, 5) kryterium równoważności (criterion equivalence) umożliwiającym rozróżnienie próby klinicznej od nieklinicznej, 6) klinicznej istotności (clinical significance) pozwalającej na ustalenie punktu odcięcia (cut-off point) oraz indeksu rzetelnej zmiany (reliable change index – RCI) oraz 7) metodzie powtórnego testowania (test-retest) w celu oszacowania miernika rzetelności OQ.

Metoda równoległa jako symulacyjna metoda bazująca na modelu Monte Carlo porównuje wartości własne wyznaczone na podstawie danych rzeczywistych z wartościami własnymi uzyskanymi na podstawie danych losowych. Dany czynnik jest uwzględniany, jeżeli związana z nim wartość własna jest większa niż 95 percentyl wartości własnej otrzymanej z danych losowych. Liczbę czynników określa punkt, w którym wartości własne wyznaczone dla rzeczywistych danych po raz ostatni będą większe od wartości własnych wyznaczonych na podstawie danych losowych [26].

Analizy czynnikowej użyto w dwóch celach: potwierdzenia liczby czynników z oryginalnego narzędzia oraz przyporządkowania pytań kwestionariusza do odpowiednich czynników. Badanie zgodności czynnikowej przeprowadzono na podstawie eksploracyjnej analizy czynnikowej (EFA z rotacją czynników: varimax, promax oraz oblimin), bazując na danych pochodzących z pierwszego i ostatniego kwestionariusza badanej grupy pacjentów, a uzyskane wyniki potwierdzono za pomocą confirmacyjnej analizy czynnikowej (CFA), szacując parametry modelu za pomocą metody największej wiarygodności (maximum likelihood – ML) oraz uogólnionej metody najmniejszych kwadratów (generalized least square – GLS) dla pierwszych czterech sesji (ze względu

na posiadaną liczbę kwestionariuszy dla grupy kontrolnej, uwzględniono wyniki tylko z czterech sesji). Zdecydowano się na zastosowanie zarówno metody ML, jak i GLS, aby móc porównywać uzyskane wyniki z rezultatami innych badaczy kwestionariusza OQ. Z metody ML skorzystali bowiem np. Bludworth i wsp. [18], natomiast z metody GLS np. de Jong i wsp. [9] czy Mueller i wsp. [21]. Ponadto w niniejszej pracy zastosowano wersję ML odporną (robust ML) na niespełnienie założeń wielowymiarowego rozkładu normalnego, błędy standardowe i statystyki testowe. Dodatkowym argumentem przemawiającym za zastosowaniem metody ML była duża liczba obserwacji.

W przypadku CFA zastosowano następujące miary dopasowania modelu do danych (goodness-of-fit criteria): χ^2 (the chi-square – χ^2), χ^2 dzielony przez liczbę stopni swobody (the chi-square divided by degrees of freedom – χ^2/df), pierwiastek średniokwadratowy reszt (the root mean square residual – RMR), pierwiastek średniokwadratowego błędu aproksymacji (the root mean square error of approximation – RMSEA), indeks dopasowania modelu do danych (the goodness-of-fit index – GFI), skorygowany indeks dopasowania modelu do danych (the adjusted goodness-of-fit index – AGFI), unormowany indeks dopasowania (the normed fit index – NFI), porównawczy indeks dopasowania (the comparative fit index – CFI), indeks Tuckera-Lewisa (the Tucker-Lewis index – TLI), krytyczne N (the critical N – CN), kryterium informacyjne Akaike (the Akaike information criterion – AIC) oraz bayesowskie kryterium informacyjne Schwarza (the Bayesian information criterion – BIC). Zaprezentowano wszystkie wspomniane powyżej miary, aby uzyskane wyniki porównać z wynikami innych autorów. Jakość dopasowania oceniono na podstawie następujących wartości granicznych: χ^2 – miara bez wartości granicznych; $\chi^2/df < 3$; RMR < 0,05; RMSEA < 0,07; GFI, AGFI, CFI, NFI oraz TLI > 0,9; CN > 200; AIC oraz BIC wskazujące najniższe wartości w przypadku porównywanych modeli [27].

W przypadku EFA liczbę czynników wyekstrahowano, korzystając z liczby czynników sugerowanych przez: PA, wartości własne (> 1), kryterium osypiska, polychoryczną macierz korelacji, jak również teorii psychologicznych i argumentów klinicznych. Pytania, których wartości ładunku czynnikowego co do modułu były ≥ 30 , pozostały w oryginalnych podskalach. W konsekwencji trzy oryginalne podskale zostały skrócone, natomiast pozostałe pytania uformowały nowe podskale. Mierniki jakości CFA potwierdziły adekwatność wprowadzonych modyfikacji.

W kolejnym kroku przeanalizowano, zgodnie z sugestią z literatury [28], model typu dwupoziomowego, rozważając kilka alternatywnych dwupoziomowych struktur kwestionariusza OQ-45. Następnie proponowane struktury zostały przeanalizowane przez ekspertów klinicystów. Na podstawie powyższej procedury uzyskano dwa modele typu dwupoziomowego, które według autorów stanowią najwłaściwsze dopasowanie do polskich danych: model dwupoziomowy z pięcioma ortogonalnymi podskalami zwany „pięcoczynnikowym modelem typu dwupoziomowego” oraz model dwupoziomowy z pięcioma podskalami ortogonalnymi oraz dwiema dodatkowymi podskalami nieortogonalnymi (typu ukośnego) zwany dalej „siedmioczynnikowym modelem typu dwupoziomowego”. Wskaźniki dopasowania modelu do danych faworyzowały model siedmioczynnikowy względem modelu pięcoczynnikowego, zatem do dalszych analiz wybrano wyłącznie ten pierwszy model.

Wyznaczając współczynniki korelacji pomiędzy podskalami, przyjęto, że wartości pomiędzy 0,31 a 0,6 wskazują na umiarkowaną korelację między nimi, natomiast wartości powyżej 0,6 dla podskal oraz wyniku całościowego OQ wskazują na silną korelację. Kryterium równoważności estymowano jako różnicę w wynikach dla grupy klinicznej i nieklinicznej. Zastosowano test λ Wilksa.

Wartości krytyczne do szacowania istotności klinicznej, punktu odcięcia dla wyników oraz RCI wyznaczono, korzystając z kryteriów podanych przez Jacobsona i Truax [29] i zmodyfikowanych przez Lamberta, Hansena i Bauer [30]. Rzetelność oszacowano na podstawie wewnętrznej zgodności (α Cronbacha) oraz metody powtórnego testowania (współczynnik korelacji Pearsona). Współczynnik powyżej 0,7 uznano jako satysfakcjonujący miernik rzetelności. Odległość czasowa w ramach metody test-retest wyniosła tydzień. Czułość na zmianę (sensitivity to change) zmierzono za pomocą statystyk: testu t-Studenta dla powtarzanych pomiarów oraz wielkości efektu (effect size). Zastosowano standaryzowaną wielkość efektu d , którą zważono rozmiarem próby. Klasyfikację miernika d przeprowadzono na podstawie gradacji Cohena: $> 0,79$ – duży, $0,5-0,79$ – średni, $< 0,5$ – mały [31].

EFA przeprowadzono za pomocą pakietów: SPSS, Statistica oraz R, CFA za pomocą: AMOS 21 oraz R 3.0, natomiast metodę powtórnego testowania i wewnętrzną zgodność obliczono przy użyciu Statistica oraz SPSS. Brakujące wartości zastąpiono z wykorzystaniem procedury Monte Carlo [32]. Braki danych względem pytania stanowiły średnio 0,93%.

Wyniki

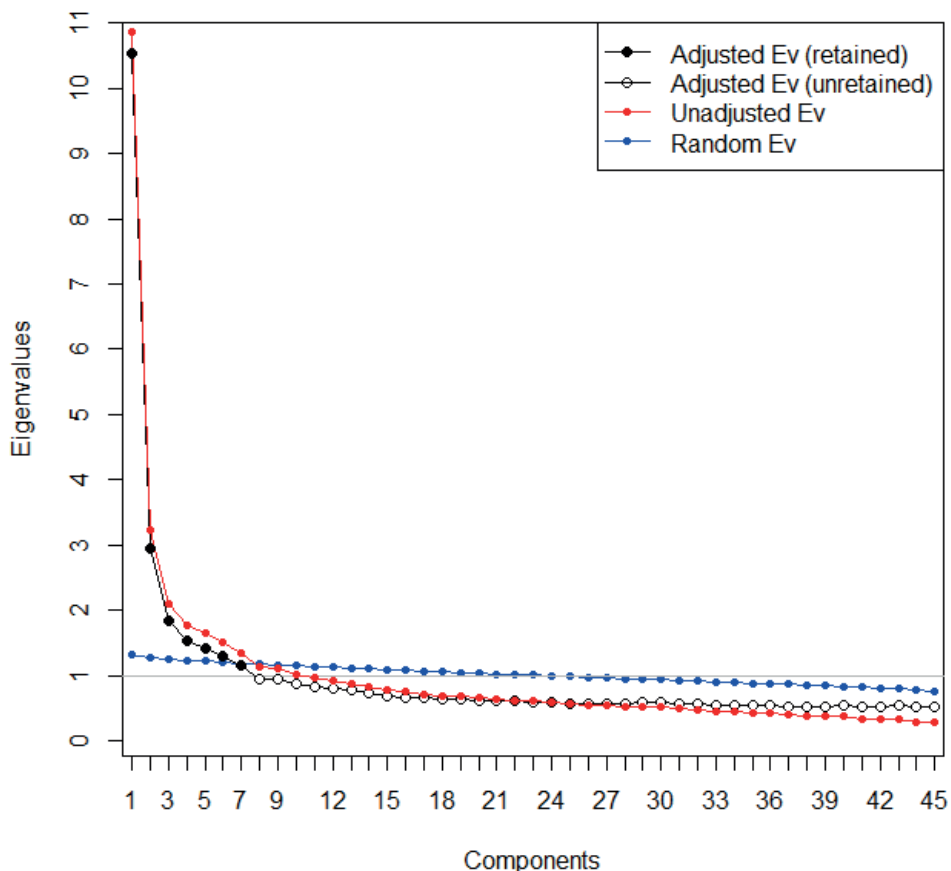
Metoda równoległa

Wyniki metody równoległej wskazują na model złożony z siedmiu podskal jako najbardziej odpowiednią strukturę polskiego OQ (potwierdzone względem danych z każdej sesji dla wszystkich trzech niezależnych podgrup pacjentów, jak również połączonej grupy wszystkich pacjentów). Uwzględniono czynniki korespondujące z wartościami własnymi większymi niż jeden. Z powodu ograniczonego miejsca zaprezentowano wyniki wyłącznie dla całej grupy badanej (rys. 1). Pierwszych siedem rzeczywistych wartości własnych jest zilustrowanych linią ciągłą, natomiast siedem losowych wartości własnych linią kropkowaną.

Analiza czynnikowa

Niekonkluzywne wyniki poprzednich badań, jak również kulturowe różnice między populacją amerykańską a polską [33] spowodowały, że stało się prawdopodobne, iż struktura czynnikowa polskiego OQ może być odmienna od oryginalnego amerykańskiego modelu. Przy użyciu eksploracyjnej analizy czynnikowej wyznaczono więc strukturę czynnikową OQ bardziej adekwatnie dopasowaną do polskich danych. Natomiast stosując konfirmacyjną analizę czynnikową, uwzględniono zarówno procedury empiryczne (tzn. modyfikacje struktury pozycji), jak

Parallel Analysis



Rysunek 1. **Metoda równoległa: wartości własne danych rzeczywistych i losowych dla zachowanych czynników (N = 1 008)**

Components – nr czynników, Eigenvalues – wartości własne, Adjusted Ev (retained) – skorygowane wartości własne (uwzględnione przy wyborze liczby czynników), Adjusted Ev (unretained) – skorygowane wartości własne (nieuwzględnione przy wyborze liczby czynników), Unadjusted Ev – nieskorygowane wartości własne, Random Ev – wartości własne na podstawie „losowych danych”

i racjonalne (tzn. adekwatność względem teorii psychologicznych oraz praktyki klinicznej) [21].

Model trójczynnikowy zaproponowany przez twórców OQ na polskich danych nie został pozytywnie zweryfikowany przez EFA – uzyskano jedynie częściowe dopasowanie pozycji między wersjami amerykańską i polską. Ponadto słabo spełniał kryteria dopasowania dla CFA. Na podstawie wartości ładunków czynnikowych polskich

danych niektóre pozycje modelu trójczynnikowego pozostały w ramach oryginalnych ortogonalnych podskal pierwszego rzędu (tab. 2), zaś inne przeniesiono z jednej oryginalnej skali do drugiej, podczas gdy pozostałe pozycje bardziej pasowały do czynników ortogonalnych drugiego rzędu. W efekcie oryginalne podskale zmodyfikowano, a ponieważ wciąż wydawały się mierzyć pierwotne fenomeny, utrzymano ich nazwy.

Tabela 2. Standaryzowane ładunki czynnikowe siedmioczynnikowego modelu typu dwupoziomowego (N = 1 008)

Pozycja	IR	SD	SR	SC	AB	SA	SR2	TOT
OQ1 ^a	0,62					0,44 ^x		0,56
OQ7 ^a	0,36							0,76
OQ12 ^c	0,48		0,39 ^x				0,4	0,6
OQ13 ^b	0,88							0,22
OQ15 ^b	0,65	0,33 ^x					0,54	0,46
OQ17 ^a	0,42							0,77
OQ18 ^a	0,64					0,48 ^x		0,52
OQ20 ^a	0,83							0,29
OQ21 ^c	0,72						0,55	0,45
OQ24 ^b	0,75					0,57 ^x		0,43
OQ30 ^a	0,5			0,43 ^x		0,41 ^x		0,49
OQ31 ^b	0,86							0,25
OQ37 ^a	0,76							0,37
OQ43 ^a	0,79							0,36
OQ2 ^b		0,52	0,36 ^x			0,42		0,58
OQ5 ^b	0,44 ^x	0,47				0,42 ^x		0,58
OQ6 ^b	0,29 ^x	0,3		0,35 ^x		0,31 ^x		0,69
OQ8 ^b	0,29 ^x	0,29						0,74
OQ9 ^b	0,39 ^x	0,46	0,32 ^x			0,49		0,51
OQ10 ^b		0,6				0,47		0,53
OQ16 ^a		0,45					0,3	0,7
OQ23 ^b		0,33						0,56
OQ25 ^b		0,48				0,35		0,65
OQ27 ^b		0,69				0,51		0,49
OQ29 ^b		0,67				0,46		0,54
OQ33 ^b		0,66				0,46		0,54
OQ34 ^b		0,69				0,52		0,48
OQ35 ^b		0,42				0,2		0,8

dalszy ciąg tabeli na następnej stronie

OQ36 ^b		0,58			0,44		0,56
OQ40 ^b	0,34 ^x	0,44				0,37 ^x	0,63
OQ41 ^b		0,62			0,44		0,56
OQ42 ^b		0,44					0,51
OQ45 ^b		0,7			0,52		0,48
OQ3 ^b	0,38 ^x		0,4			0,37	0,63
OQ4 ^c			0,51			0,39	0,61
OQ22 ^b		0,36 ^x	0,58		0,47 ^x		0,53
OQ28 ^c			0,67			0,48	0,52
OQ38 ^c			0,71				0,44
OQ14 ^c				0,35		0,21	0,79
OQ19 ^a				0,55		0,32	0,68
OQ39 ^c			0,44 ^x	0,58			0,44
OQ44 ^c				0,57		0,38 ^x	0,62
OQ11 ^b					0,56	0,35 ^x	0,65
OQ26 ^a					0,65	0,43 ^x	0,57
OQ32 ^c					0,6	0,37 ^x	0,63

Test sferyczności Bartletta: $\chi^2 = 23\ 887$; $df = 990$; $p < 0,0001$; miara KMO: 0,967.

IR – podskala Relacje Interpersonalne, SD – podskala Nasilenie Objawów, SR – podskala Rola Społeczna, SC – podskala Konflikty Społeczne, AB – podskala Zachowania Uzależnieniowe, SA – podskala Somatyzacja i Lęk, SR2 – podskala Rola Społeczna 2, TOT – wynik całościowy, a – oryginalnie część podskali IR amerykańskiego OQ-45.2, b – oryginalnie część podskali SD amerykańskiego OQ-45.2, c – oryginalnie część podskali SR amerykańskiego OQ-45.2, x – ładunki czynnikowe większe od 0,3, ale mniejsze od największego ładunku czynnikowego dla danej pozycji

Czynnik Nasilenie Objawów zawiera 18 z 25 pozycji oryginalnej podskali SD oraz jedną pozycję oryginalnie umieszczoną w podskali IR (tab. 2). W ten sposób na nowo zdefiniowana struktura podskali SD blisko przypomina podskalę Stresu Psychologicznego z badań Minami i wsp. [20], która jednak nie zawiera pozycji OQ16 („Niepokoję się kłopotami rodzinnymi”), znajdującej się w polskiej podskali SD. Zdaniem autorów niniejszego artykułu ta pozycja adekwatniej odzwierciedla naturę podskali SD niż IR, gdyż odnosi się raczej do niepokoju rozumianego jako poznawcze i emocjonalne zjawisko, które potencjalnie może ulec zaostrzeniu do poziomu stresu objawowego, niż mierzy problematyczną dynamikę systemu rodzinnego. Ponadto Polacy, z racji tego, że przeciętnie bardziej niż Amerykanie mają skłonność do wyrażania swojego niezadowolenia i do pograżania się w obawach [33], mogą być bardziej skłonni do odczytywania OQ16 raczej z perspektywy objawowej niż interpersonalnej. Pozycję OQ16 można więc uznać za adekwatny komponent podskali mierzącej objawy psychologiczne i fizyczne.

Czternaście pozycji wyraźnie ukonstytuowało czynnik Relacje Interpersonalne, który mierzy różne wymiary funkcjonowania interpersonalnego. Osiem pozycji polskiej podskali IR jest równoważnych z wersją amerykańską, dwie pozycje oryginalnie znajdowały się w podskali SR, a cztery w podskali SD (tab. 2). Obecność w zmodyfikowanej podskali IR dwóch pozycji poprzednio należących do SR (OQ12 „Mam dużo satysfakcji z pracy/nauki” oraz OQ21 „Z radością korzystam z wolnego czasu”) można tłumaczyć tym, że stopień satysfakcji z relacji interpersonalnych ludzie często postrzegają przez pryzmat jakości pracy zawodowej oraz wypoczynku [34].

Analogicznie, pozycje oryginalnie należące do podskali SD (np. OQ13 „Jestem szczęśliwą osobą”, OQ15 „Czuję się bezwartościowy”) nie tylko wydają się wprost powiązane z poziomem objawów, ale także z faktem, że jakość funkcjonowania interpersonalnego zależy od samooceny i ogólnego zadowolenia z życia [35]. Istotne jest to, że struktura pozycji polskiej wersji IR bardzo blisko przypomina podskalę Negatywnej Samooceny (Negative Self-Worth) według Kima i wsp. [19].

Kolejny czynnik składa się z pięciu pozycji odnoszących się do funkcjonowania w roli społecznej i dlatego nazwano go identycznie jak w amerykańskim OQ: Rola Społeczna. Zawiera on trzy oryginalne pozycje SR (tab. 2) oraz dwie pozycje z oryginalnej podskali SD (OQ3 „Niczym się nie interesuję” i OQ22 „Mam trudności z koncentracją”). Dwie ostatnie pozycje uznano za adekwatne dla podskali SR z racji tego, że nawiązują do trudności zwykle zakłócających funkcjonowanie społeczne [36]. Warto przy tym zauważyć, że struktura pozycji polskiej podskali SR w znacznym stopniu odzwierciedla strukturę czynnika Utrata Zainteresowań według Kima i wsp. [19].

Pozostałe pozycje to dwa czynniki ortogonalne, które nie zostały opracowane przez twórców OQ. Jeden z nich nazwano Zachowania Uzależnieniowe (Addictive Behaviors – AB), ponieważ wydaje się adekwatnie opisywać cechy uzależnionej osoby. Składa się z trzech pozycji: OQ11 „Po wypiciu dużej ilości alkoholu, następnego ranka muszę się napić, by dalej funkcjonować” (pozycja oryginalnie znajdująca się w czynniku SD), OQ26 „Złości mnie, gdy ludzie krytykują moje picie (lub branie narkotyków)” (oryginalnie w czynniku IR) oraz OQ32 „Mam problemy w pracy/szkole z powodu picia lub brania narkotyków” (oryginalnie w czynniku SR (tab. 2)). To, że amerykańska wersja kwestionariusza nie zawiera oddzielnej podskali AB, może wynikać z faktu, że OQ normalizowano w stanie Utah, uznawanym za stan o największym odsetku abstynentów w USA. Warto zauważyć, że pod względem zawartości polska podskala AB idealnie odzwierciedla podskalę ekstrapolowane przez innych badaczy i uznane przez nich za adekwatnie mierzące nadużywanie substancji [19, 20]. Można zatem założyć, że czynnik AB nie jest specyficzny tylko dla polskiego społeczeństwa, ale także dla tych populacji, gdzie picie alkoholu jest bardziej akceptowalne społecznie niż w Utah.

Kolejny nowy ortogonalny czynnik utworzono z czterech pozycji mierzących konflikt w rozmaitych sytuacjach społecznych i dlatego nazwano go Konflikt Społeczny (Social Conflict – SC). Trzy jego pozycje znajdowały się w oryginalnej wersji podskali SR: OQ14, „Za dużo pracuję/studiuję”, OQ39 „Mam zbyt wiele nieporozumień w pracy/szkole” oraz OQ44 „W pracy/szkole bywam tak rozszoszczony, że mogę zrobić coś, czego będę później żałować”, podczas gdy jedna pozycja oryginalnie należała do podskali IR: OQ19 „Często się kłóć” (tab. 2). Struktura pozycji SC mocno

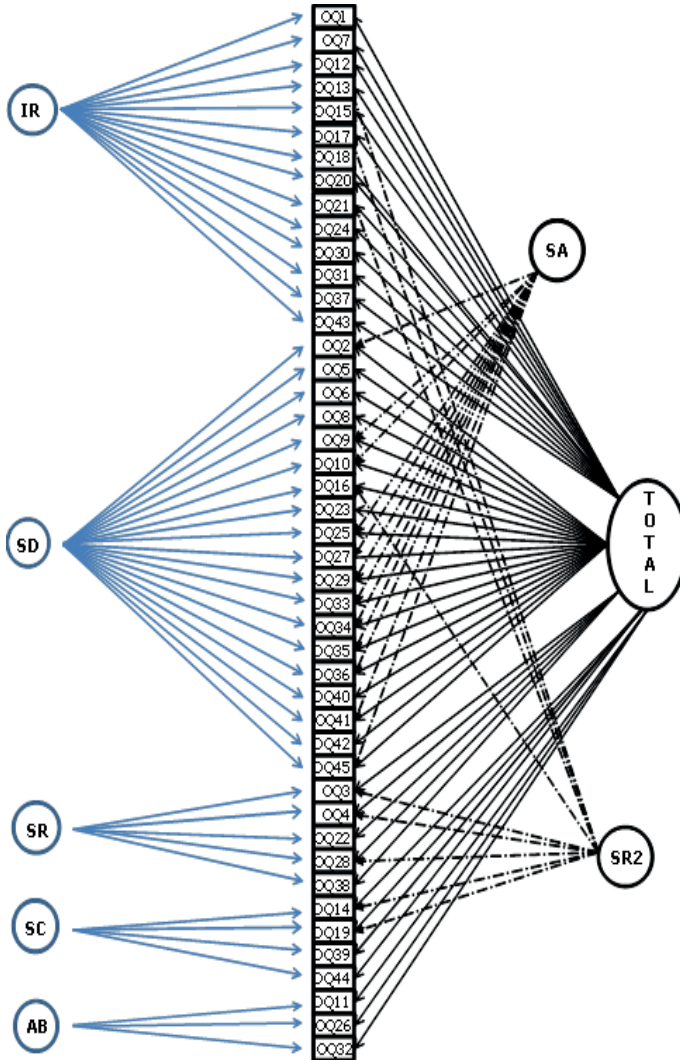
przypomina czynnik Konflikt Interpersonalny zdefiniowany przez Minami i wsp. [20], z tym, że w podskali IC zabrakło pozycji OQ14. Można zresztą dyskutować, do jakiego stopnia obecność tej właśnie pozycji w SC jest adekwatna. Wątpliwości budzi bowiem nie tylko względnie niska wartość jego ładunku czynnikowego, ale także jego treść. Z tych właśnie powodów wielu badaczy uznało OQ14 za na tyle problematyczny, że modyfikowali jego treść lub sugerowali jego usunięcie z kwestionariusza OQ [9, 12, 19–21]. Wydaje się, że oddzielna podskala mierząca konflikty społeczne niezależnie od czynników SR oraz IR mogłaby mieć istotne znaczenie kliniczne, szczególnie dla społeczeństw takich jak polskie, które pozwalają na otwarte wyrażanie frustracji i akceptują angażowanie się w konflikty społeczne [33].

Doszedłszy do wniosku, że pewne pozycje mogłyby być adekwatnie umieszczone w więcej niż jednym czynniku, zdecydowano się sprawdzić, czy model z czynnikami ukośnymi nie pasowałby lepiej do polskich danych. Użyto zatem macierzy korelacji, której wysokie wartości (wszystkie $> 0,7$) sugerowały obecność dodatkowych nieortogonalnych czynników. Następnie przypisano każdą pozycję do jednego z już wyodrębnionych czynników (IR, SD, SR, SC, AB), po czym przy użyciu rotacji ukośnej bazującej na metodzie promax (oblique rotation based on promax) oraz bezpośredniej rotacji ukośnej (direct oblimin method) powiązано poszczególne zmienne z potencjalnymi czynnikami ukośnymi. Gdy nowe czynniki ukośne zostały już statystycznie uformowane, ich zawartość została przeanalizowana zarówno z perspektywy teoretycznej, jak i klinicznej.

Pierwszy czynnik ukośny został nazwany Somatyzacja i Lęk (Somatization and Anxiety – SA), ponieważ tworzy go dwanaście pozycji mierzących objawy specyficzne dla zaburzeń lękowych i somatyzacyjnych, np. OQ10 „Czuję lęk” czy też OQ41 „Zasypiam z trudem lub źle śpię” (tab. 2). Wprawdzie wszystkie pozycje SA znajdują się także w podskali SD, jednak uznano ich wyodrębnienie za klinicznie istotne szczególnie dla tych respondentów, którzy mają rozpoznania zaburzeń nerwicowych czy lękowych. Z uwagi na to, że stanowią oni większość pacjentów psychoterapii, przynajmniej ambulatoryjnej, twierdzimy, że podskala SA może stanowić użyteczny dodatek do struktury pięcioczynnikowej kwestionariusza OQ. Fakt, że czynnik SA idealnie odzwierciedla czynnik Anxiety and Somatic Distress zaproponowany przez de Jong i wsp. [9] oraz jest bardzo zbliżony do czynnika Anxiety-Somatization opisanego przez Kim i wsp. [19], sugeruje, że może on być trafny nie tylko dla polskich respondentów.

Kolejny ukośny czynnik nazwano Rola Społeczna 2 (Social Role 2 – SR2), ponieważ utworzono go z dziewięciu pozycji mierzących pewne specyficzne aspekty pełnienia roli społecznej. Spośród pozycji znajdujących się także w czynniku SR w SR2 znalazła się np. pozycja OQ28 „Nie pracuję/nie uczę się tak dobrze jak kiedyś”; z puli pozycji obecnych także w czynniku IR, w SR2 ulokowano np. OQ21 „Z radością korzystam z wolnego czasu”; spośród pozycji pochodzących z czynnika SC w SR2 znalazła się np. pozycja OQ14, natomiast spośród pozycji obecnych w podskali SD w SR2 znalazło się stwierdzenie OQ16 (tab. 2). Pomimo tego, że w przypadku większości wymienionych pozycji wartość ich ładunków czynnikowych jest wystarczająca, to kliniczne zastosowanie SR2 wydaje się ograniczone do tych pacjentów, którzy są uwikłani w niemal dysfunkcjonalne role społeczne.

W następnym kroku przeprowadzono confirmacyjną analizę czynnikową w celu oceny: a) oryginalnego modelu trzech czynników ortogonalnych, b) pięcioczynnikowego modelu dwupoziomowego oraz c) siedmioczynnikowego modelu dwupozi-



Rysunek 2. Model siedmioczynnikowy, dwupoziomowy dla wyniku całkowitego, pięciu podskal ortogonalnych i dwóch ukośnych

IR – podskala Relacje Interpersonalne, SD – podskala Nasilenie Objawów, SR – podskala Rola Społeczna, SC – podskala Konflikty Społeczne, AB – podskala Zachowania Uzależnieniowe, SA – podskala Somatyzacja i Lęk, SR2 – podskala Rola Społeczna 2

N = 1 008

Tabela 3. Współczynniki dopasowania CFA (N = 1008)

Model	Estymacja	χ^2	df	χ^2/df	Współczynniki dopasowania									
					RMR	RMSEA	GFI	AGFI	NFI	CFI	TLI	CN	AIC	BIC
Oryginalny model trójczynnikowy	GLS	3 125 ^a	942	3,32	0,087	0,049*	0,856	0,842	0,15	0,187	0,146	314*	3 311	3 764
Oryginalny model trójczynnikowy	ML odporne	6 210 ^a	942	6,59	0,064	0,076	0,712	0,683	0,747	0,776	0,765	158	6 396	6 849
Dwupoziomowy model pięcioczynnikowy (z wynikiem całkowitym)	GLS	2 670 ^a	890	2,66*	0,056	0,041*	0,891**	0,873	0,356	0,449	0,361	392*	2 659	3 366
Dwupoziomowy model pięcioczynnikowy (z wynikiem całkowitym)	ML odporne	3 479 ^a	890	3,76	0,041*	0,053*	0,856	0,833	0,863	0,896**	0,884	278*	3 638	4 344
Dwupoziomowy model siedmioczynnikowy z pięcioma ortogonalnymi (z wynikiem całkowitym) i dwoma czynnikami ukośnymi	GLS	2 257 ^a	865	2,6*	0,054**	0,041*	0,9**	0,877	0,386	0,482	0,407	280*	2 597 ^b	3 326 ^b
Dwupoziomowy model siedmioczynnikowy z pięcioma ortogonalnymi (z wynikiem całkowitym) i dwoma czynnikami ukośnymi	ML odporne	3 082 ^a	865	3,56	0,038*	0,05*	0,869	0,843	0,874	0,91*	0,883	294*	3 418	4 237

* Spełnia rekomendowane kryteria (poziom zadowalający); ** Zbliża się do rekomendowanych kryteriów (poziom akceptowalny); ^a p < 0,001; ^b wskazanie najlepszego dopasowania spośród wszystkich testowanych modeli.

GLS – uogólniona metoda najmniejszych kwadratów, ML – metody największej wiarygodności, χ^2 – chi-kwadrat, df – liczba stopni swobody, RMR – pierwiastek średniokwadratowy reszt, RMSEA – pierwiastek średniokwadratowego błędu aproksymacji, GFI – indeks dopasowania modelu do danych, AGFI – skorygowany indeks dopasowania modelu do danych, NFI – unormowany indeks dopasowania, CFI – porównawczy indeks dopasowania, TLI – indeks Tucker-Lewisa, CN – krytyczne N, AIC – kryterium informacyjne Akaike, BIC – bayesowskie kryterium informacyjne Schwarz

mowego (tab. 3). Model trójczynnikiowy testowany na polskich danych słabo spełniał kryteria dopasowania. Nawet przy zastosowaniu metody opartej na GLS, tylko dwa z dziewięciu współczynników (RMSEA, CN) spełniały kryteria dopasowania na zadowalającym poziomie. Ponadto względnie wysokie wartości AIC oraz BIC sugerowały niewystarczające dopasowanie danych do tego modelu. W przypadku pięcioczynnikowego modelu dwupoziomowego (uwzględniającego wynik całkowity oraz pięć podskal ortogonalnych), po przetestowaniu go za pomocą CFA przy użyciu GLS, trzy współczynniki spełniały kryteria na zadowalającym poziomie (c^2/df , RMSEA, CN), a jeden współczynnik (GFI) na poziomie zbliżonym do zadowalającego. Współczynniki AIC oraz BIC sugerowały trafniejsze dopasowanie tego rozwiązania niż struktury trójczynnikiowej.

Siedmioczynnikowa struktura dwupoziomowa (dla wyniku całkowitego, pięciu ortogonalnych i dwóch ukośnych podskal), po przetestowaniu za pomocą kryteriów CFA opartych na GLS, wydaje się modelem najlepiej dopasowanym do polskich danych (rys. 2). Trzy współczynniki dopasowania (c^2/df , RMSEA, CN) spełniały bowiem zalecane kryteria, a dwa współczynniki (RMR, GFI) zbliżyły się do kryteriów CFA na zadowalającym poziomie. Również najniższe wartości współczynników AIC i BIC wskazują na siedmioczynnikowy model dwupoziomowy jako najbardziej dopasowany do polskich danych. Analogicznie, testowanie na podstawie proporcji wariancji także wskazuje na większą adekwatność dopasowania modelu siedmioczynnikowego w porównaniu z modelem pięcioczynnikowym czy też oryginalnym modelem trójczynnikiowym. Wariancja skumulowana (oszacowana na podstawie varimax rotation) modelu zakładającego trzy czynniki ortogonalne ma bowiem wartość 35,44%, podczas gdy dla czterech czynników ortogonalnych wartość wariancji wynosi 47%. Więcej, bo 50,01% wariancji wyjaśnia model pięciu czynników ortogonalnych, a jeszcze więcej struktura siedmiu czynników w strukturze dwupoziomowej (55,04%).

Korelacje

Wyznaczone korelacje między podskalami odzwierciedlają wielowymiarowość i niezależność w strukturze czynnikowej, wskazując na umiarkowaną trafność teoretyczną (moderate construct validity). Z jednej strony dwie podskale ortogonalne drugiego rzędu (SC, AB) stale uzyskują wartości poniżej 0,6 (tab. 4), co wydaje się wskazywać na niezależność tych dwóch czynników. Wszystkie pozostałe czynniki są jednak skorelowane powyżej 0,6, co nie jest zaskoczeniem w przypadku czynników skośnych (SA, SR2) z definicji zawierających pozycje znajdujące się już w innych podskalach (należy zauważyć szczególnie wysoką korelację między SA i SD). Te wartości korelacji wspierają ukośną strukturę polskiej wersji OQ-45. Natomiast wysokie wartości korelacji dla ortogonalnych czynników pierwszego rzędu (SD, IR, SR) wymagają nieco uwagi. Uzyskane wyniki mogą sugerować, że podskale nawet po zmodyfikowaniu mierzą nakładające się konstrukty. Korelacje między podskalami oraz wynikiem całkowitym wynoszą zwykle ponad 0,80, co przemawia za spójnością narzędzia. Wyjątkiem są dwa ukośne czynniki (SC, AB), najprawdopodobniej z powodu małej liczby pozycji w podskalach.

Tabela 4. Korelacje między podskalami a wynikiem całkowitym

N = 1008	TOT	IR	SD	SR	SC	AB	SA	SR2
TOT	1							
IR	0,91	1						
SD	0,95	0,78	1					
SR	0,82	0,68	0,74	1				
SC	0,47	0,33	0,41	0,36	1			
AB	0,12	0,06	0,07	0,08	0,15	1		
SA	0,91	0,71	0,98	0,7	0,39	0,07	1	
SR2	0,9	0,81	0,82	0,85	0,51	0,07	0,76	1

IR – podskala Relacje Interpersonalne, SD – podskala Nasilenie Objawów, SR – podskala Rola Społeczna, SC – podskala Konflikty Społeczne, AB – podskala Zachowania Uzależnieniowe, SA – podskala Somatyzacja i Lęk, SR2 – podskala Rola Społeczna 2, TOT – wynik całkowity

Kryterium równoważności

Tabela 5 pokazuje różnice w osiąganych wynikach (mierzonych między pierwszą a ostatnią sesją) pomiędzy polską grupą kliniczną a polską grupą niekliniczną.

Tabela 5. Średnie i odchylenia standardowe dla wyników OQ w polskiej grupie klinicznej i nieklinicznej

	Grupa kliniczna n = 582		Grupa niekliniczna (studenci) n = 426	
	Średnia	Odch. std.	Średnia	Odch. std.
IR	26,28	7,59	13,91	6,88
SD	33,8	9,99	20,51	8,86
SR	9,87	3,31	6,79	2,85
SC	4,82	2,18	4,49	1,88
AB	0,47	1,02	0,59	1,17
SA	20	6,8	12	5,6
SR2	17	4,3	12	4,1
TOT	75,22	19,39	46,28	18,23

IR – podskala Relacje Interpersonalne, SD – podskala Nasilenie Objawów, SR – podskala Rola Społeczna, SC – podskala Konflikty Społeczne, AB – podskala Zachowania Uzależnieniowe, SA – podskala Somatyzacja i Lęk, SR2 – podskala Rola Społeczna 2, TOT – wynik całkowity

Grupa polskich niepacjentów porównana z grupą polskich pacjentów funkcjonuje na istotnie lepszym poziomie na wszystkich podskalach oraz w zakresie wyniku całkowitego, współczynnik λ Wilksa wynosi 0,52, $F(7, 1000) = 133,76$, $p <$

0,00001. Współczynniki wielkości efektu dla różnicy między próbkami z populacji klinicznej i nieklinicznej są zwykle bardzo duże (np. dla podskali IR $F(1, 1006) = 745,14$, $p < 0,00001$, $d = 1,7$, 95% CI = [1,25, 2,15], a dla wyniku całkowitego $F(1, 1006) = 246,21$, $p < 0,00001$, $d = 1,53$, 95% CI = [0,37, 2,7]), z wyjątkami dla skal SC i AB, które mają małe i nieistotne wartości: dla SC $F(1, 1006) = 8,47$, $p < 0,0037$, $d = 0,16$, 95% CI = [0,03, 0,29], a dla AB $F(1, 1006) = 2,93$, $p < 0,0874$, $d = -0,11$, 95% CI = [-0,18, -0,04].

Porównanie grup polskich z grupami amerykańskimi dało następujące rezultaty: średnie wyniki polskiej grupy nieklinicznej są wyższe niż amerykańskich grup nieklinicznych, np. amerykańskiej próbki uniwersyteckiej ($t = 12,4$, $p < 0,000001$, $d = 0,68$, 95% CI = [-1,13, 1,59]) czy też amerykańskiego społeczeństwa ($t = 9,4$, $p < 0,000001$, $d = 0,48$, 95% CI = [-0,98; 1,1]). Z kolei współczynnik wielkości efektu między polską próbą kliniczną a amerykańską próbą kliniczną jest bardzo mały i statystycznie nieistotny ($t = 1,86$, $p = 0,18$, $d = 0,05$, 95% CI = [-1,7; 0,9]).

Zmiana istotna klinicznie i statystycznie

Wartości punktów odcięcia dla polskiego OQ są następujące: 20 dla podskali IR, 27 dla podskali SD, 8 dla SR, 5 dla SC, 1 dla AB, 16 dla SA oraz 14 dla SR2, a dla wyniku całkowitego 60. Wyniki powyżej punktu odcięcia plasują jednostkę w obszarze klinicznie dysfunkcyjnym. Innymi słowy, punkt odcięcia 60 oznacza czułość dla wyniku całkowitego OQ na poziomie 0,90, co oznacza, że 90% niepacjentów przynależy do grupy funkcjonalnej, a 90% grupy klinicznej jest adekwatnie diagnozowana jako dysfunkcyjna. Wartości współczynnika zmiany istotnej statystycznie wyniosły: 8,27 dla SD, 5,73 dla IR, 4,46 dla SR, 4,64 dla SC, 2,45 dla AB, 7,31 dla SA oraz 6,72 dla SR2. Współczynnik RCI dla wyniku całkowitego OQ wynosi 12,9. Innymi słowy, pacjent musi uzyskać poprawę przynajmniej o 13 punktów w skali wyniku całkowitego OQ, aby można było uznać tę zmianę za istotną klinicznie i statystycznie.

Spójność wewnętrzna oraz badanie test-retest

Współczynniki spójności wewnętrznej oraz rzetelność metody powtórnego testowania (test-retest), zarówno w przypadku poszczególnych podskal, jak i wyniku całkowitego polskiej wersji OQ, cechują się wartościami sugerującymi adekwatność siedmioczynnikowego dwupoziomowego modelu. Wartości spójności wewnętrznej (tab. 6) we wszystkich przypadkach, z wyjątkiem podskali SC, są co najmniej na poziomie 0,7. Większość wartości test-retest (tab. 7) wynosi ponad 0,7, a niektórych ponad 0,8. Niższe wartości utrzymywały się powyżej 0,6, co uznawane jest za najniższy poziom rzetelności narzędzia. Można więc stwierdzić, że wyniki metody powtórnego testowania wskazują na stabilność w czasie polskiej wersji OQ-45.

Tabela 6. Spójność wewnętrzna: α Cronbacha

Spójność wewnętrzna	Współczynniki α Cronbacha					
	Grupa kliniczna n = 582		Grupa niekliniczna n = 426		Grupa całkowita n = 1 008	
Podskale	n	α	n	α	n	α
IR	582	0,88	426	0,91	1 008	0,93
SD	582	0,91	426	0,92	1 008	0,93
SR	582	0,80	426	0,78	1 008	0,82
SC	582	0,61	426	0,60	1 008	0,61
AB	582	0,71	426	0,70	1 008	0,71
SR2	582	0,70	426	0,75	1 008	0,76
SA	582	0,84	426	0,85	1 008	0,88
TOT	582	0,93	426	0,95	1 008	0,95

IR – podskala Relacje Interpersonalne, SD – podskala Nasilenie Objawów, SR – podskala Rola Społeczna, SC – podskala Konflikty Społeczne, AB – podskala Zachowania Uzależnieniowe, SA – podskala Somatyzacja i Lęk, SR2 – podskala Rola Społeczna 2, TOT – wynik całościowy

Tabela 7. Rzetelność metody powtórnego testowania (test-retest)
– współczynniki korelacji liniowej Pearsona

	Rzetelność metody powtórnego testowania (test-retest)					
	Grupa kliniczna n = 582		Grupa niekliniczna n = 426		Grupa całkowita n = 1 008	
Podskale	n	r	n	r	n	r
IR	582	0,76	426	0,84	1 008	0,89
SD	582	0,77	426	0,84	1 008	0,87
SR	582	0,7	426	0,72	1 008	0,79
SC	582	0,63	426	0,68	1 008	0,65
AB	582	0,73	426	0,69	1 008	0,72
SR2	582	0,67	426	0,75	1 008	0,8
SA	582	0,78	426	0,8	1 008	0,86
TOT	582	0,77	426	0,87	1 008	0,89

IR – podskala Relacje Interpersonalne, SD – podskala Nasilenie Objawów, SR – podskala Rola Społeczna, SC – podskala Konflikty Społeczne, AB – podskala Zachowania Uzależnieniowe, SA – podskala Somatyzacja i Lęk, SR2 – podskala Rola Społeczna 2, TOT – wynik całościowy

Czułość na zmianę

Czułość na zmianę mierzono między pierwszą a czwartą sesją dla grupy 161 losowo wybranych pacjentów ze wszystkich podgrup klinicznych. Polska wersja OQ charakteryzuje się wysoką czułością na zmianę dla wszystkich podskal: np. SD: $t(161) = 8,43$, $p < 0,001$, $d = 0,56$; IR: $t(161) = 2,04$, $p < 0,001$, $d = 0,27$; AB $t(161) = 4,03$, $p < 0,001$, $d = 0,33$; SA: $t(161) = 8,02$, $p < 0,001$, $d = 0,53$; dla wyniku całkowitego $t(161) = 7,33$, $p < 0,001$, $d = 0,50$.

Dyskusja

Niniejsze badanie stanowi pierwszą próbę określenia właściwości psychometrycznych i trafności czynnikowej polskiego OQ-45, a także jego równoważności z wersją oryginalną narzędzia. Uzyskane wyniki wskazują na słabe dopasowanie trzech oryginalnych czynników OQ do polskich danych, dość dobre dopasowanie modelu pięcioczynnikowego oraz dobre dopasowanie modelu siedmioczynnikowego. Ta ostatnia struktura jest modelem dwupoziomym obejmującym wynik całkowity, pięć czynników ortogonalnych (podskale specyficzne dla polskiego OQ: Konflikty Społeczne i Zachowania Uzależnieniowe oraz oryginalne, lecz zmodyfikowane podskale: Nasilenie Objawów, Relacje Interpersonalne i Rola Społeczna) oraz dwa czynniki ukośne (Somatyzacja i Lęk oraz Rola Społeczna 2). Ten ostatni model, w porównaniu z innymi branymi pod uwagę rozwiązaniami, nie tylko wyjaśnia największy odsetek wariancji, dodatkowo jego wskaźniki dopasowania najczęściej spełniają kryteria zalecane dla CFA. Właściwości psychometryczne tego właśnie modelu oraz jego czułość na zmiany wydają się wystarczające, a punkty odcięcia i wskaźniki zmiany istotnej klinicznie są adekwatne i zbliżone do wskaźników narzędzia oryginalnego. Podsumowując, chociaż polski model OQ różni się od struktury oryginalnej, jego równoważność z wersją amerykańską można uznać za zadowalającą.

Dopasowanie oryginalnego modelu trójczynnikowego do polskich danych

Warto zauważyć, że dopasowanie oryginalnego trójczynnikowego modelu do polskich danych, gdy porównać je metodami ML lub GLS do innych walidacji, było tylko nieco gorsze [9, 11] lub nawet całkiem zbliżone [21]. Pomimo tego zdecydowano się nie uznawać modelu trójczynnikowego za wystarczająco adekwatnie dopasowany do naszych danych. Zdecydowały o tym następujące powody: (1) ładunki czynnikowe oparte na PA i EFA nie odzwierciedlały oryginalnego trójczynnikowego rozkładu pozycji, wręcz przeciwnie – wskazywały na istnienie nowych, specyficznych dla polskich respondentów podskal, (2) model trójczynnikowy uzyskał wyraźnie niższe współczynniki CFA w porównaniu ze strukturą siedmioczynnikową, a także (3) nowe podskale do pewnego stopnia odzwierciedlają rozkład pozycji i strukturę czynnikową uzyskaną w innych badaniach [9, 19, 20].

W świetle niejednoznacznych wyników wcześniejszych badań zajmujących się strukturą czynnikową OQ, nie jest niczym zaskakującym to, że na polskich danych

oryginalna struktura trójczynnikowa testowana za pomocą EFA nie została pozytywnie zweryfikowana, ani to, że słabo spełnia kryteria dopasowania według CFA [9, 11, 12, 18–21]. Autorzy tych badań, tak jak i twórcy OQ-45 [22], z jednej strony uznają OQ za narzędzie zawierające jeden czynnik ogólny (mierzony wynikiem całkowitym) i wiele podskal, z drugiej strony zdają sobie sprawę, że te podskale mogą nie być tak jednoznaczne jak sugerowałyby ich właściwości psychometryczne. Badacze ci wskazują jednocześnie na inne niż trójczynnikowa struktury OQ, sugerując, że potencjalnie mogą one trafniej odzwierciedlać wielowymiarowość funkcjonowania badanej jednostki. W związku z powyższym warto podkreślić wagę tego, że uzyskane wyniki na podstawie analiz EFA i CFA, jak również wyniki korelacji wspierają siedmioczynnikową dwupoziomą strukturę OQ-45 jako najbardziej odpowiednią dla polskich danych.

Właściwości psychometryczne polskiej wersji OQ

Polską wersję OQ cechują adekwatne punkty odcięcia oraz trafny RCI, co oznacza, że polski OQ może być użyteczny dla szacowania zmiany istotnej klinicznie, tj. może służyć do tworzenia trafnych kategorii pacjentów, którzy w trakcie leczenia zostali wyleczeni, których stan się poprawił, pogorszył lub pozostał bez zmian. Punkt odcięcia polskiego OQ wynosi 60, tzn. że polscy respondenci, którzy uzyskali 60 lub mniej punktów w kwestionariuszu OQ, mogą zostać sklasyfikowani jako znajdujący się w zakresie normy, zaś respondenci z 61 lub więcej punktami w zakresie dysfunkcyjnym. Punkt odcięcia polskiego OQ jest zbliżony do punktów odcięcia w innych jego wersjach (holenderska 55, amerykańska 63, hiszpańska 66, włoska 66). Punkty odcięcia trzech oryginalnych (choć zmodyfikowanych) podskal polskiego OQ odpowiadają amerykańskim (IR polska 20, amerykańska 15; SD polska 27, amerykańska 36; SR polska 8, amerykańska 12). Wartość RCI dla polskiego OQ wynosi 13, podczas gdy dla amerykańskiego i holenderskiego 14 [9, 22].

Chociaż trzy oryginalne podskale (SD, IR, SR) w polskiej wersji zostały skrócone i zmodyfikowane, to uzyskały właściwości psychometryczne porównywalne zarówno z amerykańską, jak i innymi wersjami OQ. Spójność wewnętrzna polskiego OQ jest ogólnie wyższa z wyjątkiem podskali SD, która uzyskała te same wartości co w wersji dla danych amerykańskich [22], holenderskich [9], włoskich [11] i szwedzkich [12]. Ponieważ większość współczynników przekracza próg 0,70, uważamy, że polski OQ cechuje dobra spójność wewnętrzna. Podskala SR, jedyna o wartości (α Cronbacha) poniżej 0,70, nadal ma wyraźnie wyższą spójność wewnętrzną w porównaniu z innymi wersjami językowymi, w tym amerykańską. Rzetelność metody powtórnego testowania wyniku całkowitego, jak również trzech pierwszorzędowych ortogonalnych podskal polskiego OQ, w porównaniu z danymi amerykańskimi i holenderskimi jest zwykle wyższa i to zarówno w grupie klinicznej jak i nieklinicznej. Jej zadowalające wartości potwierdzają stabilność w czasie polskiego OQ.

Podskale IR oraz SR są silniej skorelowane, a podskala SD nieco słabiej z wynikiem całkowitym, gdy porównać je z holenderską [9], niemiecką [10] czy włoską [11] wersją kwestionariusza OQ. Niestety, podręcznik OQ [22] nie zawiera korelacji między podskalami i wynikiem całkowitym dla wersji oryginalnej. Uzyskane wyniki

korelacji, wskazujące na wysoką spójność podskal z wynikiem całkowitym, potwierdzają słuszność zmodyfikowania i skrócenia oryginalnych podskal. Najsilniejsza korelacja podskali SD z wynikiem całkowitym zarówno w niniejszym badaniu, jak i w innych walidacjach (np. [9, 10, 21]) może sugerować, że OQ jest raczej narzędziem mierzącym obciążenie objawami niż kwestionariuszem oceniającym funkcjonowanie interpersonalne i pełnienie roli społecznej.

Wskaźniki confirmacyjnej analizy czynnikowej wskazują na dobre dopasowanie siedmioczynnikowego dwupoziomowego modelu do polskich danych. Ponadto dopasowanie to w porównaniu z dopasowaniem modelu trójczynnikowego w innych badaniach zawsze dawało lepsze współczynniki i częściej spełniało zalecane kryteria. Przykładowo struktura polskiego OQ w porównaniu z amerykańskim OQ charakteryzuje się przeważnie lepszymi wskaźnikami dopasowania (GFI, AGFI, NFI, CN), a jedynym gorszym wskaźnikiem jest χ^2/df [21]; w porównaniu z wersją holenderską polskie OQ osiąga lepsze wskaźniki dla χ^2/df , RMR, RMSEA, GFI, a gorsze dla NFI i CFI [9], natomiast w zestawieniu z wersją włoską polskie OQ osiąga lepsze RMR, RMSEA i AGFI, ale gorsze CFI i AIC [11]. Trzeba przy tym zaznaczyć, że pomimo tego, iż siedmioczynnikowy model, gdy go porównać z rozwiązaniami z innych publikacji, wydaje się dość adekwatnym odzwierciedleniem dostępnych danych, to wciąż trzy jego współczynniki nie zbliżają się do zalecanych wartości (χ^2 , NFI, CFI). Niemniej, biorąc pod uwagę względnie duży rozmiar badania, istotny χ^2 nie jest konieczną oznaką słabego dopasowania, jako że wskaźnik ten jest zależny od wielkości próby. Także niskie wyniki NFI, częściowo opartego na χ^2 , można wyjaśnić wielkością próby. Natomiast uzyskane niskie wyniki CFI są typowe dla narzędzi złożonych z wielu pozycji [37].

Ogólnie rzecz biorąc, powyższe wyniki mówią o stosownej równoważności między polskim modelem a innymi wersjami językowymi oraz wskazują zarówno na statystyczną, jak i kliniczną wartość polskiego zbioru podskal.

Podskale specyficzne dla polskiego OQ

Podskala Zachowań Uzależnieniowych z jej punktem odcięcia wynoszącym zero oraz wartością współczynnika RCI równą trzy cechuje się wysoką wartością kliniczną, niemalże na poziomie czułości diagnostycznej. Wystarczy bowiem uzyskać tylko jeden punkt na podskali AB, aby ustalić, że jednostka ma istotne problemy z alkoholem i/lub narkotykami. Z kolei o uzyskaniu istotnej klinicznie zmiany można mówić wtedy, gdy wynik respondenta zmieni się o wszystkie trzy możliwe punkty. Oczywiście, ponieważ czynnik AB składa się tylko z trzech pozycji, jego właściwości psychometryczne, wartość RCI, punkty odcięcia i korelacje mogą być do pewnego stopnia wypaczone. Niemniej ładunek czynnikowy, spójność wewnętrzna oraz rzetelność test-retest powodują, że można tę podskalę uznać za porównywalną z pozostałymi. Podskala AB wydaje się więc klinicznie uzasadnionym odstępstwem od oryginalnej struktury trójczynnikowej.

Kliniczna użyteczność podskali Konflikt Społeczny nie jest tak jednoznaczna jak w przypadku podskali AB, co można przypisać jej względnej heterogeniczności. Prawdopodobnie czynnik mierzący taki wielokontekstowy fenomen, jak konflikty społeczne, mógłby zyskać na dodaniu nowych pozycji i/lub modyfikacji treści już

istniejących pytań. Warto przy tym zauważyć, że właściwości psychometryczne podskali SC są zaniżone poprzez problematyczne pytanie OQ14. Ta pozycja wydaje się najmniej homogeniczna ze wszystkich składowych SC. Jej usunięcie z OQ zwiększyłoby wartość współczynnika α Cronbacha do 0,68 dla całej grupy, do 0,69 dla grupy klinicznej oraz do 0,63 dla grupy nieklinicznej. Warto przy tym zauważyć, że taki sam efekt, tzn. wzrost α Cronbacha po odrzuceniu OQ14, został zaobserwowany także dla innych walidacji [12].

Względnie słabe właściwości psychometryczne podskal SC, AB oraz SR2 można przypisać małej liczbie pozycji w każdej z tych podskal. Ich rzetelności i spójności wewnętrzne były zwykle lepsze, gdy podgrupy kliniczna i niekliniczna zostały połączone w grupę całkowitą.

Podskala SA okazała się klinicznie i statystycznie użyteczna. Jej homogeniczne pozycje wydają się adekwatnie mierzyć zjawiska powiązane z somatyzacją i lękiem. Dość interesujące jest to, że pod względem rozkładu pozycji podskala SA w pełni odzwierciedla czynnik ASD w holenderskim OQ [9], uzyskując jednak wyraźnie lepsze właściwości psychometryczne, gdy bierze się pod uwagę ładunki czynnikowe, korelacje z oryginalnymi podskalami, spójność wewnętrzną dla populacji nieklinicznej oraz rzetelność metody powtórnego testowania. Dodatkowo stwierdzono idealną równowagę podskal SA i ASD pod względem współczynnika RCI, korelacji z wynikiem całkowitym i spójności wewnętrznej dla populacji klinicznej.

Czynnik SR2 miał w porównaniu z innymi podskalami mniej lub bardziej równoważne właściwości psychometryczne. Z uwagi na jego względną heterogeniczność i niejasną kliniczną użyteczność, w kolejnych badaniach nad strukturą polskiego OQ, może on zostać zmodyfikowany lub usunięty.

Mocne strony i ograniczenia pracy

Wyniki niniejszego badania mogą być wartościowe nie tylko dla terapeutów, którzy rozważają korzystanie z polskiej wersji OQ-45 w swojej praktyce klinicznej, ale również dla badaczy, którzy tworzą narzędzia oceniające skuteczność psychoterapii. Do zalet tego badania należą: (1) rzetelny przegląd sprzecznych wyników wcześniejszych prac w zakresie struktury czynnikowej OQ, (2) duża i akceptowalnie heterogenna całościowa grupa badana, (3) warunki kliniczne odzwierciedlające psychoterapię odbywającą się w naturalistycznych warunkach (treatment-as-usual), (4) dane pozyskiwane w kilku odstępach czasowych, (5) kilka starannie przeprowadzonych analiz różnych możliwych struktur czynnikowych narzędzia, (6) pierwsze oszacowanie mierników rzetelności polskiej wersji OQ oraz (7) omówienie wpływu różnic kulturowych oraz wpływu czasu na wyniki. Pomimo powyższych zalet należy być ostrożnym w wyciąganiu zbyt generalnych wniosków z tego badania, a to ze względu na: (1) stosunkowo niewielkie rozmiary każdej z podgrup, (2) dominację młodych dorosłych w podgrupie nieklinicznej, (3) brak rozkładu normalnego, (4) małe współczynniki wielkości efektu, które świadczą o słabej czułości poszczególnych podskal oraz (5) niewiele ponad 50% wariancji (55,04%) wyjaśnianej za pomocą siedmioczynnikowej dwupoziomowej struktury kwestionariusza.

Ze względu na obszerność wykonanych w niniejszym projekcie analiz autorzy zdecydowali się omówić skuteczność oraz stabilność badanego narzędzia w innej pracy [38], w której zaproponowano mierniki oceny skuteczności oparte na łańcuchach Markowa.

Wnioski

Artykuł prezentuje wyniki badań empirycznych – przeprowadzonych z zastosowaniem analizy czynnikowej – dotyczących polskiej adaptacji amerykańskiego Kwestionariusza Oceny Skuteczności Psychoterapii. Podsumowując, siedmioczynnikowy dwupoziomowy model polskiego OQ wydaje się – przy uwzględnieniu ograniczeń niniejszego badania – zarówno empirycznie, jak i klinicznie stosowną miarą oceny specyficznych obszarów funkcjonowania pacjenta (w zakresie relacji interpersonalnych, w tym pełnienia ról społecznych i radzenia sobie z konfliktami społecznymi, nasilenia objawów, w tym somatyzacji i lęku, oraz zachowań uzależnieniowych). Właściwości psychometryczne polskiego OQ są spójne z amerykańskim oryginałem oraz z adaptacjami na inne języki. Rzetelność i trafność poszczególnych podskal, włączając w to nowe podskale, są obiecujące, co sugeruje, że polska wersja OQ jest wystarczająco czułym narzędziem dla opisu polskich pacjentów. Czułość OQ na zmiany w zadowalającym stopniu pozwala na rozróżnianie między badanymi z grup dysfunkcyjnej i nieklinicznej. Specyficzne punkty odcięcia pozwalają na oszacowanie klinicznie istotnych zmian u polskich pacjentów. Warto zauważyć jest to, że chociaż oryginalne podskale (SD, IR, SR) zostały zmodyfikowane i skrócone oraz pomimo tego, że wymodelowano czynniki dodatkowe (AB, SC, SA, SR2), równoważność polskiego OQ wobec wersji oryginalnej jest zadowalająca.

Niemniej stosując OQ w warunkach klinicznych – analogicznie jak w przypadku oryginalnego OQ – bezpieczniej jest polegać raczej na wyniku całkowitym OQ niż na poszczególnych podskalach, chociażby z racji tego, że oryginalna struktura trójczynnikowa OQ – stanowiąca punkt odniesienia dla niniejszych badań – nie wynika z eksploracyjnej analizy czynnikowej, lecz jest efektem arbitralnych decyzji autorów narzędzia [1].

Różnice między polską i amerykańską strukturą czynnikową można tłumaczyć licznymi rozbieżnościami kulturowymi między polskim a amerykańskim społeczeństwem [33]. To założenie w pewnym stopniu znajduje statystyczne odzwierciedlenie w rozbieżnościach pomiędzy podgrupami klinicznymi i nieklinicznymi. Pacjenci z obu krajów są dość podobni w swoich psychologicznych cechach mierzonych przez OQ, zaś niepacjenci z obu społeczeństw różnią się od siebie istotnie. Słabe dopasowanie oryginalnego modelu trójczynnikowego do polskich danych może natomiast do pewnego stopnia być spowodowane upływem czasu. Kwestionariusz OQ został opracowany 19 lat temu [1]. Nie można więc wykluczyć, że choć jego struktura była wtedy zapewne adekwatna, to obecnie nie aż tak precyzyjnie oddaje funkcjonowanie jednostek. Wydaje się to mieć szczególne znaczenie dla polskiego społeczeństwa, które znacząco zmienia się od ponad 25 lat, stopniowo odchodząc od wartości wspólnotowych na rzecz celów indywidualistycznych oraz porzucając zachowania zorientowane

na przetrwanie, a coraz bardziej rozwijając aktywności ukierunkowane na odniesienie sukcesu przy jednoczesnym kultywowaniu stylu życia opartego na wirtualnej technologii [39]. Jest dość prawdopodobne, że dodanie pytań dotyczących takich zagadnień jak np. równowaga między pracą a domem, korzystanie z portali społecznościowych itp. mogłoby zwiększyć adekwatność OQ nie tylko w odniesieniu do polskiej populacji.

Niniejsza adaptacja ma wstępny charakter i niewątpliwie potrzebne są kolejne badania nad polską wersją OQ w celu doprecyzowania trafności czynników specyficznych dla tej wersji, jak również w celu dalszego testowania równoważności pojęciowej z oryginalnym OQ. Do oszacowania pozostaje również trafność krzyżowa polskiego OQ z innymi narzędziami normalizowanymi w Polsce. Większa liczba badanych w każdej podgrupie mogłaby pomóc ustalić punkty odcięcia i wartości RCI dla bardziej homogennych grup, np. pacjentów oddziałów całodobowych, pacjentów ambulatoryjnych czy też pacjentów ze specyficznym rozpoznaniem. Mamy również nadzieję, że nasze wyniki rozpoczną dyskusję nad zasadnością zrewidowania treści pytań oraz modyfikacji struktury czynnikowej oryginalnego OQ.

Piśmiennictwo

1. Lambert MJ, Burlingame GM, Umphress V, Hansen NB, Vermeersch DA, Clouse GC. i wsp. *The reliability and validity of the Outcome Questionnaire*. Clin. Psychol. Psychother. 1996; 3: 249–258.
2. Hatfield DR, Ogles BM. *The use of outcome measures by psychologists in clinical practice*. Prof. Psychol. Res. Pract. 2004; 35: 485–491.
3. Lambert MJ, Harmon C, Slade K, Whipple JL, Hawkins EJ. *Providing feedback to psychotherapists on their patients' progress: Clinical results and practice suggestions*. J. Clin. Psychol. 2005; 61: 165–174.
4. Slade KL, Lambert MJ, Harmon SC, Smart DW, Bailey R. *Improving psychotherapy outcome: The use of immediate electronic feedback and revised clinical support tools*. Clin. Psychol. Psychother. 2008; 15: 287–303.
5. Simon W, Lambert MJ, Harris MW, Busath G, Vazquez A. *Providing patient progress information and clinical support tools to therapists: Effects on patients at risk of treatment failure*. Psychother. Res. 2012; 22: 638–647.
6. Crits-Christoph P, Ring-Kurtz S, Hamilton JL, Lambert MJ, Gallop R, McClure B. i wsp. *A preliminary study of the effects of individual patient-level feedback in outpatient substance abuse treatment programs*. J. Subst. Abuse Treat. 2012; 42: 301–309.
7. Simon W, Lambert MJ, Busath G, Vazquez A, Berkeljon A, Hyer K. i wsp. *Effects of providing patient progress feedback and clinical support tools to psychotherapists in an inpatient eating disorders treatment program: A randomized controlled study*. Psychother. Res. 2013; 23: 287–300.
8. Shimokawa K, Lambert MJ, Smart DW. *Enhancing treatment outcome of patients at risk of treatment failure: Meta-analytic and mega-analytic review of psychotherapy quality assurance system*. J. Cons. Clin. Psychol. 2010; 78: 298–311.
9. de Jong K, Nugter MA, Polak MG, Wagenborg JEA, Spinhoven P, Heiser WJ. *The Outcome Questionnaire (OQ-45) in a Dutch population: A cross-cultural validation*. Clin. Psychol. Psychother. 2007; 14: 288–301.

10. Lambert MJ, Hannöver W, Nisslmüller K, Richard M, Kordy H. *Fragebogen zum ergebnis von psychotherapie: Zur reliabilität und validität der deutschen übersetzung des Outcome Questionnaire 45.2*. Z. Klin. Psychol. Psychother. 2002; 31: 40–46.
11. Lo Coco G, Chiappelli M, Bensi L, Gullo S, Prestano C, Lambert MJ. *The factorial structure of the Outcome Questionnaire-45: A study with an Italian sample*. Clin. Psychol. Psychother. 2008; 15: 418–423.
12. Wennberg P, Philips B, de Jong K. *The Swedish version of the Outcome Questionnaire (OQ-45): Reliability and factor structure in a substance abuse sample*. Psychol. Psychother. Theory Res. Pract. 2010; 83: 325–329.
13. Aleksandrowicz JW, Bierzyński K, Filipiak J, Kowalczyk E, Martyniak J, Mahoń S. i wsp. *Kwestionariusze objawowe „S” i „O” – narzędzia służące do diagnozy i opisu zaburzeń nerwicowych*. Psychoterapia 1981; 37: 11–27.
14. Sobański JA, Klasa K, Cyranka K, Mielimaka M, Dembińska E, Müldner-Nieckowski Ł. i wsp. *Skuteczność intensywnej psychoterapii na oddziale dziennym oceniana za pomocą Kwestionariusza Osobowości Nerwicowej KON-2006*. Psychiatr. Pol. 2014 [E-pub ahead of print; DOI: 10.12740/psychiatriapolska.pl/online-first/6].
15. Sobański JA, Klasa K, Rutkowski K, Dembińska E, Müldner-Nieckowski Ł. *Profil objawów pacjentów, którzy przegrali psychoterapię kompleksową w dziennym oddziale leczenia nerwic. Badanie retrospektywne*. Psychiatr. Psychoter. 2013; 9(4): 14–30.
16. Sobański JA, Müldner-Nieckowski Ł, Klasa K, Dembińska E, Rutkowski K, Cyranka K. *Traumatic childhood sexual events and secondary sexual health complaints in neurotic disorders*. Arch. Psychiatrii Psychother. 2013; 15(3): 19–32.
17. Sobański JA, Cyranka K, Rodziński P, Klasa K, Rutkowski K, Dembińska E. i wsp. *Czy cechy osobowości i nasilenie objawów nerwicowych wiążą się ze sprawozdawaniem myśli samobójczych u pacjentów dziennego oddziału leczenia zaburzeń nerwicowych?* Psychiatr. Pol. 2014 [E-pub ahead of print; DOI: 10.12740/psychiatriapolska.pl/online-first/5].
18. Bludworth JL, Tracey TJG, Glidden-Tracey C. *The bi-level structure of the Outcome Questionnaire-45*. Psychol. Assess. 2010; 22: 350–355.
19. Kim SH, Beretvas SN, Sherry AR. *A validation of the factor structure of OQ-45 scores using factor mixture modeling*. Meas. Eval. Couns. Dev. 2010; 42: 275–295.
20. Minami T, Davies DR, Tierney SC, Bettmann JE, McAward SM, Averill LA. i wsp. *Preliminary evidence on the effectiveness of psychological treatments delivered at a university counseling center*. J. Couns. Psychol. 2009; 56: 309–320.
21. Mueller RM, Lambert MJ, Burlingame GM. *Construct validity of the outcome questionnaire: A confirmatory factor analysis*. J. Pers. Assess. 1998; 70: 248–262.
22. Lambert MJ, Morton JJ, Hatfield D, Harmon C, Hamilton S, Reid RC. i wsp. *Administration and Scoring Manual for the Outcome Questionnaire – 45*. Orem, UT: American Professional Credentialing Services; 2004.
23. Derogatis LR. *The SCL-90 manual: Scoring, administration and procedures for the SCL-90*. Baltimore: Johns-Hopkins University School of Medicine, Clinical Psychometrics Unit; 1977.
24. Beck AT, Steer RA, Garbin MG. *Psychometric properties of the Beck Depression Inventory: Twenty-five years later*. Clin. Psychol. Rev. 1988; 8: 77–100.
25. Spielberger CD. *Manual for the State-Trait Anxiety Inventory: STAI (Form Y)*. Palo Alto, CA: Consulting Psychologists Press; 1983.
26. Timmerman M, Lorenzo-Seva U. *Dimensionality assessment of ordered polytomous items with parallel analysis*. Psychol. Methods 2011; 16: 209–220.

27. Hooper D, Coughlan J, Mullen M. *Structural equation modelling: Guidelines for determining model fit*. Electron. J. Business Res. Methods 2008; 6: 53–60.
28. Osborne JW, Costello AB, Kellow JT. *Best practices in exploratory factor analysis*. W: Osborne JW. red. *Best practices in quantitative methods*. Newbury Park, CA: Sage; 2008. s. 86–124.
29. Jacobson NS, Truax P. *Clinical significance: A statistical approach to defining meaningful change in psychotherapy research*. J. Cons. Clin. Psychol. 1991; 59: 12–19.
30. Lambert MJ, Hansen NB, Bauer S. *Assessing the clinical significance of outcome results*. W: Nezu AM, Maguth-Nezu C. red. *Evidence-based outcome research: A practical guide to conducting randomized controlled trials for psychological interventions*. New York: Oxford University Press; 2008. s. 359–378.
31. Cohen J. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum; 1988.
32. Zalewska M, Niemiro W, Samoliński B. *MCMC imputation in autologistic model*. Monte Carlo Methods and Appl. 2010; 16: 421–438.
33. Zalewski T. *Inne Stany. Czym różnią się od nas Amerykanie?* Warszawa: Biblioteka „Polityki”; 2011.
34. Kahn WA. *Psychological conditions of personal engagement and disengagement at work*. Acad. Manage. J. 1990; 4: 692–724.
35. Ayduk O, Mendoza-Denton R, Mischel W, Downey G, Peake PK, Rodriguez M. *Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity*. J. Pers. Soc. Psychol. 2000; 5: 776–792.
36. Mannheim K. *The conditions of social performance*. London: Routledge; 2002.
37. Marsh HW, Hau K, Grayson D. *Goodness of fit in structural equation models*. W: Maydeu-Olivares A, McArdle JJ. red. *Contemporary psychometrics: A festschrift for Roderick P. McDonald*. Mahwah, NJ: Erlbaum; 2005. s. 275–340.
38. Śliwka P, Simon W. *Markov Chains as a Tool Measuring Effectiveness of a Psychotherapy Process*. W: BIOMAT 2014. Proceedings of the International Symposium on Mathematical and Computational Biology, edited by: Rubem P Mondaini, p. 232–244, DOI: 10.1142/9789814667944_0015, 2015
39. Aleksandrowicz JW, Sobański JA. *Kwestionariusz objawowy S-III*. Psychiatr. Pol. 2011; 45(4): 515–526.

Adres: Witold Simon
Konsonans. Ośrodek Psychoterapii
02-796 Warszawa, ul. Wąwózowa 22/61

Otrzymano: 19.04.2015
Zrecenzowano: 1.06.2015
Otrzymano po poprawie: 10.08.2015
Przyjęto do druku: 11.08.2015