

KRZYSZTOF FRONCZYK

Wyższa Szkoła Finansów i Zarządzania w Warszawie

Wydział Psychologii

IDENTYFIKACJA ODPOWIADANIA LOSOWEGO LUB NIEUWAŻNEGO W KWESTIONARIUSZU NA PRZYKŁADZIE NEO-FFI

W artykule zaprezentowano dwa mało znane wskaźniki odpowiadania losowego lub nieuważnego: indeks sabotowania Cattella oraz wskaźnik indywidualnego losowego wyniku oczekiwanego. Oba wskaźniki służą do identyfikacji osób, które w kwestionariuszach wielowymiarowych udzielają odpowiedzi niezwiązanych z treścią pytań, np. odpowiedzi losowych. Weryfikacja trafności tych wskaźników polegała na sprawdzeniu, w jakim stopniu pozwalają one odróżnić wyniki rzeczywiste uzyskane w kwestionariuszu NEO-FFI od danych losowych wygenerowanych przez komputer. W badaniu wzięły udział 943 osoby oraz wykorzystano wygenerowane losowo odpowiedzi, symulujące odpowiedzi 1000 osób. Okazało się, że na podstawie obu wskaźników jednocześnie, przy użyciu regresji logistycznej, możliwe było dość dobre odróżnienie danych rzeczywistych od losowych (poprawna identyfikacja w 86% wszystkich wyników). Wynik ten jest dość wysoki, biorąc pod uwagę, że być może część osób wypełniających kwestionariusz NEO-FFI mogła również w rzeczywistości odpowiadać losowo.

Słowa kluczowe: odpowiadanie losowe, odpowiadanie nieuważne, wskaźnik sabotowania Cattella, wskaźnik indywidualnego losowego wyniku oczekiwanego, NEO-FFI.

W kwestionariuszowych badaniach osobowości występuje wiele czynników zniekształcających uzyskiwane wyniki. Nichols, Greene i Schmolck (1989) wyróżnili dwa zasadnicze typy zniekształceń odpowiedzi. Pierwszy z nich to zniekształcanie związane z treścią pozycji testowych (*content responsive faking*). Typ

ten charakteryzuje się tym, że udzielane odpowiedzi są zależne od treści pytań, ale nie odzwierciedlają rzeczywistej samowiedzy osób badanych, na przykład symulowanie lub minimalizowanie objawów oraz odpowiadanie zgodnie z oczekiwaniami społecznymi (*fake bad* lub *fake good*). Jest to pewien sposób autoprezentacji uwarunkowany sytuacyjnie lub osobowościowo (Zawadzki, 2006).

Niniejszy artykuł jest poświęcony drugiemu typowi zniekształceń wyróżnionych przez Nichols i współpracowników (1989). Są to zniekształcenia niezwiązane z treścią pozycji testowych (*content nonresponsivity, noncontent responding*). Do tego rodzaju zniekształceń zalicza się zarówno braki odpowiedzi na pytania kwestionariusza, jak i pewne style odpowiadania, takie jak odpowiadanie losowe lub nieuważne (Beach, 1989). Taki sposób odpowiadania może być nazywany za Cattellem sabotowaniem badania (Cattell, Eber i Tatsuoka, 1970).

Przez odpowiadanie nieuważne zwykle rozumie się udzielanie niezgodnych odpowiedzi na pytania o podobnej treści lub na pary pytań, których odpowiedzi są skorelowane. Natomiast odpowiadanie losowe określane jest jako odpowiadanie niezależne od treści pozycji testowych, tak jakby badany losowo zaznaczał odpowiedzi na pytania, bez czytania ich treści (Evans i Dinning, 1983). Ponadto w odpowiadaniu losowym nie ma jakiegokolwiek konsekwencji, a więc zniekształcenia tego typu nie mogą być interpretowane jako przejaw pewnego sposobu autoprezentacji (Nichols i Greene, 1997). Zarówno konceptualne, jak i empiryczne rozróżnienie odpowiadania losowego od nieuważnego jest dość trudne. Większość istniejących wskaźników jest podobnie wrażliwa na oba z nich. W niniejszym artykule odpowiadanie losowe i nieuważne będzie więc traktowane łącznie.

Odpowiedzi losowe zwiększają wariancję błędu, co zmniejsza skorelowanie pozycji testowych, przyczyniając się do obniżenia zgodności wewnętrznej kwestionariusza. Podobnie na poziomie indywidualnym, im więcej dana osoba udzieliła losowych odpowiedzi, tym mniejsza jest zgodność jej odpowiedzi. W artykule szerzej zostaną omówione dwa spośród szeregu różnych wskaźników diagnozowania udzielania odpowiedzi losowych. Jest nim wskaźnik sabotowania (*sabotage*), opracowany przez Cattella i współpracowników (1970), oraz wskaźnik indywidualnego losowego wyniku oczekiwanego (*fixed individualized chance score*) (Haertzen i Ross, 1978). Każdy z tych wskaźników został skonstruowany w odmienny sposób.

Wskaźnik sabotowania Cattella

Zastosowany w 16-czynnikowym kwestionariuszu osobowości Cattella wskaźnik sabotowania opracowano na podstawie założenia, że każdą skalę kwestionariusza można podzielić na dwie części (Cattell i in., 1970). Wynik obliczony na podstawie połowy pytań danej skali można estymować, stosując metodę regresji, na podstawie drugiej połowy pytań tej skali. Takie szacowanie jest możliwe przy założeniu, że badani odpowiadają, odwołując się do wiedzy na własny temat, nie odpowiadają w sposób losowy oraz nie ulegają żadnym stylom odpowiadania. Na podstawie uzyskanych danych empirycznych można więc obliczyć dwie grupy wyników: rzeczywiste wyniki dla każdej z dwóch części danej skali oraz wyniki szacowane na podstawie drugiej części. Łączny wskaźnik sabotowania jest obliczany jako suma kwadratów różnic między rzeczywistymi wynikami jednej połowy każdej ze skal kwestionariusza a wynikami oszacowanymi na podstawie regresji liniowej wyniku z pierwszej połowy względem wyniku z drugiej. Innymi słowy, wskaźnik ten jest sumą obliczoną dla poszczególnych skal kwestionariusza kwadratów reszt z regresji jednej części danej skali względem drugiej. Im niższa wartość tego wskaźnika, tym mniejsze prawdopodobieństwo, że osoba, dla której obliczono wskaźnik, odpowiadała w sposób losowy bądź nieuważny. Niski wynik oznacza bowiem dość dużą zgodność między wyróżnionymi częściami poszczególnych skal.

W praktyce Cattell i współpracownicy (1970) zalecali, aby do obliczania wskaźnika sabotowania wykorzystywać uśrednioną korelację par połówek wszystkich skal, a obliczenia przeprowadzać na wynikach znormalizowanych. Wynikało to prawdopodobnie z chęci uproszczenia obliczeń.

W przypadku Kwestionariusza Osobowości Cattella wskaźnik sabotowania nie przyniósł jednak zbyt dobrych rezultatów, jeśli chodzi o identyfikację odpowiedzi losowych (O'Dell, 1971). Część wyników została zaklasyfikowana jako wyniki losowe, mimo że były to rzeczywiste wyniki osób badanych. W porównaniu z nim lepsze rezultaty uzyskano na podstawie innych skal kontrolnych (Irvine i Gendreau, 1974). Możliwe jest jednak, że w przypadku innych kwestionariuszy, charakteryzujących się większą zgodnością wewnętrzną swoich skal, wskaźnik ten przyniósłby lepsze rezultaty.

Nie jest także wykluczone, że wynik uzyskany przez O'Della (1971) jest pochodną uproszczonej procedury obliczania wskaźnika sabotowania. Można zakładać, że wykorzystanie dokładnych wartości współczynników regresji poszczególnych połówek każdej ze skal kwestionariusza (a nie wartości uśrednionych dla wszystkich skal) powinno poprawić możliwości stosowania omawiane-

go wskaźnika. Ponieważ oprócz wspomnianych dwóch analiz (O'Dell, 1971; Irvine i Gendreau, 1974) nie są dostępne żadne inne wyniki dotyczące tego zagadnienia, warto te badania kontynuować.

Wskaźnik indywidualnego losowego wyniku oczekiwanego

Możliwe jest jeszcze jedno, odmienne podejście do diagnozowania odpowiadania losowego bądź nieuważnego. W podejściu tym wyznacza się wartości oczekiwane poszczególnych skal kwestionariusza przy założeniu, że badany odpowiada losowo. Jeśli na poziomie indywidualnym proporcja odpowiedzi danego typu (np. odpowiedzi „tak”) przy odpowiadaniu losowym jest taka sama w całym kwestionariuszu, jak i w każdej z jego skal, to można obliczyć wyniki oczekiwane każdej ze skal (FIC, *fixed individualized chance score*) na podstawie proporcji poszczególnych rodzajów odpowiedzi w całości kwestionariusza, niezależnie od ich przynależności do skal treściowych, co zastosowali Haertzen i Ross (1978). W najprostszym przypadku, gdy weźmie się pod uwagę kwestionariusz o dwukategorialnym formacie odpowiedzi („tak”-„nie”), można obliczyć wartość oczekiwaną danej skali w następujący sposób:

$$FIC_i = (P_T \times T_i) + (Q \times N_i)$$

gdzie:

FIC_i – to indywidualny wynik oczekiwany i -tej skali (*fixed individualized chance score*)

P_T – proporcja odpowiedzi twierdzących w całym kwestionariuszu, udzielonych przez daną osobę niezależnie od skali

T_i – liczba odpowiedzi twierdzących diagnostycznych w i -tej skali

$Q = 1 - P_T$, czyli proporcja odpowiedzi przeczących w całym kwestionariuszu, udzielonych przez daną osobę niezależnie od skali

N_i – liczba odpowiedzi przeczących diagnostycznych w i -tej skali

Indywidualne wyniki oczekiwane przy założeniu losowego sposobu odpowiadania, uzyskane w każdej ze skal, mogą być porównane z rzeczywistymi wynikami. Różnice w każdej ze skal między wynikiem oczekiwanym a wynikiem rzeczywistym, podniesione do kwadratu i zsumowane, dają łączny wskaźnik odpowiadania indywidualnego losowego wyniku oczekiwanego wszystkich skal. Im wyższa wartość tego wskaźnika, tym mniejsze prawdopodobieństwo, że dana osoba odpowiadała w sposób losowy, gdyż jej wyniki znacznie odbiegają od profilu wyznaczonego na podstawie indywidualnych wyników oczekiwanych.

Podobnie jak w przypadku wskaźnika sabotowania Cattella, tak i w przypadku wskaźnika indywidualnego losowego wyniku oczekiwanego, nie ma zbyt wiele publikacji na jego temat. Wprawdzie w polskiej literaturze naukowej wskaźnik ten był już opisywany (Paluchowski, 1983), jednak liczba badań testujących jego trafność ogranicza się do analiz przeprowadzonych przez Haertzena i Rossa (1978) oraz Rossa i Haertzena (1979). Badacze ci wykazali bardzo wysoką trafność swojego wskaźnika z wykorzystaniem mało znanego kwestionariusza *Social Experience Questionnaire*. Analizowano dwa rodzaje danych: dane pochodzące od rzeczywistych osób badanych oraz dane wygenerowane losowo. Wskaźnik ten umożliwił prawidłowe zidentyfikowanie wszystkich danych wygenerowanych losowo, a błędnie zakwalifikowano jako losowe jedynie 3,7% rzeczywistych danych. Ten obiecujący rezultat uzyskano jednak na dość specyficznych, ograniczonych liczebnościowo danych. Nie są znane inne prace z wykorzystaniem wskaźnika indywidualnego losowego wyniku oczekiwanego. Konieczne są więc dalsze badania na większych grupach badanych z wykorzystaniem częściej stosowanych kwestionariuszy.

Metodologia weryfikacji trafności wskaźników odpowiadania losowego

W celu weryfikacji możliwości diagnostycznych różnych metod identyfikacji odpowiadania losowego bądź nieuważnego przeprowadza się często badania symulacyjne, w których generuje się komputerowo odpowiedzi udzielane losowo. Są one wykorzystywane jako dane kryterialne względem rzeczywistych danych dla różnych wskaźników odpowiadania losowego (np. Archer i Elkins, 1999; Pinsoneault, 2002, 2005). O wysokiej trafności danego wskaźnika świadczą jego możliwości różnicowania między rzeczywistymi a symulowanymi danymi, co można sprawdzić, stosując np. analizę dyskryminacji lub analizę regresji, tak jak np. w badaniach Baera, Ballengera, Berry'ego i Weltera (1997) oraz Archera i Elkinsa (1999).

Dane symulacyjne czasem są krytykowane ze względu na ich inny charakter niż dane uzyskiwane w warunkach rzeczywistych. Ponadto prawdopodobnie niektórzy ludzie, mimo że odpowiadają zgodnie z samowiedzą, to ich odpowiedzi są nietypowe, tak jakby losowe. W praktyce badawczej jest to jednak najprostsza metoda uzyskania danych, co do których istnieje pewność, że powstały w sposób losowy. Oczywiście można poprosić osoby badane, aby odpowiadały w sposób losowy czy też nie czytając treści pytań. Jednak wobec takich danych

można by postawić podobny zarzut sztuczności sytuacji, w jakiej powstały, i ich oderwania od realnych warunków przeprowadzania badań diagnostycznych.

Cel badania i założenia badawcze

Celem badania była weryfikacja trafności oraz wybór najlepszego spośród dwóch wskaźników odpowiadania losowego bądź nieuważnego w kwestionariuszu NEO-FFI: wskaźnika sabotowania oraz wskaźnika indywidualnego losowego wyniku oczekiwanego, a także łącznego zastosowania obu tych wskaźników. Kwestionariusz ten wybrano ze względu na jego dość dużą popularność w praktyce psychologicznej oraz na to, że w przeciwieństwie do wielu inwentarzy psychologicznych nie ma on żadnych skal kontrolnych ani innych wskaźników stylu odpowiadania.

Zdecydowano się na wykorzystanie wskaźnika sabotowania zaproponowanego przez Cattella i współpracowników (1970) oraz wskaźnika indywidualnego losowego wyniku oczekiwanego Haertzena i Rossa (1978), które na potrzeby niniejszego badania poddano autorskim modyfikacjom. W przeciwieństwie do oryginalnej propozycji Cattella, wskaźnik sabotowania obliczono na danych surowych, a nie znormalizowanych. Dodatkowo wyznaczono dokładnie współczynniki regresji poszczególnych połówek skal względem siebie, a nie posługiwano się uśrednionymi współczynnikami. Każdą ze skal NEO-FFI podzielono na dwie połowy, włączając do jednej z nich pozycje o numerach parzystych, a drugiej – nieparzystych. Obliczając wskaźnik sabotowania, dokonano regresji – dla każdej z pięciu skal NEO-FFI oddzielnie – w której część nieparzysta danej skali była zmienną zależną, a część parzysta – niezależną.

Wcześniej omówiony wskaźnik indywidualnego losowego wyniku oczekiwanego dotyczył jedynie pozycji o odpowiedziach binarnych. W celu wykorzystania go w omawianym badaniu został on rozszerzony na narzędzia składające się z pozycji o wielokategorialnym formacie odpowiedzi (zmodyfikowany indywidualny wynik oczekiwany, MFIC). W tym celu konieczne było obliczenie indywidualnych proporcji poszczególnych kategorii odpowiedzi w całym kwestionariuszu, a następnie przemnożenie tych proporcji przez liczbę pozycji punktowanych wprost oraz oddzielnie – przez liczbę pozycji punktowanych odwrotnie. W kolejnym kroku każdy z tych iloczynów był przemnożony przez liczbę punktów przyznawanych w danym sposobie punktowania. Ostateczny wynik uzyskuje się przez zsumowanie wszystkich iloczynów obliczonych dla poszczególnych kategorii oraz dla obu rodzajów punktowania. W skrócie przedstawia to następujący wzór:

$$MFIC_i = \sum_{j=1}^k w_j P_j T_{ij} + \sum_{j=1}^k z_j P_j N_{ij}$$

gdzie:

$MFIC_i$ – zmodyfikowany indywidualny wynik oczekiwany i -tej skali

w_j – liczba punktów przypisana odpowiedzi kategorii j przy kluczu polegającym na zliczaniu wprost

z_j – liczba punktów przypisana odpowiedzi kategorii j przy kluczu polegającym na zliczaniu przez inwersję

P_j – proporcja odpowiedzi kategorii j w całym kwestionariuszu udzielonych przez daną osobę niezależnie od skali

T_{ij} – liczba odpowiedzi kategorii j w i -tej skali, której przypisana jest liczba punktów w

N_{ij} – liczba odpowiedzi kategorii j w i -tej skali, której przypisana jest liczba punktów z

W celu przetestowania trafności wskaźników zostaną wykorzystane dane empiryczne uzyskane w warunkach neutralnych oraz dane kryterialne wygenerowane losowo. Założono, że odpowiedzi losowe na pytania NEO-FFI będą pochodzić z rozkładu równomiernego (zwanego też jednorodnym, jednostajnym lub prostokątnym), czyli takiego, w którym wszystkie wartości występują z jednakową częstością tzn. wylosowanie każdej z możliwych odpowiedzi będzie jednakowo prawdopodobne (Krysicki, Bartos, Dyczka, Królikowska i Wasilewski, 2007). Takie założenie przyjmuje większość badaczy zajmujących się symulowaniem komputerowym odpowiadania losowego (Karabatsos, 2003).

Metoda

W badaniach wykorzystano polską wersję Inwentarza Osobowości NEO-FFI Costy i McCrae (Zawadzki, Strelau, Szczepaniak i Śliwińska, 1998), który służy do oceny nasilenia pięciu cech osobowości: neurotyczności, ekstrawersji, otwartości na doświadczenia, ugodowości i sumienności. Poszczególne skale NEO-FFI cechują się akceptowalną dokładnością pomiaru (zgodnością wewnętrzną). W niniejszym badaniu uzyskano wartości wskaźników α Cronbacha mieszczące się w przedziale od 0,59 do 0,82.

W badaniu wzięło udział 996 osób, z czego do analizy zakwalifikowano wyniki 943 osób, które udzieliły pełnych odpowiedzi na wszystkie 60 pytań kwestionariusza NEO-FFI. W badaniach uczestniczyli studenci I roku studiów licencjackich na kierunku pielęgniarstwo, z 9 uczelni medycznych w kraju (Białystok, Bydgoszcz, Katowice, Kraków, Lublin, Łódź, Poznań, Warszawa i Wrocław), prowadzących kształcenie na poziomie licencjackim i magisterskim. Większość badanych była w wieku 19-21 lat (80,1%). W badaniu ogromną większość

(92,5%) stanowiły kobiety. Dane te zebrano przy okazji badań nad motywami podejmowania studiów pielęgniarstwa (Kądalska i Fronczyk, 2006). Danymi kryterialnymi w stosunku do tych danych empirycznych były wygenerowane z rozkładu jednostajnego dane symulujące odpowiedzi 1000 osób na poszczególne pozycje kwestionariusza NEO-FFI.

Aby zweryfikować możliwości diagnostyczne wskaźników, obliczono trzy równania regresji logistycznej, w których zmienną zależną była przynależność do danych losowych bądź danych empirycznych. W dwóch równaniach jako predyktor traktowano kolejne wskaźniki osobno, natomiast w trzecim – oba wskaźniki łącznie. Założono, że wartość wskaźnika, dla której prawdopodobieństwo przynależności do kategorii osób odpowiadających losowo wynosi przynajmniej 0,5, będzie świadczyła o zaklasyfikowaniu danej osoby badanej do tej kategorii sposobu odpowiadania. Dla każdego ze wskaźników osobno oraz dla obu wskaźników łącznie obliczono także: procent prawidłowych klasyfikacji, czułość, swoistość, dodatnią wartość predykcyjną (PPV – *positive predictive value*), ujemną wartość predykcyjną (NPV, *negative predictive value*), a także współczynnik zgodności Kappa Cohena. Przez czułość rozumie się prawdopodobieństwo, z jakim można stwierdzić, że osoba przejawiająca dany styl odpowiadania zostanie prawidłowo sklasyfikowana (trafne akceptacje). Specyficzność to prawdopodobieństwo, że osoba odpowiadająca zgodnie z samowiedzą nie zostanie zdiagnozowana jako przejawiająca dany styl (trafne odrzucenia). Wskaźnik NPV to prawdopodobieństwo, że wśród osób, które zostały zdiagnozowane jako udzielające odpowiedzi zgodnych z samowiedzą, są osoby, które rzeczywiście tak odpowiadają, natomiast wskaźnik PPV to prawdopodobieństwo odpowiadania losowego przez osobę, która została w ten sposób zdiagnozowana (Baer i in., 1997; Pinsoneault, 2002). Współczynnik Kappa Cohena opiera się na porównaniu zgodności zaklasyfikowania dokonanego na podstawie analizowanego wskaźnika oraz rzeczywistej przynależności do danych losowych lub empirycznych ze zgodnością w sytuacji, gdyby pomiędzy tymi oboma klasyfikacjami nie było żadnej zależności (badani zostaliby przydzieleni do jednej z kategorii w sposób losowy) (Zieliński, 2008). Dzięki tym obliczeniom zdecydowano, który ze wskaźników jest najbardziej trafny. Wyznaczono dla niego krzywą *Receiver Operating Characteristic* (ROC), która obrazuje relację pomiędzy prawdopodobieństwem fałszywego alarmu i prawdopodobieństwem trafnego zdiagnozowania odpowiadania losowego dla wszystkich możliwych wartości punktów odcięcia. Skonstruowana jest w następujący sposób: na osi pionowej przedstawiono prawdopodobieństwo trafnego zdiagnozowania odpowiadania losowego, natomiast oś pozioma ukazuje prawdopodobieństwo fałszywych alarmów. Moc klasyfikacyjną

testu określa obszar pod krzywą ROC. Maksymalna jego wartość liczbową wynosi 1, co oznacza trafną klasyfikację we wszystkich możliwych przypadkach.

Wszystkie obliczenia wykonano w środowisku obliczeniowym R, wykorzystując podstawowy zestaw pakietów. Jedynie do obliczenia wartości α Cronbacha zastosowano pakiet psych, a krzywą ROC wyznaczono dzięki kodowi podanemu przez Ćwika i Mielniczuka (2009). Kod generujący dane losowe przy zastosowaniu komendy: runif umieszczono w Załączniku 1. Natomiast w Załączniku 2 podano kod obliczeniowy obu wskaźników odpowiadania losowego.

Wyniki

Na wstępie zostaną przedstawione obliczenia związane z tworzeniem wskaźnika sabotowania. W Tabeli 1 przedstawiono wyniki pięciu analiz regresji dla połówek każdej ze skal. Jak widać, w przypadku wszystkich skal ich połówki są istotnie ze sobą związane. We wszystkich przypadkach wartość statystyki t wykorzystywana do oceny istotności współczynnika regresji (β) oraz odpowiadający jej poziom prawdopodobieństwa p potwierdzają, że współczynnik ten istotnie różni się od zera.

Tabela 1

Wyniki pięciu analiz regresji połówek skal NEO-FFI

	Współczynnik regresji	Błąd standardowy	t	p
Zmienna zależna: neurotyzm – pozycje nieparzyste				
Stała	4,22	0,31	13,60	< 0,005
neurotyzm – pozycje parzyste	0,69	0,03	27,33	< 0,005
Zmienna zależna: ugodowość – pozycje nieparzyste				
Stała	7,32	0,36	20,09	< 0,005
ugodowość – pozycje parzyste	0,54	0,02	22,37	< 0,005
Zmienna zależna: ekstrawersja – pozycje nieparzyste				
Stała	7,14	0,37	19,49	< 0,005
ekstrawersja – pozycje parzyste	0,55	0,02	23,05	< 0,005
Zmienna zależna: otwartość – pozycje nieparzyste				
Stała	9,98	0,45	22,25	< 0,005
otwartość – pozycje parzyste	0,39	0,04	11,08	< 0,005
Zmienna zależna: sumienność – pozycje nieparzyste				
Stała	5,88	0,39	15,20	< 0,005
sumienność – pozycje parzyste	0,65	0,02	27,53	< 0,005

W dalszym kroku dla każdej osoby obliczono reszty z wyżej opisanych regresji. Zsumowane kwadraty tych reszt dają wskaźnik sabotowania. Parametry równania regresji uzyskane w próbie osób badanych posłużyły do wyznaczenia wskaźnika sabotowania dla danych losowych. Wskaźnik indywidualnego losowego wyniku oczekiwanego obliczono według wcześniej opisanej procedury. Nie wymagał on estymacji żadnych parametrów na całości próby.

Wskaźnik sabotowania, jak i wskaźnik indywidualnego losowego wyniku oczekiwanego są względnie słabo ze sobą skorelowane. W grupie danych losowych ich korelacja nie jest istotna statystycznie ($r = 0,03$), natomiast w przypadku danych pochodzących od osób badanych korelacja ta jest wprawdzie istotna ($p < 0,001$), jednak jej wartość jest nieduża ($r = 0,19$), co przy tak licznej próbie trudno uznać za świadczące o jakimkolwiek rzeczywistym związku omawianych zmiennych.

W dalszej kolejności (Tabela 2) analizowano rozkład wyników obu wskaźników oraz różnice ich średnich ze względu na rodzaj danych (rzeczywiste bądź losowe). Uzyskane rozkłady znacznie odbiegają od rozkładu symetrycznego. Szczególnie zaznacza się to w danych rzeczywistych. Średnie obu wskaźników znacznie różnią się w obu grupach danych.

Tabela 2

Właściwości rozkładu wskaźników odpowiadania losowego

	Dane	<i>M</i>	<i>SD</i>	Skośność	Kurtoza	<i>W</i> Wilcoxona	<i>p</i>
Wskaźnik sabotowania	rzeczywiste	45,56	37,09	2,44	9,79	180379	< 0,005
	symulowane komputerowo	97,08	56,89	1,03	1,28		
Wskaźnik indywidualnego losowego wyniku oczekiwanego	rzeczywiste	759,44	200,20	1,02	0,98	431056	< 0,005
	symulowane komputerowo	530,67	87,60	0,49	0,31		

Kolejnym krokiem było wyznaczenie trzech równań regresji logistycznej, w których predyktorami były kolejno: wskaźnik sabotowania, wskaźnik indywidualnego losowego wyniku oczekiwanego oraz oba wskaźniki jednocześnie. Parametry tych równań zebrano w Tabelach 3-5.

Tabela 3

Parametry regresji logistycznej klasyfikującej odpowiadanie losowe na podstawie wskaźnika sabotowania

	Współczynnik regresji	Błąd standardowy	Test z	p	Iloraz szans
Stała	-1,70	0,1	-16,59	< 0,005	
Wskaźnik sabotowania	0,03	0,001	18,26	< 0,005	1,03

Tabela 4

Parametry regresji logistycznej klasyfikującej odpowiadanie losowe na podstawie wskaźnika indywidualnego losowego wyniku oczekiwanego

	Współczynnik regresji	Błąd standardowy	Test z	p	Iloraz szans
Stała	8,75	0,41	21,45	< 0,005	
Wskaźnik indywidualnego losowego wyniku oczekiwanego	-0,01	0,001	-20,98	< 0,005	0,99

Tabela 5

Parametry regresji logistycznej klasyfikującej odpowiadanie losowe na podstawie wskaźnika sabotowania oraz wskaźnika indywidualnego losowego wyniku oczekiwanego

	Współczynnik regresji	Błąd standardowy	Test z	p	Iloraz szans
Stała	7,08	0,45	15,83	< 0,005	
Wskaźnik indywidualnego losowego wyniku oczekiwanego	-0,01	0,001	-18,87	< 0,005	0,99
Wskaźnik sabotowania	0,03	0,001	15,37	< 0,005	1,03

Uzyskane wyniki wskazują, że im wyższa wartość wskaźnika sabotowania oraz im mniejsza wartość wskaźnika indywidualnego losowego wyniku oczekiwanego, tym większe prawdopodobieństwo, że dana osoba odpowiadała w sposób losowy. Wskazują na to dodatnie wartości parametrów dla wskaźnika sabotowania i ujemne dla wskaźnika indywidualnego losowego wyniku oczekiwanego. Taki wynik uzyskano, biorąc pod uwagę oba wskaźniki oddzielnie (Tabele 3 i 4), jak i razem (Tabela 5) przy wysokim poziomie istotności statystycznej. Do podobnych wniosków skłaniają także wartości ilorazów szans. Są one wyższe od jedności dla wskaźnika sabotowania i niższe od jedności dla wskaźnika indywidualnego losowego wyniku oczekiwanego. Różnice względem jedności nie są

zbyt duże, jednak nie przesądzają one o znaczeniu omawianych wskaźników. Po prostu oba wskaźniki osiągają dość duże wartości średnie oraz odchylenia standardowe, stąd przyrost wartości o jednostkę powiązany jest z małym przyrostem prawdopodobieństwa udzielania odpowiedzi losowych.

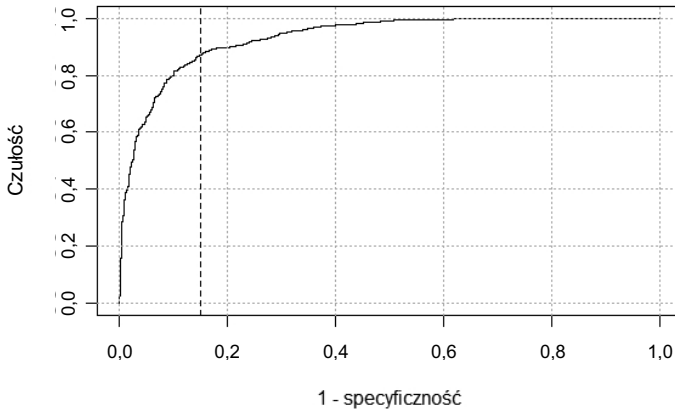
Wartości wskaźnika dopasowania modelu informują, że zarówno wykorzystanie jedynie wskaźnika sabotowania ($AIC = 1653,79$), jak i jedynie wskaźnika indywidualnego losowego wyniku oczekiwanego ($AIC = 2129,2$) daje gorsze dopasowanie niż jednoczesne użycie obu tych wskaźników ($AIC = 1279,57$). Trafność klasyfikacji opartej na każdym z wskaźników osobno, jak i na obu wskaźnikach łącznie oceniano dzięki obliczeniu szeregu miar poprawności klasyfikacji. Zaprezentowano je w Tabeli 6, z której wynika, że klasyfikacja oparta na obu wskaźnikach łącznie daje najlepsze rezultaty. Wskazują na to wartości miar jakości klasyfikacji. Ogółem 86% wszystkich obserwacji jest poprawnie zaklasyfikowanych. Porównując klasyfikację opracowaną na podstawie rozłączone użycie obu wskaźników trudno jest wskazać, który z nich byłby lepszy. Wskaźnik indywidualnego losowego wyniku oczekiwanego charakteryzuje się wprawdzie większą czułością, nieco lepszą ogólną liczbą prawidłowych klasyfikacji oraz lepszym wskaźnikiem NPV, ale daje niższy wskaźnik PPV oraz mniejszą swoistość niż wskaźnik sabotowania. Współczynnik Kappa oznacza, że wskaźnik sabotowania jest mniej trafny, jest to jednak wartość akceptowalna, choć niewysoka (Zieliński, 2008). Nieco wyższa wartość cechuje wskaźnik indywidualnego losowego wyniku oczekiwanego, choć dalej należy ona do przedziału wartości akceptowalnych. Łączne użycie obu wskaźników daje wartość Kappa z pogranicza wartości wysokich i akceptowalnych.

Tabela 6

Parametry trafności klasyfikacji odpowiadania losowego na podstawie różnych wskaźników

	Wskaźnik sabotowania	Wskaźnik indywidualnego losowego wyniku oczekiwanego	Wskaźnik sabotowania i wskaźnik indywidualnego losowego wyniku oczekiwanego
Czułość	0,69	0,84	0,87
Swoistość	0,79	0,73	0,85
PPV	0,78	0,77	0,86
NPV	0,71	0,81	0,86
% prawidłowych klasyfikacji	0,74	0,78	0,86
Kappa Cohena	0,48	0,57	0,72

Podsumowując, wydaje się, że najlepsze rezultaty w diagnozowaniu odpowiadania losowego daje łączne użycie wskaźnika sabotowania oraz wskaźnika indywidualnego losowego wyniku oczekiwanego. Dla tych dwóch wskaźników łącznie wyznaczono więc krzywą ROC (zob. Wykres).



Zastosowane wskaźniki mają dużą moc różnicującą dane losowe od rzeczywistych, co wynika z silnego wygięcia krzywej. Powierzchnia pod tą krzywą wynosi 0,93, więc jest bardzo duża. Na wykresie pionowa linia przerywana odpowiada specyficzności uzyskanej w badaniu dla obu wskaźników łącznie, czyli dla wartości 0,85. Wybrana wartość funkcji logistycznej (0,5), powyżej której dany wzorec odpowiadania zostaje zakwalifikowany jako losowy, wydaje się umiarkowanie rygorystyczna. Wartość specyficzności nie jest zbytnio wysoka względem wartości czułości.

DYSKUSJA

Oba wskaźniki, sabotowania oraz indywidualnego losowego wyniku oczekiwanego, mają bardzo skośne rozkłady, co widać szczególnie w przypadku danych pochodzących od rzeczywistych osób badanych. Jest to zrozumiałe, gdyż należy oczekiwać, że większość ludzi odpowiada w sposób spójny, nielosowy, a więc powinny przeważać wartości wskazujące na odpowiadanie zgodne z samowiedzą. Zastanawiające jest, dlaczego wskaźnik sabotowania ma także rozkład skośny, choć w mniejszym stopniu, w danych losowych symulowanych komputerowo. W tym przypadku należałoby oczekiwać rozkładu symetrycznego

czy może skośnego, ale w drugą stronę, ponieważ wyższe wyniki wskazują na odpowiadanie losowe. Być może taki rezultat podważa nieco trafność wskaźnika sabotowania. Niemniej jednak różnice między średnimi obu wskaźników pochodzącymi z danych rzeczywistych i z danych symulowanych osiągnęły bardzo wysoki poziom istotności statystycznej.

Uzyskane wyniki wskazują, że zastosowanie wyłącznie wskaźnika sabotowania Cattella i współpracowników (1970) albo wyłącznie wskaźnika indywidualnego losowego wyniku oczekiwanego Haertzena i Rossa (1978) daje umiarkowane możliwości identyfikowania osób udzielających odpowiedzi losowe. Dopiero łączne posłużenie się tymi obydwoma wskaźnikami daje dużo lepsze rezultaty. Dzieje się tak prawdopodobnie dlatego, że oba wskaźniki ujmują nieco inny aspekt losowości danych, na co wskazuje ich słabe skorelowanie. Wskaźnik sabotowania jest wrażliwy bardziej na brak wewnętrznej spójności odpowiedzi w ramach każdej ze skal, natomiast wskaźnik indywidualnego losowego wyniku oczekiwanego dotyczy proporcji poszczególnych kategorii odpowiedzi z uwzględnieniem indywidualnych preferencji poszczególnych odpowiedzi.

W innych badaniach wykazano, że jednoczesne użycie kilku wskaźników daje lepsze wyniki niż zastosowanie jednego wskaźnika, jeśli chodzi o poprawność klasyfikacji stylów odpowiadania lub odpowiadania losowego. Prawdliwość tę wykazano w przypadku kwestionariusza MMPI (Archer i Elkins, 1999; Baer, Kroll, Rinaldo i Ballenger, 1999). Wynik otrzymany w niniejszym badaniu nie jest więc czymś wyjątkowym. Może on odzwierciedlać ogólną prawidłowość dotyczącą trudności w odróżnianiu odpowiadania zgodnego z samowiedzą od udzielania odpowiedzi losowych.

W badaniu uzyskano dość wysoką proporcję prawidłowych klasyfikacji. Można by oczekiwać jeszcze większej proporcji, jednak zarówno dane rzeczywiste, jak i dane symulowane komputerowo nie były w pełni „czyste”. Otóż wśród osób badanych mogły zdarzyć się takie, które z pewnych powodów odpowiadały w sposób losowy. Natomiast wśród danych wygenerowanych komputerowo mogły zdarzyć się takie konfiguracje odpowiedzi, które – przez zwykły przypadek – były kombinacjami, które mogłyby być odpowiedziami rzeczywistych osób badanych. Ta częściowa (nawet jeśli dotyczyło to tylko niewielkiej części danych) niejednoznaczność mogła przyczynić się do błędnych klasyfikacji.

Jak wspomniano we wprowadzeniu, wskaźnik sabotowania byłby prawdopodobnie bardziej trafną metodą identyfikacji osób odpowiadających losowo w przypadku kwestionariuszy o wyższych wskaźnikach zgodności wewnętrznej. Z badań symulacyjnych wiadomo bowiem, że odpowiadanie losowe obniża zgodność wewnętrzną kwestionariuszy (Fong, Ho i Lam, 2010). Można ponadto

przypuszczać, że w inwentarzach składających się z większej liczby skal wskaźnik ten daje lepsze możliwości różnicowania osób udzielających odpowiedzi losowych od odpowiadających zgodnie z samowiedzą. Im większa liczba skal, tym mniejsze prawdopodobieństwo pojawienia się wielu przypadkowych rozbieżności wśród odpowiedzi nielosowych i błędnej ich kwalifikacji jako losowych. Podobnie większa liczba skal oraz pozycji stwarza lepsze możliwości identyfikowania odpowiadania losowego względem odpowiadania zgodnego z samowiedzą w przypadku wskaźnika indywidualnego losowego wyniku oczekiwanego.

Jak możemy stwierdzić na podstawie podręcznika testowego (Zawadzki i in., 1998) do polskiej adaptacji zastosowanego w niniejszym badaniu kwestionariusza NEO-FFI, zgodność wewnątrzna skal tego kwestionariusza waha się w przedziale od 0,68 do 0,82. Dość zbliżone wartości uzyskano w prezentowanym badaniu. Nie są to wartości zbyt wysokie. Wraz z niezbyt dużą liczbą pięciu skal mogło to się przyczynić do uzyskania umiarkowanie dobrego rezultatu diagnozowania odpowiadania losowego.

Najważniejszym rezultatem badania jest wykazanie możliwości różnicowania odpowiadania losowego od odpowiadania zgodnego z samowiedzą w kwestionariuszach osobowości. Co więcej, przedstawione dane umożliwiają takie diagnozowanie w praktyce psychologicznej. Wprawdzie byłoby to związane z kłopotliwymi obliczeniami, ale wykorzystanie arkusza kalkulacyjnego znacznie je uprości.

Praktyczne wykorzystanie wyników badania wymaga jednak dodatkowej weryfikacji. Wartości punktów odjęcia wyznaczanych w różnych badaniach bardzo często nie są sobie równe. Wynika to ze specyfiki próby osób badanych oraz specyfiki symulowania danych losowych. W celu praktycznego wdrożenia omawianych wskaźników konieczne byłoby przeprowadzenie dodatkowych badań w planie walidacji krzyżowej lub zastosowanie bootstrappingu, polegającego na pobieraniu wielokrotnie podprób spośród osób badanych i weryfikowaniu wskaźników odpowiadania losowego. Uśredniony wynik tych wielokrotnych analiz jest wynikiem ostatecznym.

Prezentowane badanie ma pewne ograniczenia. Po pierwsze, w badaniu wykorzystano dane losowe pochodzące jedynie z rozkładu równomiernego. W rzeczywistości nie jest takie pewne, czy każda alternatywa wybierana jest z jednakowym prawdopodobieństwem. Jest to pewne uproszczenie, gdyż przy odpowiadaniu czysto losowym może wystąpić preferencja którejś z kategorii odpowiedzi (van Ijzendoorn, 1984).

Inne ograniczenie może być związane ze specyfiką badanych osób. Jak wspomniano, byli nimi studenci pielęgniarstwa, w większości kobiety. Nie jest wykluczone, że ludzie młodzi oraz kobiety mają większą motywację do rzetelnego wypełniania kwestionariuszy, niż ma to miejsce w populacji generalnej. Mogło to owocować mniejszym odsetkiem nieuważnie wypełnionych kwestionariuszy wśród rzeczywistych danych, a przez to podnosić wykrywalność odpowiadania losowego.

Podsumowując należy podkreślić bardzo obiecujące możliwości identyfikowania odpowiadania losowego bądź nieuważnego na podstawie dość rzadko wykorzystywanych wskaźników sabotowania Cattella oraz indywidualnego losowego wyniku oczekiwanego. Kolejne badania z wykorzystaniem innych kwestionariuszy, prowadzone na bardziej zróżnicowanych próbach badanych, używanych w różnych sytuacjach oraz przy zastosowaniu danych kryterialnych wygenerowanych nie tylko z rozkładu jednostajnego, powinny przyczynić się do dalszej weryfikacji możliwości diagnostycznych obu wskaźników.

LITERATURA CYTOWANA

- Archer, R. P. i Elkins, D. E. (1999). Identification of random responding on the MMPI-A. *Journal of Personality Assessment*, 73, 407-421.
- Baer, R. A., Ballenger, J., Berry, D. T. R. i Wetter, M. W. (1997). Detection of random responding on the MMPI-A. *Journal of Personality Assessment*, 68, 139-151.
- Baer, R. A., Kroll, L. S., Rinaldo, J. i Ballenger, J. (1999). Detecting and discriminating between random responding and overreporting on the MMPI-A. *Journal of Personality Assessment*, 72, 308-320.
- Beach, D. A. (1989). Identifying the random responder. *The Journal of Psychology*, 123, 101-103.
- Cattell, R. B., Eber, H. W. i Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire (16PF)*. Champaign, IL: IPAT.
- Ćwik, J. i Mielniczuk, J. (2009). *Statystyczne systemy uczące się. Ćwiczenia w oparciu o pakiet R*. Warszawa: Oficyna Wydawnicza PW.
- Evans, R. G. i Dinning, W. D. (1983). Response consistency among high F scale scorers on the MMPI. *Journal of Clinical Psychology*, 39, 246-248.
- Fong, D. Y. T., Ho, S. Y. i Lam, T. H. (2010). Evaluation of internal reliability in the presence of inconsistent responses. *Health and Quality of Life Outcomes*, 8, 27.
- Haertzen, C. A. i Ross, F. E. (1978). Using four chance profiles to estimate carelessness. *Psychological Reports*, 41, 1079-1087.
- Irvine, M. J. i Gendreau, P. (1974). Detection of the fake 'good' and 'bad' response on the sixteen personality factor inventory in prisoners and college students. *Journal of Consulting and Clinical Psychology*, 42, 465-466.
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 16, 277-298.

- Kądalska, E. i Fronczyk, K. (2006). Motywy wyboru studiów licencjackich na kierunku pielęgniarstwo w Polsce. *Pielęgniarstwo XXI wieku, 1-2*, 111-115.
- Krysicki, W., Bartos, J., Dyczka, W., Królikowska, K. i Wasilewski, M. (2007). *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*. Warszawa: Wydawnictwo Naukowe PWN.
- Nichols, D. S. i Greene, R. L. (1997). Dimensions of deception in personality assessment: The example of the MMPI-2. *Journal of Personality Assessment, 68*, 251-266.
- Nichols, D., Greene R. i Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology, 45*, 239-250.
- O'Dell, J. W. (1971). Method for detecting random answers on personality questionnaires. *Journal of Applied Psychology, 55*, 380-383.
- Paluchowski, W. J. (1983). Źródła zakłóceń w kwestionariuszowym badaniu osobowości i ich kontrola W: W. J. Paluchowski (red.), *Z zagadnień diagnostyki osobowości* (s. 249-271). Wrocław: Zakład Narodowy im. Ossolińskich.
- Pinsonneault, T. B. (2002). A variable response inconsistency scale and a true response inconsistency scale for the Millon Adolescent Clinical Inventory. *Psychological Assessment, 14*, 320-330.
- Pinsonneault, T. B. (2005). Detecting random, partially random, and nonrandom Minnesota Multiphasic Personality Inventory-Adolescent protocols. *Psychological Assessment, 17*, 476-480.
- Ross, F. E. i Haertzen, C. A. (1979). The use of chance profiles for detecting carelessness: The effect of determining the true response rate from items in scales. *Journal of Psychology, 101*, 27-35.
- Van Ijzendoorn, M. H. (1984). Answers without questions: A note on response style in questionnaires. *Perceptual and Motor Skills, 59*, 827-831.
- Zawadzki, B. (2006). *Kwestionariusze osobowości: strategie i procedura konstruowania*. Warszawa: Scholar.
- Zawadzki, B., Strelau, J., Szczepaniak, P. i Śliwińska, M. (1998). *Inwentarz Osobowości NEO-FFI Costy i McCrae. Adaptacja polska. Podręcznik*. Warszawa: Pracownia Testów Psychologicznych PTP.
- Zieliński, A. (2008). Błąd klasyfikacji w badaniach epidemiologicznych. *Przegląd Epidemiologiczny, 62(2)*, 461-470.

ZAŁĄCZNIK 1

KOD PROGRAMU R GENERUJĄCY DANE LOSOWE
KWESTIONARIUSZA NEO-FFI

```
dane.losowe<-matrix(NA,1000, 60)
for (i in 1:60)
{
dane.losowe[,i]<-runif(1000,0,1)
}
odpowiedzi.losowe<-
ifel-
se(dane.losowe<=1/5,0,ifelse(dane.losowe<=2/5,1,ifelse(dane.losowe<=3/5,2,ifelse(dane.losowe<=
4/5,3,4))))
```

ZAŁĄCZNIK 2

KOD PROGRAMU R OBLICZANIA WSKAŹNIKA SABOTOWANIA
I INDYWIDUALNEGO LOSOWEGO WYNIKU OCZEKIWANEGO

```
lm.neu<-lm(neu_nparz~neu_parz)
lm.ekstr<-lm(ekstr_nparz~ekstr_parz)
lm.otw<-lm(otw_nparz~otw_parz)
lm.sum<-lm(sum_nparz~sum_parz)
lm.ugo<-lm(ugo_nparz~ugo_parz)
sabotowanie<-
lm.neu$residuals^2+lm.ekstr$residuals^2+lm.otw$residuals^2+lm.sum$residuals^2+lm.u
go$residuals^2
zlicz1<-rowSums(NEO_FFI== 1)/60
zlicz2<-rowSums(NEO_FFI== 2)/60
zlicz3<-rowSums(NEO_FFI== 3)/60
zlicz4<-rowSums(NEO_FFI== 4)/60
zlicz5<-rowSums(NEO_FFI== 5)/60
FIC_E=8*(zlicz2+2*zlicz3+3*zlicz4+4*zlicz5)+4*(4*zlicz1+3*zlicz2+2*zlicz3+zlicz4)
FIC_N=8*(zlicz2+2*zlicz3+3*zlicz4+4*zlicz5)+4*(4*zlicz1+3*zlicz2+2*zlicz3+zlicz4)
FIC_O=5*(zlicz2+2*zlicz3+3*zlicz4+4*zlicz5)+7*(4*zlicz1+3*zlicz2+2*zlicz3+zlicz4)
FIC_S=8*(zlicz2+2*zlicz3+3*zlicz4+4*zlicz5)+4*(4*zlicz1+3*zlicz2+2*zlicz3+zlicz4)
FIC_U=4*(zlicz2+2*zlicz3+3*zlicz4+4*zlicz5)+8*(4*zlicz1+3*zlicz2+2*zlicz3+zlicz4)
D=(neu-FIC_N)^2+(ekstr-FIC_E)^2+(otw-FIC_O)^2+(sum-FIC_S)^2+(ugo-FIC_U)^2
```