

OMEGA MCDONALDA JAKO ALTERNATYWA DLA ALFA CRONBACHA W SZACOWANIU RZETELNOŚCI TESTU

Barbara Ciżkowicz

Institut Psychologii, Uniwersytet Kazimierza Wielkiego w Bydgoszczy
Institute of Psychology, Kazimierz Wielki University in Bydgoszcz

MCDONALD'S OMEGA AS AN ALTERNATIVE TO CRONBACH'S ALPHA IN SCALE SCORE RELIABILITY ESTIMATION

Summary. This article presents the advantages and disadvantages of Cronbach's alpha and McDonald's omega as the reliability coefficients. The article also shows an example a method of calculating omega using SPSS. Use of alpha requires, apart from Classical Test Theory, additionally the unidimensionality of the scale and meeting the conditions of the *essentially tau-equivalent model*. Failure to meet these assumptions makes the value of the coefficient alpha biased estimator of reliability. It is emphasized that use of alpha should be preceded by factor analysis. Particularly in the case of multidimensionality scale, it is proposed to use other coefficient such as, inter alia omega, which use requires less restrictive assumptions. Existing theoretical analysis and simulation studies reveal the consequences of non-compliance with the assumptions while estimating reliability.

Key words: reliability, multidimensionality, bifactor model, McDonald's omega, Cronbach's alpha

Wprowadzenie

Ważnym pojęciem w psychometrii i teorii testu jest rzetelność wyników testu. W Klasycznej Teorii Testu (KTT) rzetelność jest zdefiniowana jako stosunek wariancji wyników prawdziwych do wariancji wyników otrzymanych w teście (Brzeziński, 1996; Niemierko, 1999; Hornowska, 2001). Ponieważ wariancji wyniku prawdziwego nie można bezpośrednio obserwować, rzetelność musi być estymowana. Najczęściej test przeprowadzany jest jednokrotnie. W tym przypadku standardowym podejściem do szacowania rzetelności jest metoda wewnętrznej zgodności, a powszechnie stosowanym współczynnikiem w tej metodzie jest współczynnik alfa (α ; Cronbach,

Adres do korespondencji: Barbara Ciżkowicz, e-mail: cizbar@ukw.edu.pl

1951; Cortina, 1993; Osburn, 2000; Sijtsma, 2009). Został on zaproponowany przez Guttmana (1945) jako λ_3 , a później spopularyzowany przez Cronbacha (1951) jako α (Revelle, Zinbarg, 2009). Z uwagi na powszechność stosowania współczynnika α Cronbacha poświęcono liczne opracowania psychometryczne (Woodward, Bentler, 1979; Crocker, Algina, 1986; Zimmerman, Zumbo, Lalonde, 1993; Raykov, 1997b; Revelle, Zinbarg, 2009; Dunn, Baguley, Brunsten, 2014; Gignac, 2014; Bonett, Wright, 2015; Green, Yang, 2015). W wielu artykułach przedstawiane są analizy wskazujące kiedy współczynnik α powinien, a kiedy nie powinien być stosowany (Cortina, 1993; Schmitt, 1996; Osburn, 2000; Peters, 2014). Wynika z nich, że α Cronbacha jest użytecznym¹ współczynnikiem, ale posiada pewne ograniczenia. Najczęściej α zarzuca się, że nie bada wewnętrznej spójności testu (Schmitt, 1996; Graham, 2006; Sijtsma, 2009a), w sposób punktowy szacuje wartość rzetelności bez podawania przedziałów ufności (Kelley, Pornprasertmanit, 2016) oraz powoduje znaczne niedoszacowanie rzetelności, gdy niespełnione są założenia niezbędne do jej stosowania (Graham, 2006; Sijtsma, 2009a). Dlatego też proponowane są inne współczynniki rzetelności, wśród których wymieniana jest omega McDonalda (ω ; Sijtsma, 2009a, 2009b). Dotyczy to szczególnie testów wielowymiarowych (Schmitt, 1996; Osburn, 2000).

Rzetelność w KTT – podstawowe założenia

W KTT przyjęto założenie, że wynik otrzymany testu (X) jest sumą wyniku prawdziwego (T) i błędu pomiaru (E): $X = T + E$. Ponadto w KTT zakłada się, że błędy pomiaru mają rozkład normalny, są losowe: ($M_E = 0$) i nieskorelowane: ($r_{E_iE_j} = 0$), nieskorelowane są również wyniki prawdziwe z błędami pomiaru: ($r_{TE_i} = 0$) oraz wyniki pomiaru wyrażone są na skali co najmniej interwałowej.

Rzetelność (ρ_{xx}), która informuje o stopniu w jakim wynik danego testu odzwierciedla poziom mierzonej cechy (Hornowska, 2001), w KTT została zdefiniowana jako stosunek wariancja wyniku prawdziwego ($Var(T)$) do wariancji wyniku otrzymanego ($Var(X)$; Lord, Novick, 1968). Innymi słowy, jest to część całkowitej zmienności zawartej w odpowiedziach badanych, która jest przypisana do cechy mierzonej (Dunn, Baguley, Brunsten, 2014). Zasadniczo koncepcja rzetelności zakłada, że pozycje testu są jednowymiarowe, a naruszenie tego założenia powoduje znaczne niedoszacowanie rzetelności (McDonald, 1999).

Na gruncie KTT opracowano szereg metod empirycznego szacowania rzetelności takich, jak test-retest, form alternatywnych, zgodności wewnętrznej (współczynnik α i ω). Wszystkie one wymagają spełnienia założeń KTT. Współczynnik α jest szczególnym przypadkiem współczynnika ω . Jego stosowanie wymaga do-

¹ Liczba cytowań artykułu Cronbacha (1951) według Google Scholar w dniu 4.01.2017 r. sięgnęła 30 978.

datkowo spełnienia założenia dotyczącego liniowego związku między wynikami prawdziwymi pozycji ($T_i = T_j + a_{ij}$) dla wszystkich par pozycji, gdzie a_{ij} jest stałą, dla konkretnej pary pozycji. Spełnienie tego założenia zapewnia równość wariancji wyników prawdziwych wszystkich pozycji testu.

Zamieszczone poniżej rozważania odnoszą się do szacowania rzetelności metodą zgodności wewnętrznej. Przedstawione w nich porównanie współczynników α i ω ma uzasadnić sens równoczesnego stosowania tych współczynników. W tym celu wskazano, wynikające z analiz teoretycznych oraz badań empirycznych, konsekwencje niespełniania założeń dla relacji między α , ω i ρ_{xx} oraz zwrócono uwagę na znaczenie struktury testu w szacowaniu rzetelności. Przedstawiono też sposób obliczania ω na podstawie wyników analizy czynnikowej.

Szacowanie rzetelności w obrębie analizy czynnikowej

W KTT szacowanie rzetelności metodą zgodności wewnętrznej może odbywać się z zastosowaniem analizy czynnikowej (Terry, Kelly, 2012). W takim przypadku identyfikacja modeli pomiarowych wymaga przyjęcia określonych założeń. Do szacowania rzetelności stosowane są modele pomiarowe, które różnią się restrykcyjnością założeń. Do najczęściej wykorzystywanych modeli można zaliczyć modele: równoległy, τ -równoważny, zasadniczo τ -równoważny i generyczny (Dunn, Baguley, Brunnsden, 2014). Poniżej przedstawiono ich krótką charakterystykę.

Model **równoległy** (*parallel*) jest najbardziej restrykcyjnym modelem pomiarowym stosowanym w określaniu składowych wyniku prawdziwego. Zakłada on, że wszystkie pozycje testu są w badaniu dokładnie sobie równoważne, co oznacza, że muszą mierzyć tę samą zmienną (jednowymiarowość), na tej samej skali, z tą samą dokładnością i z taką samą wariancją błędów (Raykov, 1997a, 1997b). Model równoległy musi więc spełniać następujące warunki:

- a) ładunki czynnikowe dla wszystkich pozycji testu są sobie równe $\lambda_1 = \lambda_2 = \dots = \lambda_k$
- b) dodatkowo równe są wariancje błędów $\text{Var}(e_1) = \text{Var}(e_2) = \dots = \text{Var}(e_k)$

W modelu **τ -równoważnym** (*tau-equivalent*) utrzymane są warunki dotyczące wyniku prawdziwego (równe ładunki czynnikowe), jednak wariancje błędów mogą być różne dla różnych pozycji. Jeszcze mniej wymagające założenia występują w modelu **zasadniczo τ -równoważnym** (*essentially tau-equivalent*). Oprócz różnych wariancji błędów dopuszcza się w nim wyniki prawdziwe różniące się o stałą addytywną (dodawaną), czyli jeżeli zapiszemy ładunek czynnikowy i -tej pozycji testu jako funkcję liniową j -tej pozycji ($\lambda_i = b_i + c_i \lambda_j$), to stałe addytywne w tym równaniu (b_i) mogą mieć wartości różne od zera, a stałe multiplikatywne (mnożone; c_i) muszą mieć wartość 1. Dopuszczenie różnic wyników prawdziwych tylko o stałą addytywną gwarantuje nadal stałość wariancji wyników prawdziwych dla różnych pozycji testu.

Założenia modelu **generycznego** (*congeneric model*) są najmniej restrykcyjne. Oprócz różnic w wariancjach błędów dopuszczalne są różne wartości ładunków czynnikowych pozycji testu zarówno co do stałej addytywnej (b_i), jak i stałej multiplikatywnej (c_i). Tak więc model generyczny jest najbardziej ogólnym modelem stosowanym w szacowaniu rzetelności. Wcześniej omówione modele są jego szczególnymi przypadkami powstającymi poprzez wprowadzanie dodatkowych ograniczeń (Graham, 2006; Dunn, Baguley, Brunnsden, 2014).

Do szacowania rzetelności mogą być stosowane różne współczynniki w zależności od założeń spełnianych przez wyniki testu. Stąd w modelu równoległym można szacować rzetelność używając wzoru proroczego Spearmana-Brauna, w którym rzetelność testu jest rzetelnością pojedynczej pozycji (identyczną dla wszystkich pozycji testu), a krotność zmiany długości testu jest liczbą pozycji w teście (McDonald, 1999, s. 123-124). W modelu zasadniczo τ -równoważnym do szacowania rzetelności używany jest współczynnik α Cronbacha, a w modelu generycznym – współczynnik ω McDonalda (Graham, 2006; Green, Yang, 2015).

α Cronbacha w modelu analizy czynnikowej

Przyjęte w KTT założenie, że wynik otrzymany (x_i) dla i -tej pozycji testu jest sumą dwóch nieobserwowalnych bezpośrednio składowych: wyniku prawdziwego (t_i) i błędu pomiaru (e_i), można zapisać (Lord, Novick, 1968): $x_i = t_i + e_i$.

W modelu analizy czynnikowej, wynik otrzymany można przedstawić jako funkcję ładunków czynnikowych i reszt (Geen, 2015):

$$x_i = a_i + \lambda_i f + e_i, \quad \text{co oznacza, że} \quad t_i = a_i + \lambda_i f$$

gdzie: a_i – stała dla i -tej pozycji; f – wynik dla danego czynnika; λ_i – ładunek czynnikowy na danym czynniku dla i -tej pozycji; e_i – reszta dla i -tej pozycji.

Jeśli model zasadniczo τ -równoważny jest poprawny, wówczas na podstawie ładunków czynnikowych można obliczyć α Cronbacha² według poniższej zależności (McDonald, 1999; Green, Yang, 2015):

$$\alpha = k^2 \lambda^2 / \sigma^2(X) \tag{1}$$

gdzie: k – liczba pozycji testu; λ – ładunek czynnikowy; $\sigma^2(X)$ – wariancja wyników testu ω **McDonalda**.

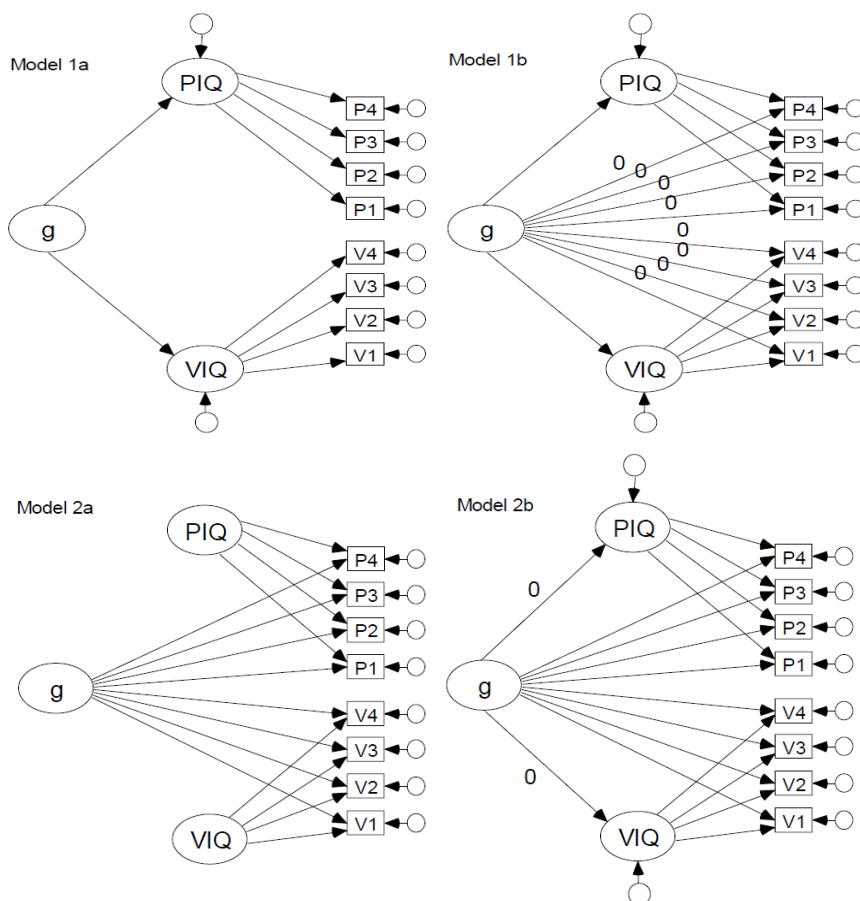
² McDonald (1999) przedstawia dokładne związki kowariancji i wariancji pozycji testu z ładunkami czynnikowymi i wariancją specyficzną w modelu czynnikowym (s. 108) oraz ich uproszczenia związane z przyjęciem założeń modelu zasadniczo τ -równoważnego (s. 115, 121-122).

Zaproponowany przez McDonalda (1978, 1999) współczynnik ω jest szacowaniem rzetelności na podstawie analizy czynnikowej w modelu generycznym.

Dla testu jednoczynnikowego ω stanowi stosunek wariancji przypadającej na daną cechę mierzoną do wariancji wyników obserwowanych (McDonald, 1985; Graham, 2006; Brunner, Nagy, Wilhelm, 2012; Terry, Kelly, 2012).

$$\omega = \frac{(\sum_{i=1}^k \lambda_i)^2}{(\sum_{i=1}^k \lambda_i)^2 + \sum_{i=1}^k (1 - h_i^2)} \quad (2)$$

gdzie: λ_i – standaryzowany ładunek czynnikowy i -tej pozycji; k – liczba pozycji testu; $(1-h_i^2)$ – standaryzowana wariancja błędów (Gignac, 2014).



Rycina 1. Modele hierarchiczne czynnika wyższego rzędu (modele 1a i 1b) i podwójnego czynnika (modele 2a i 2b); g – czynnik ogólny; PIQ i VIQ – podczynniki (Gignac, 2008)

Do obliczania ω w testach heterogenicznych używane są modele czynnika wyższego rzędu (*higher-order models*; rycina 1, model 1a i b) lub modele podwójnego czynnika (*bifactor model*; rycina 1, model 2a i b), które różnią się założeniami dotyczącymi powiązań między czynnikami (Rindskopf, Rose, 1988; Gignac, 2008; Grygier, 2016). W modelu czynnika wyższego rzędu związek czynnika ogólnego z obserwowalnymi wskaźnikami zachodzi tylko poprzez czynniki niższego rzędu. Model 1a (rycina 1) przedstawia pośrednie powiązanie czynnika drugiego rzędu ze wskaźnikami poprzez czynniki pierwszego rzędu. W modelu 1b zaznaczone zostały powiązania bezpośrednie, jednak ich ładunki czynnikowe ustalono na 0. Tak więc jest on równoważny modelowi 1a. W modelu podwójnego czynnika jest odwrotnie. Związek czynnika ogólnego ze wskaźnikami zachodzi tylko w relacji bezpośredniej bez udziału pośredniczącej roli czynników pierwszego rzędu (model 2a i b). Każda pozycja testu jest wskaźnikiem czynnika ogólnego oraz dodatkowo wiąże się (ładuje) ze specyficzną podskalą, do której dana pozycja należy (Gibbons, Hedeker, 1992).

Dla testów heterogenicznych współczynnik ω został rozszerzony o możliwość oceny rzetelności całego testu (ω_t), nasycenia wyników testu czynnikiem ogólnym (ω_h hierarchiczna) oraz rzetelności podczynników (ω_s) przy kontroli czynnika ogólnego. W tym kontekście ω można zapisać (McDonald, 1999; Zinbarg i in., 2005; Gignac, 2014; Green, Yang, 2015):

$$\omega_t = \frac{(\sum_{i=1}^k \lambda_{gi})^2 + (\sum_{i=1}^{s_1} \lambda_{s1i})^2 + (\sum_{i=1}^{s_2} \lambda_{s2i})^2 + \dots + (\sum_{i=1}^{s_p} \lambda_{spi})^2}{(\sum_{i=1}^k \lambda_{gi})^2 + (\sum_{i=1}^{s_1} \lambda_{s1i})^2 + (\sum_{i=1}^{s_2} \lambda_{s2i})^2 + \dots + (\sum_{i=1}^{s_p} \lambda_{spi})^2 + \sum_{i=1}^k (1 - h_i^2)} \quad (3)$$

$$\omega_h = \frac{(\sum_{i=1}^k \lambda_{gi})^2}{(\sum_{i=1}^k \lambda_{gi})^2 + (\sum_{i=1}^{s_1} \lambda_{s1i})^2 + (\sum_{i=1}^{s_2} \lambda_{s2i})^2 + \dots + (\sum_{i=1}^{s_p} \lambda_{spi})^2 + \sum_{i=1}^k (1 - h_i^2)} \quad (4)$$

Analogicznie do równania (3) i (4) można dokonać obliczeń dla podskal (Rodriguez, Reise, Haviland, 2016):

$$\omega_{t,s1} = \frac{(\sum_{i=1}^{s_1} \lambda_{gi})^2 + (\sum_{i=1}^{s_1} \lambda_{s1i})^2}{(\sum_{i=1}^{s_1} \lambda_{gi})^2 + (\sum_{i=1}^{s_1} \lambda_{s1i})^2 + \sum_{i=1}^{s_1} (1 - h_i^2)} \quad (5)$$

$$\omega_{h,1} = \frac{(\sum_{i=1}^{s_1} \lambda_{s1i})^2}{(\sum_{i=1}^{s_1} \lambda_{gi})^2 + (\sum_{i=1}^{s_1} \lambda_{s1i})^2 + \sum_{i=1}^{s_1} (1 - h_i^2)} \quad (6)$$

gdzie: λ_g – standaryzowane ładunki czynnika ogólnego; $\lambda_{s1}, \lambda_{s2}, \dots, \lambda_{sp}$ – standaryzowane ładunki podczynników przy kontroli czynnika ogólnego; $1-h_i^2$ – standaryzowana wariancja błędów; k – liczba pozycji testu; s_1, \dots, s_p – liczba pozycji w $1, \dots, p$ podczynniku (podskali).

Podstawową ideą ω jest oszacowanie odsetka całkowitej wariancji otrzymanych wyników testu przypisanej do czynnika ogólnego i wszystkich podczynników (ω_t) oraz odsetka całkowitej wariancji wyników testu związanej tylko z czynnikiem ogólnym (ω_h). Tak więc ω_h informuje o nasyceniu wyników testu czynnikiem ogólnym. Ponadto równania (5) i (6) pozwalają obliczyć rzetelność każdej podskali (ω_s) przy

uwzględnieniu czynnika ogólnego. Należy zwrócić uwagę, że w tym przypadku wszystkie składniki w równaniach obejmują tylko pozycje należące do danej podskali.

Relacje między α Cronbacha a ω McDonalda – analiza teoretyczna

Analiza teoretyczna przeprowadzona przez Zinbarga i współpracowników (2005) pozwoliła określić relacje między rzetelnością prawdziwą (ρ) oraz współczynnikami α , ω_t i ω_h . Poddali oni analizie cztery przypadki różniące się spełnianiem warunku homogeniczności testu (homogeniczny *vs* heterogeniczny) oraz równości ładunków czynnika ogólnego (wymagana tylko dla α). Wyniki przedstawiono w tabeli 1.

Tabela 1. Relacje między rzetelnością prawdziwą (ρ) a współczynnikami α , ω_t i ω_h względem spełniania założeń – analiza teoretyczna (Zinbarg i in., 2005, s. 128)

Test	Nierówne ładunki czynnika ogólnego (model generyczny)	Równe ładunki czynnika ogólnego (zasadniczo τ -równoważny)
	przypadek I	przypadek II
Wielowymiarowy	$\alpha < \omega_t \leq \rho$ $\omega_h < \omega_t \leq \rho$	$\omega_h < \alpha < \omega_t \leq \rho$
	przypadek III	przypadek IV
Jednowymiarowy	$\alpha < \omega_h = \omega_t \leq \rho$	$\alpha = \omega_h = \omega_t \leq \rho$

Testy jednowymiarowe w modelu zasadniczo τ -równoważnym oznaczają spełnienie założeń zarówno dla α , jak i ω (przypadek IV, tabela 1). W efekcie, bez względu na zastosowany współczynnik, ich wartości są sobie równe i stanowią dolną granicę oszacowania rzetelności (ρ). Naruszenie założeń α (przypadek III) lub α i ω (przypadek I i II) powoduje, w każdym przypadku, niższe oszacowanie rzetelności przez współczynnik α niż przez ω_t . Podkreślane jest niedoszacowanie rzetelności przez α szczególnie w przypadku testów heterogenicznych (I i II), dla których ω_t daje zawsze wyższą wartość rzetelności niż α . Jednak wnioskowanie o mierzonej zmiennej na podstawie wyników testu wymaga również zagwarantowania odpowiedniego poziomu ω_h .

Przyjmuje się, że ω_h , tak jak α , powinno osiągać wartości między 0,70 a 0,95 (Moumni i in., 2016, s. 910). Niestety, nie ma jasnych wytycznych dla dopuszczalności (punktów odcięcia) ω_h i ω_s (Reise i in., 2013b). Podkreśla się, że jeśli głównym celem stosowania testu jest pomiar jednego, ogólnego konstruktów należy maksymalizować ω_h . Jeśli zaś wyniki podskal są bardziej przydatne do celów badawczych, należy zadbać o wysoką wartość ω_s (Green, Yang, 2015). Reise i współpracownicy (2013a, s. 137)

proponują dla ω_n i ω , próbnie przyjąć jako minimalną wartość 0,50 i uznać, że wartość bliższa 0,75 byłaby bardzo korzystna. Podkreślają jednak, że są to ich subiektywne wytyczne.

Konsekwencje niespełniania założeń dla współczynników α i ω – dane empiryczne

W literaturze przedmiotu prezentowane są analizy pozwalające ocenić wrażliwość wskaźników rzetelności na łamanie założeń. W większości przypadków analizy te przeprowadzane są na danych fikcyjnych, co często skutkuje testowaniem większych odchyień od założeń niż te, z którymi mamy powszechnie do czynienia (Alessandri i in., 2015). Dlatego też dla każdego przykładu podano informacje o generowanych danych. Poniżej przedstawiono cztery przykłady analizy rzetelności wyników testu z zastosowaniem współczynników α i ω .

Przykład 1. Obciążenie α Cronbacha jako estymatora rzetelności badano na danych fikcyjnych dwóch testów (Osburn, 2000). Testy różniły się długością (4 i 8 pozycji) oraz interkorelacją pozycji (tabela 2). Współczynniki α dla obu testów są równe prawdziwej rzetelności, jeśli pozycje są τ -równoważne. Jeśli pozycje spełniają założenie modelu generycznego, α pozostaje nadal bardzo bliska prawdziwej rzetelności. Dla krótszego testu niedoszacowanie jest nieco większe, ale i tak stanowi zaledwie 1% prawdziwej rzetelności. Należy jednak podkreślić, że przyjęta na etapie generowania danych stała multiplikatywna ma niską wartość ($c_i = 1,5$).

Tabela 2. α Cronbacha – obciążenie estymatora jako efekt niespełniania założeń (Osburn, 2000, s. 349)

Rzetelność	Modele pomiarowe		Heterogeniczność – 2 czynniki skorelowane		
	τ -równoważny	generyczny	słabo (.80)	umiarkowanie (.40)	silnie (.20)
4 pozycje wysoko skorelowane (ok. .50)					
ρ	,798	,786	,781	,760	,703
α	,798	,778	,752	,696	,547
$(\alpha/\rho)*100$		(99%)	(96%)	(92%)	(78%)
8 pozycji niżej skorelowanych (ok. .25)					
ρ	,725	,712	,704	,679	,613
α	,725	,709	,692	,654	,554
$(\alpha/\rho)*100$		(99,6%)	(98%)	(96%)	(90%)

Oznaczenia: ρ – rzetelność prawdziwa.

Wpływ heterogeniczności testu na oszacowanie rzetelności sprawdzano dla testów dwuczynnikowych o trzech poziomach skorelowania czynników (tabela 2). Każdy czynnik ma po dwie lub po cztery pozycje odpowiednio dla krótszego i dłuższego testu. Im niższa wartość korelacji, tym czynniki są bardziej ortogonalne. Analizy pokazały, że w testach heterogenicznych niedoszacowanie rzetelności jest tym większe, im krótszy test i bardziej ortogonalne czynniki. Dla testu zbudowanego z czterech pozycji tworzących dwa silnie heterogeniczne czynniki współczynnik α daje wartość rzetelności o 22% niższą od prawdziwej. Niedoszacowania zmniejszają się, gdy maleje ortogonalność czynników. Na uwagę zasługuje też wyraźnie mniejsza czułość α na odstępstwa od założeń dla dłuższego testu. Dla testu 8-pozycyjnego posiadającego wyraźnie wyodrębnione dwa czynniki niedoszacowanie sięga już tylko 10% prawdziwej rzetelności.

Przykład 2. Wrażliwość α na odstępstwa od założeń modelu zasadniczo τ -równoważnego prezentuje Graham (2006) na przykładzie danych fikcyjnych testu składającego się z pięciu pozycji wygenerowanych dla 60 osób. Pierwsza analiza obejmuje pięć pozycji testu (x_1 - x_5) różniących się tylko stałą addytywną, co oznacza spełnienie założeń modelu zasadniczo τ -równoważnego (tabela 3). W dalszych analizach pozycję x_5 zastępowano kolejno pozycją x_6 i x_7 , które różniły się od pozycji x_5 stałą multiplikatywną (c_i) równą odpowiednio 5 i 10. Należy przypomnieć, że wartość $c_i \neq 1$ dopuszczalna jest tylko w modelu generycznym.

W tabeli 3 zamieszczono oszacowanie rzetelności przy użyciu modelu zasadniczo τ -równoważnego (α Cronbacha) i generycznego (ω McDonalda). W przypadku danych spełniających założenia modelu zasadniczo τ -równoważnego (x_1 - x_5) dopasowania obu modeli są tak samo dobre ($p = 0,249$), a oszacowane rzetelności są sobie równe. W przypadku zamiany w teście pozycji x_5 kolejno na x_6 i x_7 (model generyczny) model zasadniczo τ -równoważny jest źle dopasowany, a α jest obciążonym estymatorem rzetelności, przy czym obciążenie wzrasta wraz ze wzrostem odchyień od założeń (większe, gdy większa wartość stałej multiplikatywnej). Należy podkreślić, że w praktyce α Cronbacha stosuje się dla testów składających się z pozycji o tych samych skalach odpowiedzi.

Tabela 3. Dopasowanie modeli i estymowana rzetelność (Graham, 2006, s. 939)

Pozycje testu	Model	Rzetelność	CFI	RMSEA	χ^2 (df)	p
x_1 - x_5	τ -równoważny	,91	1,000	,000	5,5(9)	
	generyczny	,91	1,000	,000	0,1(5)	,249
x_1 - x_4 , x_6	τ -równoważny	,76	0,630	,473	127,8(9)	
	generyczny	,97	1,000	,000	< 0,1(5)	< ,001
x_1 - x_4 , x_7	τ -równoważny	,56	0,556	,537	161,9(9)	
	generyczny	,99	1,000	,000	< 0,1(5)	< ,001

Oznaczenia: CFI – względny indeks dopasowania; RMSEA – pierwiastek średniego kwadratu błędu aproksymacji (Konarski, 2009).

Przykład 3. Revelle i Zinbarg (2009) porównali wartości współczynników α Cronbacha i ω McDonalda dla sześciu różnych zbiorów danych, przy czym α obliczano tradycyjną metodą i metodą składowych głównych (α_{pc}), a ω wyznaczano dla całego testu (ω_t) oraz dla czynnika ogólnego (ω_h). Do wyznaczania ω badacze zastosowali eksploracyjną analizę czynnikową.

Analizy obejmują sześć zbiorów zaczerpniętych od Sijtsma (2009). W kolumnie S-1 (tabela 4) współczynniki rzetelności obliczono dla testu składającego się z ośmiu pozycji ocenianych na skali od 0 do 3 i przeprowadzonego na próbie $N = 828$ (dane rzeczywiste). Analizie poddano też dwa czynniki (S-1a i S-1b) ujawnione w analizie składowych głównych (*Principal Component Analysis*). Każdy z tych czynników zawiera po cztery pozycje o ładunkach wyższych niż 0,50.

Analizy S-2a, S-2b i S-2c przeprowadzono na danych fikcyjnych. Test S-2a ma strukturę trójczynnikową (po 2 pozycje w czynniku), S-2b – dwuczynnikową (po 3 pozycje), a S-2c – jednoczynnikową. Testy składały się z 6 pozycji o wariancjach $\sigma_i^2 = 0,25$ i kowariancjach:

- (1) dodatnich i równych dla pozycji należących do jednego czynnika;
- (2) zerowych pomiędzy pozycjami z różnych czynników;
- (3) o stałej sumie w różnych macierzach, co implikuje równe wartości α .

Z wartości zamieszczonych w tabeli 4 wynika, że α Cronbacha obliczana metodą tradycyjną i metodą składowych głównych daje bardzo zbliżone wartości i to bez względu na rodzaj danych i stopień niespełniania założeń ($\alpha_{pc} - \alpha \leq 0,02$). Ponadto we wszystkich przykładach ω_t zapewnia wyższe oszacowanie rzetelności niż α . Wybór właściwego współczynnika jest więc bardzo ważny i powinien być poprzedzony analizą struktury testu.

Tabela 4. Oszacowanie rzetelności – α i ω (Revelle, Zinbarg, 2009, s. 8)

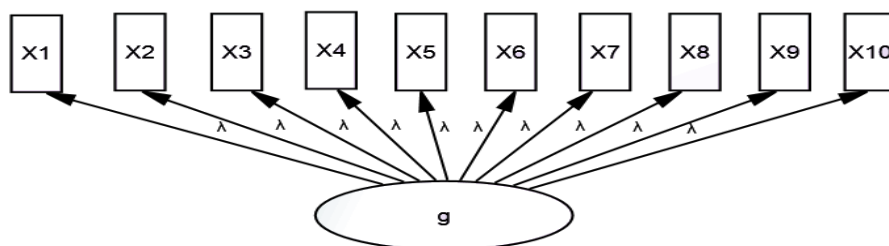
Współczynnik	S-1	S-1a	S-1b	S-2a	S-2b	S-2c
L.poz.	8	4	4	6	6	6
α	,785	,749	,676	,533	,533	,533
α_{pc}	,787	,749	,676	,553	,533	,553
ω_t	,844	,893	,859	,889	,669	,561
ω_h	,593	,643	,676	,049	,000	,532
$\alpha_{pc} - \alpha$,002	0	0	,02	0	,02
$\omega_t - \alpha$,059	,144	,183	,356	,136	,028

Oznaczenia: l.poz. – liczba pozycji w teście; α – wyznaczona metodą tradycyjną; α_{pc} – wyznaczona metodą składowych głównych; ω_t – dla całego testu; ω_h – hierarchiczna.

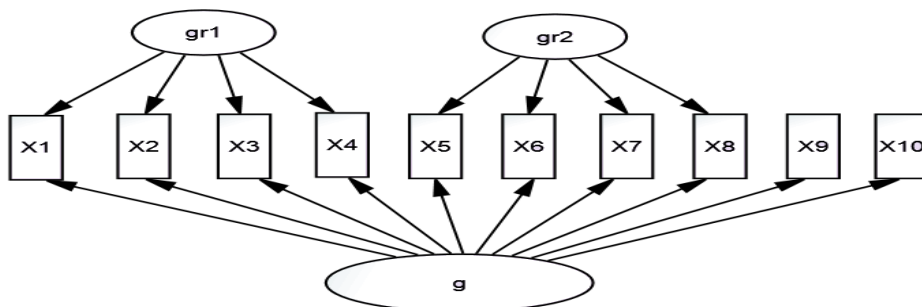
Przykłady testów różniących się liczbą czynników (S-2a, S-2b i S-2c) dające te same wartości α potwierdzają słuszność zarzutów stawianych α , że nie informuje ona o wewnętrznej strukturze testu (Stijtsma, 2009; Reise, Bonifay, Haviland, 2013a). Inaczej jest w przypadku ω . Wartości ω_t i ω_h testów S-2a, S-2b i S-2c ulegają zmianie

wraz ze zmianą struktury. Dla tych testów zarówno α , jak i ω wskazują na niedostateczną rzetelność testu do wnioskowania o ogólnym konstrukcie. Ponadto przykład ten jasno pokazuje, że dla testów heterogenicznych raportowanie ω_i bez podania ω_i jest informacją niepełną i może być mylącą.

Przykład 4. Analizy przedstawione przez Greena i Yang (2015) przeprowadzono na danych fikcyjnych dla $N = 500$ osób. Test składał się z 10 pozycji (skala Likerta 1-7). Dane wygenerowano tak, by w teście wystąpiły dwa czteropozycyjne czynniki (x_1 - x_4 i x_5 - x_8). Wyniki analizy modelu zasadniczo τ -równoważnego (rycina 2) i modelu podwójnego czynnika z dwoma podczynnikiem (rycina 3) zamieszczono w tabelach 5 i 6.



Rycina 2. Model zasadniczo τ -równoważny – wszystkie ładunki czynnikowe równe



Rycina 3. Model podwójnego czynnika (*bifactor*) – g czynnik ogólny; gr1, gr2 – podgrupy

Tabela 5 zawiera wyniki CFA dla dwóch modeli. Rzetelność oszacowana na podstawie równania (1) w modelu zasadniczo τ -równoważnym ($\alpha = 0,797$) tylko nieznacznie różni się od obliczonej metodą tradycyjną ($\alpha = 0,795$). Jednak wartości CFI $< 0,90$ oraz RMSEA $> 0,1$ wskazują na złe dopasowanie modelu do danych (Kornarski, 2009). W tej sytuacji nie należy interpretować ani szacowanych parametrów modelu, ani współczynników opartych na tych oszacowaniach (Green, Yang, 2015, s. 17). Dobrze dopasowany okazał się model podwójnego czynnika, a rzetelność estymowana współczynnikiem ω_i (3) jest wyższa od α o 0,048.

Tabela 5. Rzetelność i wskaźniki dopasowania $N = 500$ (Green, Yang, 2015, s. 18)

Model	Analiza rzetelności			Wskaźniki dopasowania modelu			
	Wariancja wyniku prawdziwego	Wariancja błędu	Rzetelność	χ^2	df	CFI	RMSEA
Zasadniczo τ -równoważny	41,518	10,579	,797	436,669	44	,686	,134
Podwójnego czynnika	43,521	8,107	,843	29,743	27	,998	,014

Rozwinięcie ω dla testów heterogenicznych pozwala dodatkowo ocenić nasylenie wyników testu czynnikiem ogólnym związanym z mierzoną zmienną oraz z podczynnikiem. Wyniki modelu podwójnego czynnika dla analizowanego przykładu (rycina 3) zamieszczono w tabeli 6.

Tabela 6. Wariancja i rzetelność w modelu podwójnego czynnika (Green, Yang, 2015, s. 18)

Wyniki obserwowane	Wariancja				Rzetelność dla czynników			
	czynnika ogólnego	podskali x_1-x_4	podskali x_5-x_8	błędu	całkowita	wszystkich	ogólnego	podczynników
x_1-x_{10}	29,690	7,123	6,711	8,117	51,628	,843	,575	,268
x_1-x_4	3,716	7,120		3,064	13,900	,779	,267	,512
x_5-x_8	4,140		6,711	3,050	13,900	,781	,298	,483
x_9-x_{10}	2,210			1,993	4,203	,526	,526	

Rzetelność całej skali ($\omega_t = 0,843$) wskazuje, że wariancja wyników prawdziwych stanowi 84,3% wariancji wyników otrzymanych. Współczynnik ω_h , obliczony zgodnie z równaniem (4), dla czynnika ogólnego całego testu ma wartość 0,575, co oznacza, że czynnik ogólny wyjaśnia 57,5% wariancji otrzymanych wyników testu. Stosunek ω_h (0,575) do ω_t (0,843) wskazuje, że 68,2% rzetelnej wariancji wyników związanej jest z czynnikiem ogólnym. Pozostała część wariancji wyników (0,268) jest wariancją związaną z dwoma podczynnikiem występującymi w skali. Stanowi ona 31,8% wariancji prawdziwej ($100 \cdot 0,268 / 0,843$). Współczynnik ω_{si} (6) dostarcza dodatkowo informacji o jakości wyników podskal. Rzetelność całej podskali (x_1-x_4) jest zadowalająca ($\omega_t = 0,779$), przy czym około jedna czwarta wariancji wyników otrzymanych tej podskali jest związana z czynnikiem ogólnym (0,267), a 51,2% – z podskala.

Jak twierdzi Reise (2012), ω_i jest przydatnym wskaźnikiem jednowymiarowości wyników testu, jednak tylko, gdy przyjęty model ma sens teoretyczny i pasuje do danych. Wartość $\omega_i > 0,7$ można uznać za dobry wskaźnik jednowymiarowości testu (Reise, Bonifay, Haviland, 2013a; Reise i in., 2013b). Porównanie rzetelności wielowymiarowego, złożonego wyniku całkowitego (ω_t) z ω_i wskazuje stopień, w jakim wynik całkowity odzwierciedla rzetelną wariancję mierzonego konstrukt. Brak jednak uzasadnionych wskazań minimalnych wartości ω_{it} , które należy uznać za wystarczające, by wyniki całej skali traktować jako dobry wskaźnik badanego konstrukt.

Przykład obliczania ω w SPSS

Do obliczania ω można stosować zarówno eksploracyjną (EFA), jak i konfirmacyjną analizę czynnikową (CFA). EFA jest analizą preferowaną, gdy nie można określić modelu z powodu niewystarczającej wiedzy na temat modelu pomiarowego lub modele okazały się źle dopasowane w CFA (Green, Yang, 2015).

Komercyjne pakiety statystyczne nie oferują gotowej procedury obliczania ω . Współczynnik ω można otrzymać na dwa sposoby. Jednym z nich jest wykorzystanie pakietu R (*R psych package*; Revelle, 2017), który dostarcza wartości ω , drugi to obliczanie ω na podstawie wyników analizy czynnikowej przeprowadzonej w komercyjnych pakietach statystycznych. W tym przypadku otrzymujemy wartości ładunków czynnikowych i błędów i konieczne jest wykonanie dodatkowych obliczeń z wykorzystaniem równań (3)-(6).

Poniżej przedstawiono wyniki analizy testu składającego się z 15 pozycji przeprowadzonego na próbie 660 osób (dane rzeczywiste). W CFA wykonanej w Amos 21 testowano model podwójnego czynnika z pięcioma podczynnikami, z których każdy zawierał po 3 pozycje. Model okazał się dobrze dopasowany do danych: $\chi^2(40) = 106,12, p < 0,001$; RMSEA = 0,051 [90%CI: 0,039-0,062]; CFI = 0,975.

W tabeli 7 zamieszczono standaryzowane ładunki czynnikowe dla czynnika ogólnego oraz pięciu podczynników. W ostatniej kolumnie znajdują się standaryzowane wariancje błędów dla każdej pozycji. Dane te pozwalają obliczyć współczynniki ω McDonalda zgodnie z równaniami (3)-(6).

Tabela 7. Wyniki testowania modelu podwójnego czynnika w CFA ($N = 660$)

Nr poz.	Standaryzowane ładunki czynnikowe					Wariancja 1- h^2
	g	F1	F2	F3	F4	
P1	,58		,53			,38
P2	,63			,27		,53
P3	,74				,27	,38
P4	,69	,56				,22
P5	,67				,35	,43
P6	,64		,54			,29

cd. tabeli 7

P7	,66			,66			,11
P8	,64					,18	,56
P9	,65	,66					,15
P10	,63				,31		,50
P11	,66		,51				,30
P12	,67			,57			,22
P13	,69					,39	,37
P16	,58	,50					,41
P15	,54				,56		,40
Σ	9,67	1,72	1,58	1,5	1,22	,84	5,25
Σ^2	93,51	2,96	2,50	2,25	1,49	,71	-

Oznaczenia: P1-P15 – numery pozycji testu; g – czynnik ogólny; F1-F5 – podczynniki; $1-h^2$ – standaryzowana wariancja błędów; Σ – sumowanie; Σ^2 – kwadraty sum.

W dwóch ostatnich wierszach tabeli 7 zamieszczono odpowiednio sumę standaryzowanych ładunków czynnikowych dla czynnika ogólnego i podczynników oraz kwadraty tych sum. Zsumowano również wariancję błędów.

Zgodnie z równaniem

$$(3) \quad \omega_t = \frac{93,51 + 2,96 + 2,50 + 2,25 + 1,49 + 0,71}{93,51 + 2,96 + 2,50 + 2,25 + 1,49 + 0,71 + 5,25} = 0,95;$$

$$(4) \quad \omega_h = \frac{93,51}{93,51 + 2,96 + 2,50 + 2,25 + 1,49 + 0,71 + 5,25} = 0,86;$$

$$(5) \quad \omega_{t,F1} = \frac{2,96 + 3,68}{2,96 + 3,68 + 0,78} = 0,90;$$

$$(6) \quad \omega_{h,F1} = \frac{2,96}{2,96 + 3,68 + 0,78} = 0,40;$$

Należy zwrócić uwagę, że w obliczeniach dotyczących ω dla podczynników (równania (5) i (6)) uczestniczą tylko pozycje tego podczynnika i to zarówno przy uwzględnianiu czynnika ogólnego, jak i wariancji błędów.

Otrzymane wartości oznaczają, że wszystkie wspólne źródła wariancji wyjaśniają (ω_t) 95% wariancji wyników otrzymanych, a (ω_h) 86% całkowitej wariancji wyników testu można przypisać czynnikowi ogólnemu. Pierwiastek kwadratowy z ω_h (0,93) jest korelacją między czynnikiem ogólnym a wynikiem obserwowanym. Ważne jest też porównanie ω_h z ω_t . Można zauważyć, że 91% (stosunek ω_h do ω_t) rzetelnej wariancji testu związanej jest z czynnikiem ogólnym. Tylko 9% (różnica między ω_t i ω_h) wariancji można przypisać wielowymiarowości testu spowodowanej podczynnikiem (Rodriguez, Reise, Haviland, 2016). Na podstawie uzyskanych wy-

ników można stwierdzić, że test jest wysoce rzetelny ($\omega_i = 0,95$). Wartość ω_i wskazuje również na wysokie nasycenie wyników testu czynnikiem ogólnym, co upoważnia do wnioskowania o poziomie mierzonej cechy na podstawie wyników testu.

Do podobnych konkluzji prowadzi α , która obliczona metodą tradycyjną dla tego testu ma wartość 0,92.

Podsumowanie

Współczynnik ω McDonalda jest jednym ze współczynników szacowania rzetelności metodą zgodności wewnętrznej. W literaturze wskazywane są jego zalety w porównaniu z powszechnie używaną α Cronbacha. Dlatego też zestawienie zalet i wad obu współczynników powinno ułatwić czytelnikowi decyzję w doborze właściwego.

W opracowaniu przedstawiono podstawowe założenia teoretyczne oraz równania niezbędne do obliczania wartości współczynników ω i α . Z przedstawionych porównań współczynników na poziomie teoretycznym wynika, że ω jest lepszym oszacowaniem rzetelności niż α . W dalszej części przytoczono przykłady pokazujące wpływ łamania założeń na stronniczość ω i α w szacowaniu rzetelności. Należy jednak podkreślić, że przykłady te dotyczą w znacznej części danych fikcyjnych, które istotnie odbiegają od typowych danych empirycznych. Nakazuje to ostrożność w formułowaniu na ich podstawie kategoriycznych wniosków. Ponadto na przykładzie danych rzeczywistych dla heterogenicznego testu pokazano sposób obliczania wartości ω na podstawie wyników CFA i równań (3)-(6). Wszystko to pozwala na sformułowanie poniższych spostrzeżeń.

Współczynnik α ma wiele zalet. Procedura jego obliczania jest dostępna we wszystkich pakietach statystycznych, a notacja matematyczna prosta i intuicyjnie zrozumiała. Zapewnia on stosunkowo dokładne oszacowanie rzetelności jeśli spełnione są wymagane założenia lub jeśli odstępstwa od założeń są niewielkie (Green, Yang, 2015).

Współczynnik α posiada też pewne wady. Założenia przyjmowane dla α są bardziej restrykcyjne (model zasadniczo τ -równoważny) niż dla ω (Zinbarg i in., 2005). Szacowanie rzetelności powinno być poprzedzone analizą czynnikową, ponieważ α nie jest odpowiednim współczynnikiem do szacowania rzetelności pomiaru konstruktywów o strukturze niejednorodnej (Brunner, Nagy, Wilhelm, 2012; Green, Yang, 2015). Relacja między α a ładunkami czynnikowymi (1) potwierdza, podkreślany w literaturze, brak w α informacji o strukturze testu (Schmitt, 1996; Osburn, 2000; Graham, 2006; Revelle, Zinbarg, 2009; Peters, 2014). Stąd niewłaściwe jest interpretowanie α jako wskaźnika wewnętrznej spójności i jednorodności testu (Stijtsma, 2009; Reise, Bonifay, Haviland, 2013a). Kolejnym zarzutem stawianym α jest punktowe oszacowanie rzetelności w standardowych procedurach SPSS i Statistica (Terry, Kelly, 2012; Kelley, Pornprasertmanit, 2016), chociaż wyznaczanie przedziałów ufności możliwe jest w pakiecie R (Dunn, Baguley, Brunnsden, 2014; Bonett, Wright, 2015).

Ponadto naruszenie założeń modelu zasadniczo τ -równoważnego skutkuje niedoszacowaniem rzetelności przez współczynnik α (Graham, 2006; Sijtsma, 2009a).

Współczynniki ω dostarczają szczególnie użytecznych informacji o strukturze testu (Revelle, 2017). Dla testów wielowymiarowych pozwalają one ocenić nie tylko rzetelność całego testu (ω_t), ale i wpływ czynnika ogólnego (ω_h) i podczynników (ω_s) na rzetelność testu (Schmitt, 1996; Osburn, 2000). Ponadto jasno zdefiniowany związek między współczynnikami ω i modelem pomiarowym zmniejsza prawdopodobieństwo ich błędnej interpretacji. Założenia konieczne do stosowania ω są mniej restrykcyjne (model generyczny) niż w przypadku α (Zinbarg i in., 2005). Również mniejsza jest wrażliwość ω na odstępstwa od założeń, co w konsekwencji prowadzi do mniejszego niedoszacowania rzetelności (Revelle, Zinbarg, 2009; Green, Yang, 2015).

Jednak współczynnik ω , nie jest wolny od wad. Bazuje on na analizie czynnikowej, co jest przyczyną pewnych utrudnień. Po pierwsze stosowanie ω wymaga od użytkowników znajomości analizy czynnikowej (EFA i/lub CFA). Po drugie szacowanie parametrów w modelach pomiarowych wymaga dużych prób ($N \geq 200$; Yang, Green, 2010). Po trzecie modele pomiarowe mogą okazać się źle dopasowane do danych, co w CFA nie uprawnia do ich interpretowania, a w EFA może powodować stronniczość ω (Green, Yang, 2015). Po czwarte dla ω_h i ω_s brakuje uzasadnionych wskazań jakie wartości tych współczynników można uznać za zadowalające (Reise, Bonifay, Haviland, 2013a). Ostatnim z wymienionych utrudnień jest brak gotowej procedury w SPSS i Statistica udostępniającej wartości ω . Jest ona dostępna w otwartym oprogramowaniu R (Revelle, 2017), jednak jej użytkowanie nie jest intuicyjne.

Omega jest wskazywana przez część badaczy jako lepszy wskaźnik zgodności wewnętrznej zarówno w porównaniu z α , jak i innymi współczynnikami (Raykov, 1997a, 1997b; Zinbarg i in., 2005; Graham, 2006; Revelle, Zinbarg, 2009). Na podstawie przedstawionych przykładów wydaje się najbardziej zasadne używanie zarówno α , jak i ω szczególnie na etapie opracowania testu. Jednak dopóki ω pozostanie niedostępna w pakietach takich, jak SPSS i Statistica wątpliwe jest, by stała się ona powszechnie stosowaną. Zwraca na to uwagę Sijtsma, pisząc: „oprogramowanie powinno być tak łatwe w stosowaniu, jak SPSS, przynajmniej w przypadku kiedy chcemy, żeby nasze narzędzia (psychometryczne) miały wpływ na istotne badania i nie «zgubiły się» w obszarze «niczym» między teorią i praktyką” (2009b, s. 172).

Literatura cytowana

- Alessandri, G., Vecchione, M., Eisenberg, N., Laguna, M. (2015). On the factor structure of the Rosenberg (1965) General Self-Esteem Scale. *Psychological Assessment*, 27 (2), 621-635, doi: 10.1037/pas0000073
- Bonett, D., Wright, T. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior*, 36, 3-15, doi: 10.1002/job.1960

- Brunner, M., Nagy, G., Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, 80 (4), 796-846, doi: 10.1111/j.1467-6494.2011.00749.x
- Brzeziński, J. (1996). *Metodologia badań psychologicznych*. Warszawa: Wydawnictwo Naukowe PWN.
- Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104, doi: 10.1037/0021-9010.78.1.98
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of a test. *Psychometrika*, 16, 297-334, doi: 10.1007/BF02310555
- Dunn, T., Baguley, T., Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105, 399-412, doi: 10.1111/bjop.12046
- Gibbons, R., Hedeker, D. (1992). Full-information item bi-factor analysis. *Psychometrika*, 57, 423-436, doi: 10.1007/BF02295430
- Gignac, G. (2008). Higher-order models versus direct hierarchical models: g as superordinate or breadth factor? *Psychology Science Quarterly*, 50, 21-43.
- Gignac, G. (2014). On the inappropriateness of using items to calculate total scale score reliability via coefficient alpha for multidimensional scale. *European Journal of Psychological Assessment*, 30 (2), 130-139, doi: 10.1027/1015-5759/a000181
- Graham, J. (2006). Congeneric and (Essentially) Tau-Equivalent Estimates of score reliability What they are and how to use them. *Educational and Psychological Measurement*, 66 (6), 930-944, doi: 10.1177/0013164406288165
- Green, S., Yang, Y. (2015). Evaluation of dimensionality in the assessment of internal consistency reliability: Coefficient alpha and omega coefficients. *Educational Measurement: Issues and Practice*, 34 (4), 14-20, doi: 10.1111/emip.12100
- Grygiel, P. (2016). Test podłużnej niezmienności modelu podwójnego czynnika na przykładzie Kwestionariusza poczucia integracji rówieśniczej. *Edukacja*, 2 (137), 79-99.
- Gu, F., Little, T., Kingston, N. (2013). Misestimation of reliability using coefficient alpha and structural equation modeling when assumptions of tau-equivalence and uncorrelated errors are violated. *Methodology*, 9 (1), 30-40, doi: 10.1027/1614-2241/a000052
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282, doi: 10.1007/BF02288892
- Hornowska, E. (2001). *Testy psychologiczne. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe Scholar.
- Kelley, K., Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods*, 21 (1), 69-92, doi: 10.1037/a0040086
- Konarski, R. (2009). *Modele równań strukturalnych. Teoria i praktyka*. Warszawa: Wydawnictwo Naukowe PWN.
- Lord, F., Novick, M. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

- McDonald, R. (1978). Generalizability in factorable domains: Domain validity and generalizability. *Educational and Psychological Measurement*, 38, 75-79, doi: 10.1177/001316447803800111
- McDonald, R. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.
- McDonald, R. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McDonald, R. (2010). Structural models and the art of approximation. *Perspectives on Psychological Science*, 5, 675-686, doi: 10.1177/1745691610388766
- Mellenbergh, G. (1996). Measurement precision in test score and item response models. *Psychological Methods*, 1, 293-299, doi: 10.1037/1082-989X.1.3.293
- Moumni, M., Van Eck, M., Wendt, K., Reininga, I., Mekkink, L. (2016). Structural validity of the dutch version of the Patient-Rated Wrist Evaluation (PRWE-NL) in patients with hand and wrist injuries. *Physical Therapy*, 96 (6), 908-916, doi: 10.2522/ptj.20140589
- Muthén, L., Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599-620.
- Niemierko, B. (1999). *Pomiar wyników kształcenia*. Warszawa: Wydawnictwa Szkolne i Pedagogiczne.
- Novick, M., Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1-13, doi: 10.1007/BF02289400
- Osburn, H. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, 5 (3), 343-335, doi: 10.1037/1082-989X.5.3.343
- Peters, G. (2014). The alpha and the omega of scale reliability and validity. *The European Health Psychologist*, 16 (2), 56-69.
- Raykov, T. (1997a). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21, 173-184, doi: 10.1177/01466216970212006
- Raykov, T. (1997b). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, 32, 329-353.
- Reise, S. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47, 667-696, doi: 10.1080/00273171.2012.715555
- Reise, S., Bonifay, W., Haviland, M. (2013a). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95 (2), 129-140, doi: 10.1080/00223891.2012.725437
- Reise, S., Scheines, R., Widaman, K., Haviland, M. (2013b). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, 73, 5-26, doi: 10.1177/0013164412449831
- Revelle, W. (2017). An overview of the psych package, <https://cran.r-project.org/web/packages/psych/vignettes/overview.pdf>
- Revelle, W., Zinbarg, R. (2009). Coefficients alpha, beta, omega and the glb: Comments on Sijtsma. *Psychometrika*, 74, 145, doi: 10.1007/s11336-008-9102-z
- Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47, 361-372.

- Rindskopf, D., Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, 23, 51-67, doi: 10.1207/s15327906mbr2301_3
- Rodriguez, A., Reise, S., Haviland, M. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21 (2), 137-150, doi: 10.1037/met0000045
- Schmitt, N. (1996). Uses and abuses of coefficient Alpha. *Psychological Assessment*, 8 (4), 350-353.
- Sijtsma, K. (2009a). On the use, the misuse, and the very limited usefulness of Cronbach's α . *Psychometrika*, 74 (1), 107-120, doi: 10.1007/s11336-008-9101-0
- Sijtsma, K. (2009b). Reliability beyond theory and into practice. *Psychometrika*, 74 (1), 169-173, doi: 10.1007/S11336-008-9103-Y
- Terry, L., Kelly, K. (2012). Sample size planning for composite reliability coefficients: Accuracy in parameter estimation via narrow confidence intervals. *British Journal of Mathematical and Statistical Psychology*, 65, 371-401, doi: 10.1111/j.2044-8317.2011.02030.x
- Woodward, J., Bentler, P. (1979). A statistical lower bound to population reliability. *Psychological Bulletin*, 6, 1323-1326, doi: 10.1037/0033-2909.85.6.1323
- Yang, Y., Green, S.B. (2010). A note on structural equation modeling estimates of reliability. *Structural Equation Modeling*, 17, 66-81, doi: 10.1080/10705510903438963
- Zimmerman, D., Zumbo, B., Lalonde, C. (1993). Coefficient alpha as an estimate of test reliability under violation of two assumptions. *Educational and Psychological Measurement*, 53, 33-49, doi: 10.1177/0013164493053001003
- Zinbarg, R., Revelle, W., Yovel, I., Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70 (1), 123-133, doi: 10.1007/s11336-003-0974-7

Streszczenie. W artykule przedstawiono zalety i wady alfy Cronbacha oraz omegi McDonalda jako współczynników rzetelności testów. Na przykładzie pokazano też sposób obliczania omegi wykorzystując pakiet SPSS. Stosowanie alfy, oprócz założeń Klasycznej Teorii Testu, wymaga dodatkowo jednowymiarowości testu oraz równości wyników prawdziwych pozycji. Niespełnienie tych założeń powoduje, że wartość współczynnika alfa jest obciążonym estymatorem rzetelności. Istnieje potrzeba przeprowadzania analizy czynnikowej jako poprzedzającej stosowanie alfy Cronbacha. Szczególnie w przypadku testów heterogenicznych proponowane jest stosowanie innych współczynników. Wśród nich wymieniana jest omega McDonalda, która przyjmuje mniej restrykcyjne założenia. Przedstawione w artykule analizy teoretyczne i badania symulacyjne pozwalają ustosunkować się do konsekwencji niespełniania założeń w szacowaniu rzetelności.

Słowa kluczowe: rzetelność, wielowymiarowość, model podwójnego czynnika, omega McDonalda, alfa Cronbacha

Data wpłynięcia: 20.10.2017

Data wpłynięcia po poprawkach: 15.03.2018

Data zatwierdzenia tekstu do druku: 31.03.2018