

Pouzdanost mjerenja u psihologiji: Razvoj metode, zaludenost Cronbachovim alfa koeficijentom i preporuke za ispravnu procjenu pouzdanosti

Josip Novak

Sveučilište u Zagrebu, Filozofski fakultet, Odsjek za psihologiju, Zagreb, Hrvatska

Sažetak

Svrha je ovoga preglednog rada uputiti na bogatu povijest koncepta pouzdanosti mjerenja i strjelovit napredak u procjeni pouzdanosti koji se dogodio od 1940-ih, što je rezultiralo velikim brojem pristupa procjeni pouzdanosti. Na početku je rada naveden kratak povijesni pregled nastanka koncepta pouzdanosti i metoda procjene pouzdanosti. Potom su opisani koeficijenti unutarnje konzistencije prateći njihov kronološki razvoj. Kritizirana je primjena Cronbachova α koeficijenta bez provjere potrebnih pretpostavki te su obrazložene česta pogrešna shvaćanja vezana uz Cronbachov α koeficijent. Navedene su pretpostavke potrebne za opravdanu primjenu α koeficijenta, pristranosti procjene u slučaju nezadovoljenosti potrebnih pretpostavki, ispravan način izvještavanja o pouzdanosti, kao i moguće primjene koeficijenta u svrhe različite od procjene pouzdanosti. Međutim, u praksi su pretpostavke za opravdanu primjenu α koeficijenta vrlo rijetko zadovoljene i često je opravdana primjena drugih koeficijenata, poput λ_2 , λ_4 , ρ_{NDG} ili ω , za kojeg se pokazalo da je u većini uvjeta najmanje pristran. Ponuđene su smjernice za primjenu navedenih koeficijenata. Najnoviji je pristup procjeni pouzdanosti strukturalnim modeliranjem, koje omogućava jednostavnu provjeru navedenih pretpostavki i u kombinaciji s ω i ω_H koeficijentima rezultira vjerodostojnom procjenom pouzdanosti, a opravdan je u slučaju kompleksnih mjernih modela. Naposljetku, istaknuti su konceptualni problemi vezani uz procjenu pouzdanosti te su predložene smjernice za daljnja istraživanja pouzdanosti mjerenja.

Ključne riječi: psihometrija, pouzdanost mjerenja, Cronbachov alfa koeficijent, McDonaldov omega koeficijent, teorija testova, strukturalno modeliranje

✉ Josip Novak, Filozofski fakultet, Odsjek za psihologiju, I. Lučića 3, 10000 Zagreb, Hrvatska. E-pošta: josip.novak92@gmail.com

Napomena:

Zahvaljujem doc. dr. sc. Blažu Rebernjaku na iznimno korisnim povratnim informacijama.

Nastanak koncepta pouzdanosti mjerenja

Pouzdanost mjerenja nužna je za svako valjano znanstveno zaključivanje. Međutim, iako je u psihologiji pouzdanost mjerenja doživjela nagli razvoj od početka njezina istraživanja, spoznaje nakon nastanka Cronbachova α koeficijenta (koji uopće nije Cronbachov izum) nisu poznate većini istraživača, kao ni funkcioniranje samog α koeficijenta, koji je pogrešno upotrebljavan. Prilikom obrazovanja psihologa u program uglavnom nisu uključene određene informacije o pouzdanosti mjerenja, koje su ne samo korisne već i potrebne budućim istraživačima i praktičarima. U literaturi općenito, a prema tome i u Hrvatskoj, dosad nije pridano dovoljno pozornosti ovoj problematici i stoga je svrha ovog rada prikazati povijesni razvoj pouzdanosti mjerenja, navesti česte miskoncepcije vezane uz najčešće upotrebljavan Cronbachov α koeficijent te skupiti postojeće i dopuniti preporuke za ispravnu procjenu pouzdanosti, kao i usmjerenje budućih istraživanja.

Istraživanje pouzdanosti mjerenja u psihologiji započelo je sa Spearmanom (1904, 1910), koji je uveo koncepte pravog rezultata i pogreške mjerenja, kao i korekciju za atenuaciju i tzv. *split-half* procjenu pouzdanosti. Prema njemu, pouzdanost je sredstvo pronalaska prave korelacije između „pravih“ vrijednosti ispravljanjem atenuacije opažene korelacije, koja je rezultat slučajnih pogrešaka nastalih tijekom procesa mjerenja. Predložio je formulu utemeljenu na pretpostavci da je navedene slučajne pogreške moguće mjeriti pomoću veličine diskrepancija između sukcesivnih mjerenja istih konstrukata.

$$r_{x'y'} = \frac{r_{xy}}{\sqrt{r_{xx'}r_{yy'}}} \quad (1)$$

S obzirom na to da je pretpostavljena prava korelacija umanjena ili atenuirana, ako je dobivena povezanost podijeljena korijenom umnoška pouzdanosti dviju varijabli, tada dezatenuirana korelacija predstavlja odnos varijabli izmjerenih bez pogreške. Procjena pouzdanosti u nazivniku odnosi se na korelacije opaženih rezultata X i Y s njihovim paralelnim¹ formama X' i Y' . Spearman (1904) je smatrao kako je za utvrđivanje „pravih objektivnih vrijednosti“ potrebno pronaći prosječnu korelaciju dvaju ili više neovisno dobivenih nizova rezultata. U slučajevima kada je dostupan isključivo jedan test, prvotno je rješenje bilo procijeniti pouzdanost na temelju korelacije između dvaju dijelova testa ($k = 2$), uz korekciju koja uzima u obzir da se radi o polovicama testa (Brown, 1910; Spearman, 1910).

$$\rho_{xx'}^* = \frac{kr_{xx'}}{1 + (k - 1)r_{xx'}} \quad (2)$$

¹ Paralelni su testovi oni koji mjere isti konstrukt i stoga imaju jednake aritmetičke sredine i varijance rezultata te nekorelirane pogreške jednakih varijanci (Lord i Novick, 1968), kao i jednake povezanosti s vanjskim varijablama.

Kada je varijabilitet rezultata polovica testa nejednak, tada nije opravdano koristiti Spearman-Brownovu formulu. Stoga su Flanagan (1937) i Rulon (1939) predložili drugačiju formulaciju *split-half* pouzdanosti, koja uzima u obzir i varijabilitet rezultata. Njihove interpretacije pouzdanosti zajedno sa Spearmanovom koncepcijom čine temelj za kasniju interpretaciju koncepta pouzdanosti. Postoji i integrirana varijanta formule, u kojoj su korištene korelacija i standardne devijacije. Flanagan-Rulonova formula poseban je oblik kasnije najkorištenijeg, α koeficijenta (Cronbach, 1951).

$$\rho_{Flanagan} = 2 \left(1 - \frac{var_1 + var_2}{var_X} \right) \quad (3)$$

$$\rho_{Rulon} = 1 - \frac{var_e}{var_X} \quad (4)$$

$$\rho_{Flanagan-Rulon} = \frac{(4r_{1,2}SD_1SD_2)}{var_X} \quad (5)$$

Spearmanova i Flanagan-Rulonova koncepcija pouzdanosti potakle su teorijsku raspravu o značenju pravog rezultata i pogreške mjerenja te procjeni i interpretaciji pouzdanosti kao koncepta (Cronbach, 1943, 1947, 1951; Gulliksen, 1950; Guttman, 1945; Kelley, 1921; Kuder i Richardson, 1937; Loevinger, 1947; Lord, 1955a, 1955b, 1958; Thurstone, 1931; Tryon, 1957). Tijekom desetljeća ova je rasprava rezultirala matematičkom formulacijom klasične teorije testova (Lord i Novick, 1968), još poznate pod nazivima klasična teorija mjerenja, odnosno teorija pravog rezultata.

Pouzdanost u okviru klasične teorije testova i metode procjene pouzdanosti

Klasična teorija testova (KTT) temelji se na pretpostavci da je bruto rezultat (X_N) suma varijance pravog rezultata (X_T) i varijance pogreške (e).

$$X_N = X_T + e \quad (6)$$

U KTT-u vrijede aksiomi da je aritmetička sredina pogreške nula, odnosno da bi opaženi rezultat sudionika u slučaju beskonačnog broja primjena testa bio jednak pravom rezultatu, te da je varijanca pogreške nepovezana s varijancom pravih rezultata, kao ni s varijancom pogreške u paralelnim mjerenjima (Lord, 1959a, 1959b; Novick, 1966).

S obzirom na to da pravu varijancu nije moguće jednoznačno odrediti, korištene su različite metode procjene pouzdanosti. Navedene metode ne zahvaćaju iste izvore varijance pogreške, pa podrazumijevaju različite vrste pouzdanosti: pouzdanost ponovljenog mjerenja ili test-retest pouzdanost (kojom dobivamo koeficijente

stabilnosti), pouzdanost paralelnih/alternativnih formi² (kojom dobivamo koeficijente ekvivalentnosti) te pouzdanost unutarnje ili interne konzistencije (kojom dobivamo koeficijente unutarnje konzistencije). Navedeni koeficijenti nisu jednoznačni i uzajamno zamjenjivi te imaju različite vrijednosti u slučaju korekcije za atenuaciju, određivanja standardne pogreške mjerenja i ostalih praktičnih primjena (Cronbach, 1947).

Pouzdanost ponovljenog mjerenja

Pouzdanost se ponovljenog mjerenja odnosi na primjenu istog testa u dvije vremenske točke i izračun koeficijenta korelacije između opaženih rezultata. Iako je ovom metodom moguće odrediti trajna svojstva uzorka (Cronbach, Gleser, Nanda i Rajaratnam, 1972), ona ima određene nedostatke, poput efekta prenošenja. Točnije, mjerenje u prvoj točki može kontaminirati rezultate u drugoj točki mjerenja zbog retencije ili uvježbavanja (Cronbach i Furby, 1970). Drugi je nedostatak vremenski interval između točaka mjerenja. Kratak interval može povećati vjerojatnost efekta prenošenja zbog retencije, uvježbavanja i sl., dok dugačak interval može povećati vjerojatnost efekta prenošenja zbog pribavljenih informacija, a i u interakciji je s razvojnim promjenama kod sudionika³. Prema tome, ova je metoda opravdana u slučaju primjene mjera stabilnih karakteristika (Allen i Yen, 1979). Nadalje, najveći je nedostatak što nije moguće razlučiti stvarnu promjenu i nepouzdanost, odnosno test-retest-korelacija predstavlja 1) stupanj u kojemu pogreška mjerenja utječe na rezultate testa i 2) količinu promjene u pravim rezultatima. Osim toga, test-retest korelacija djelomično je ovisna o autokorelacijama čestica, što umjetno povećava procjenu pouzdanosti. Naposljetku, relativno visok koeficijent stabilnosti može biti opažen uz nisku unutarnju konzistenciju (Nunnally i Bernstein, 1994).

Pouzdanost paralelnih/alternativnih formi

Pouzdanost paralelnih/alternativnih formi je korelacija opaženih rezultata sudionika u dvije paralelne forme testa. Budući da često nije moguće potvrditi paralelnost dvaju testova, uglavnom su korištene alternativne forme. Alternativne su forme bilo koja dva testa koji su konstruirani kako bi bili paralelni⁴ i ako udovoljavaju uvjetima, opravdano je prijeći u izračun korelacija između njihovih

² Drugi je naziv pouzdanost ekvivalentnih/alternativnih formi.

³ Nije moguće jednoznačno definirati kratak i dugačak interval zbog specifičnosti pojedinih konstrukata, iako se u literaturi većinom kratkim intervalom smatra period od 14 dana, a dugačkim period od šest mjeseci i više (npr. Nunnally i Bernstein, 1994).

⁴ Ovdje je važno razlučiti paralelne i alternativne forme testa. Budući da u praksi istinski paralelne forme ne postoje, korištene su tzv. alternativne forme, koje imaju približno jednake aritmetičke sredine, varijance i korelacije s drugim mjerama i zato je opravdano tretirati ih kao paralelne (Allen i Yen, 1979).

rezultata. Nedostaci su ovog pristupa jednaki kao kod test-retest procjene pouzdanosti (Allen i Yen, 1979) uz dodatak pogreške zbog uzorkovanja čestica, ali ova procjena pouzdanosti nije pod utjecajem efekta prenošenja (Cohen i Swerdlik, 2018).

Postoji i tzv. pouzdanost odgođenih paralelnih/alternativnih formi (Webb, Shavelson i Haertel, 2006), koja podrazumijeva primjenu alternativnih formi testova u dvije vremenske točke. U ovakvu nacrtu dugoročni opći efekt pridonosi konzistentnosti, dok ostali efekti pridonose pogrešci. Moguće je usporediti korelacije alternativnih formi dobivene nakon kraćeg i dužeg vremenskog intervala, pri čemu veća korelacija u slučaju kraćeg vremenskog intervala indicira da je varijacija u osobini tijekom vremena značajan izvor nepouzdanosti, odnosno da se mjera više odnosi na stanje nego na osobinu (Nunnally i Bernstein, 1994). Procjena pouzdanosti dobivena ovim pristupom naziva se koeficijent stabilnosti i ekvivalentnosti (Cronbach, 1947) i uglavnom iznosi manje u odnosu na koeficijent stabilnosti i koeficijent ekvivalentnosti određene za alternativne forme testa primijenjene na istom uzorku (Crocker i Algina, 1986).

Pouzdanost unutarnje konzistencije

Unutarnja se konzistencija odnosi na procjenu pouzdanosti na temelju primjene testa u jednoj vremenskoj točki. Moguće ih je podijeliti u dvije skupine, od kojih je jedna temeljena na *split-half* metodi, a druga na analizi struktura varijanci-kovarijanci čestica (Crocker i Algina, 1986). U početku je najčešći pristup bila *split-half* procjena pouzdanosti, temeljena na podjeli testa u dva jednaka dijela koje je moguće tretirati kao alternativne forme. Procjena pouzdanosti unutarnje konzistencije nije podložna nedostacima prethodno navedenih pristupa. Međutim, ova metoda nije opravdana ako test nije moguće podijeliti na paralelne dijelove zbog broja čestica ili njihovih karakteristika (Allen i Yen, 1979). Zbog jednostavnosti izračuna najčešće je korištena procjena pouzdanosti kao unutarnje konzistencije (pogledati Hogan, Benjamin i Brezinski, 2000).

Procjena pouzdanosti unutarnje konzistencije

Kuder i Richardson (1937) osmislili su metodu za procjenu pouzdanosti testa koji se sastoji od dihotomnih čestica bez njegova razdvajanja u polovice. Sugeriraju kako je putem *split-half* metode test podijeljen u dva arbitrarno određena dijela, a tu je podjelu moguće učiniti na $\frac{J}{2\binom{J}{2}}$ načina, od kojih svaki daje različitu procjenu

pouzdanosti. Logika je koeficijenta sljedeća: 1) iznaći prosječne korelacije između J čestica u testu, 2) pretpostaviti da je dobivena prosječna korelacija koeficijent pouzdanosti za svaku pojedinu česticu j te 3) primijeniti Spearman-Brownovu formulu u svrhu procjene pouzdanosti testa koji se sastoji od J dihotomnih čestica

koristeći njihove indekse lakoće (p_j) i težine (q_j) prilikom izračuna. S obzirom na to da se radilo o 20. deriviranoj formuli, koeficijent je nazvan Kuder-Richardsonov 20 koeficijent pouzdanosti (ρ_{KR20} ; Allen i Yen, 1979), a kasnije je nazvan koeficijent pouzdanosti α -20 ili skraćeno $\alpha(20)$, prema Cronbachu (1951).

$$\rho_{KR20} = \frac{J}{J-1} \left(1 - \frac{\sum_{j=1}^J p_j q_j}{var_x} \right) \quad (7)$$

Ako su čestice visoke homogenosti, ρ_{KR20} i *split-half* procjena pouzdanosti bit će podjednake, a ako su čestice heterogene, ρ_{KR20} imat će niže vrijednosti od *split-half* procjene pouzdanosti. Drugim riječima, ρ_{KR20} može biti smatran i prosjekom svih mogućih *split-half* podjela.

S obzirom na to da je izračun ρ_{KR20} u slučaju velikog broja čestica kompliciran, predložena je još jedna formula (Kuder-Richardsonov 21 koeficijent pouzdanosti; ρ_{KR21}), jednostavnija za izračun, ali manje precizna jer pretpostavlja da su indeksi lakoće svih čestica jednaki (μ).

$$\rho_{KR21} = \frac{J}{J-1} \left(1 - \frac{\mu(J-\mu)}{Jvar_x} \right) \quad (8)$$

Vrijednosti ρ_{KR21} u pravilu su niže od vrijednosti ρ_{KR20} , ali su razlike zanemarive ako su psihometrijske karakteristike zadovoljavajuće (Lord, 1959c). Derivacije ovih formula potakle su velik broj istraživača na razvijanje nadogradnji navedenih, kao i vlastitih varijanti formula (Kuder, 1991), a konačan je rezultat bio povećanje interesa za proučavanje pouzdanosti mjerenja.

Guttman (1945) je pod utjecajem Spearmana predstavio niz λ koeficijenata pouzdanosti, koji se odnose na procjenu donje granice pouzdanosti. Njegova formula pouzdanosti temelji se na sljedećem: 1) pogreška mjerenja eksplicitno je definirana za svakog sudionika u univerzumu ponovljenih mjerenja, iz čega proizlaze tri zasebna izvora varijacije: ponovljena mjerenja, sudionici i čestice, a nepouzdanost je definirana kao varijacija rezultata u ponovljenim mjerenjima, 2) čestice ne mogu zamijeniti ponovljena mjerenja, a u slučaju neovisnosti dvaju mjerenja, koeficijent pouzdanosti jednak je korelaciji rezultata iz navedenih mjerenja, 3) zbog teške ostvarivosti neovisnosti mjerenja bitno je usmjeriti se na informacije koje je moguće dobiti iz jednog mjerenja, što je procjena donje granice pouzdanosti. Prema tome, ovim su pristupom uklonjene pretpostavke o potrebi za mjerenjem u barem dvije vremenske točke i metodološke komplikacije koje proizlaze iz pretpostavke o neovisnosti mjerenja. Prema Guttmanu (1945) razvijeni koeficijenti predstavljaju sukcesivna poboljšanja procjene pouzdanosti utemeljene na navedenim prednostima, a logike svakoga pojedinog koeficijenta utjecale su na razvoj kasnije deriviranih koeficijenata.

Ako se u jednoj točki mjerenja uzme u obzir da postoji varijanca rezultata sudionika za svaku česticu j , kao i varijanca ukupnog rezultata svih sudionika (var_X), tada se na vrlo jednostavan način može doći do procjene donje granice pouzdanosti λ_1 .

$$\lambda_1 = 1 - \frac{\sum_{j=1}^J var_j}{var_X} \quad (9)$$

Međutim, iako je neovisnost mjerenja s obzirom na sudionike jednostavno postići, neovisnost mjerenja s obzirom na čestice⁵ nije realističan uvjet ako se radi o procjeni zajedničke, prave varijance. Prema tome, bolja je varijanta λ_1 koeficijenta pouzdanosti λ_2 , koja uključuje dvostruku sumu kvadrata $\frac{J(J-1)}{2}$ kovarijanci čestica j i preostalih čestica $j-1$ u jednoj točki mjerenja.

$$\lambda_2 = \lambda_1 + \frac{\sqrt{\frac{J}{J-1} 2 \sum_{j=1}^J kov_{j,j-1}^2}}{var_X} \quad (10)$$

Idući koeficijent (λ_3) deriviran je iz λ_2 radi lakšeg izračuna, odnosno izostavlja sumu kvadrata kovarijanci čestica, a preciznija je procjena pouzdanosti od λ_1 . Poznat je i pod nazivom Cronbachov α koeficijent jer je Cronbach (1951) zagovarao njegovo korištenje umjesto Spearman-Brownove formule.

$$\lambda_3 = \frac{J}{J-1} \lambda_1 \quad (11)$$

Odnos navedena tri koeficijenta je sljedeći: $\lambda_1 < \lambda_3 \leq \lambda_2$.

Koeficijent λ_4 sličan je *split-half* koeficijentu pouzdanosti utoliko što zahtijeva podjelu rezultata u dvije polovice sa zasebnim varijancama (var_1 i var_2). Predstavlja donju granicu pouzdanosti bez obzira na korištenu metodu podjele testa i ne zahtijeva pretpostavke o unidimenzionalnosti i τ -ekvivalentnosti⁶. Međutim, poželjno je podijeliti test na način da je koeficijent maksimaliziran jer to predstavlja najprecizniju procjenu donje granice pouzdanosti i stoga je nazvan maksimalnim

⁵ Neovisnost mjerenja s obzirom na sudionike ili neovisnost opažanja odnosi se na osiguravanje da mjerenje rezultata svakog sudionika u uzorku nije ni na koji način povezano ili pod utjecajem mjerenja rezultata drugih sudionika. Neovisnost se mjerenja s obzirom na čestice odnosi na impliciranost u formuli λ_1 koeficijenta da su čestice tretirane kao linearna kombinacija međusobno nepovezanih čestica (jer su u brojniku samo varijance, ali ne i kovarijance).

⁶ τ -ekvivalentnost odnosi se na uvjete u kojima su čestice jednako saturirane određenim konstruktom, ali njihove pogreške nisu jednakih vrijednosti.

split-half koeficijentom. Ako su varijance polovica jednake, λ_4 je numerički jednak iznosu koji bi se dobio primjenom Spearman-Brownove formule na tim polovicama.

$$\lambda_4 = 2 \left(1 - \frac{\text{var}_1 + \text{var}_2}{\text{var}_X} \right) \quad (12)$$

Odnos λ_3 i λ_4 je sljedeći: $\lambda_3 = E(\lambda_4)$, odnosno λ_3 predstavlja očekivanu, tj. prosječnu vrijednost λ_4 koeficijenta na temelju svih mogućih podjela. Prema tome, λ_4 predstavlja precizniju procjenu pouzdanosti od λ_3 .

Koeficijent λ_5 vrijednosti u dijagonali matrice opaženog Z broja kovarijanci zamjenjuje dvostrukim korijenom najveće moguće sume kvadriranih kovarijanci čestica. Ovaj koeficijent podrazumijeva poboljšanje u odnosu na λ_2 koeficijent utoliko što maksimalizira kovariranje čestica kao što je u λ_4 maksimalizirano kovariranje polovica.

$$\lambda_5 = \lambda_1 + \frac{2 \sqrt{\sum_{z=1}^Z \text{kov}_{j,j-1}^2}}{\text{var}_X} \quad (13)$$

Posljednji od koeficijenata, λ_6 , temelji se na multiploj korelaciji. Može biti izračunan pomoću sume varijanci pogreške procjene čestice j na temelju linearne multiple regresije preostajućih čestica $j-1$. λ_6 bit će veći od λ_2 i λ_3 kada su čestice relativno nisko korelirane, ali s visokim multiplim korelacijama.

$$\lambda_6 = 1 - \frac{\sum_{j=1}^J \text{var}_{e_{j,j-1}}}{\text{var}_X} \quad (14)$$

Cronbachov α koeficijent (Cronbach, 1951), zapravo Guttmanov λ_3 koeficijent i generalizirani ρ_{KR20} , jedan je od najvažnijih i najpervazivniji koeficijent pouzdanosti (Cortina, 1993; Schmitt, 1996; Sijtsma, 2009a; Yang i Green, 2011). Počiva na sljedećim pretpostavkama (Graham, 2006; Novick i Lewis, 1967; Raykov, 1997a, 1997b; Revelle i Zinbarg, 2009; Sheng i Sheng, 2012; Yang i Green, 2011): 1) esencijalna τ -ekvivalentnost⁷, odnosno pretpostavka da sve čestice mjere isti konstrukt na istoj skali uz jednak stupanj preciznosti, 2) normalitet distribucija i kontinuirana skala čestica, 3) normalitet distribucije ukupnih rezultata, 4) nekoreliranost pogrešaka čestica te 5) unidimenzionalnost. Moguće ga je interpretirati kao prosjek svih mogućih *split-half* koeficijenata pouzdanosti procijenjenih na temelju Flanagan-Rulonove koncepcije pouzdanosti i stoga

⁷ Za detaljnije pojašnjenje koncepta τ -ekvivalentnosti i njegova razlučivanja od paralelnosti i kongeneričnosti preporučeno je pogledati Lord i Novick (1968).

predstavlja očekivanu korelaciju dvaju testova jednake duljine koji mjere isti konstrukt različitim česticama uzorkovanih iz istog univerzuma čestica. Drugim riječima, u logici *split-half* procjene pouzdanosti ide korak dalje tretirajući svaku česticu kao subtest. Prednost je ovog koeficijenta u odnosu na prethodne *split-half* procjene pouzdanosti ta što je kao *split-half split-halfova* stabilniji jer mu vrijednost ne varira ovisno o metodi podjele testa u dva dijela. Ako je primijenjen na test koji se sastoji od dihotomnih čestica, dovest će do identične vrijednosti kao i ρ_{KR20} (Feldt, 1969). Postoje dvije varijante formule sirovoga Cronbachova α koeficijenta:

$$\alpha = \frac{J}{J-1} \left(1 - \frac{\sum_{j=1}^J var_j}{var_x} \right) \quad (15)$$

$$\alpha = \frac{J}{J-1} \left(1 - \frac{\sum_{j=1}^J kov_{j,j-1}}{var_x} \right) \quad (16)$$

U formuli (15) u brojniku se nalaze varijance čestica j (var_j), a u (16) kovarijance čestica j s preostalim česticama $j-1$ ($kov_{j,j-1}$), dok je u nazivniku ukupna varijanca opaženih rezultata (var_x). Obje formule dovode do jednakog rezultata. Međutim, varijable često nisu mjerene na istoj skali i stoga je u takvim slučajevima opravdano koristiti Pearsonov r koeficijent korelacije kao standardiziranu mjeru kovariranja, iz čega proizlazi potreba za standardiziranom varijantom formule.

Standardizirani α koeficijent još se naziva generaliziranom Spearman-Brownovom formulom, a od sirovog α koeficijenta razlikuje se po tome što umjesto kovarijanci upotrebljava prosječnu interkorelaciju čestica ($r_{j,j-1}$). Ipak, vrijednosti mu nisu jednake onima dobivenim primjenom Spearman-Brownove formule, osim u iznimnom slučaju u kojemu bi korelacije između polovica testa bile jednake prosječnoj interkorelaciji čestica. Neki od autora navode da se ovaj oblik koeficijenta može smatrati mjerom saturacije prvim faktorom (Hattie, 1985), što je ispravno isključivo kada su sve interkorelacije čestica jednake njihovoj prosječnoj interkorelaciji i tada je standardizirani α koeficijent izravno povezan s karakterističnim korijenom nerotirane prve glavne komponente (Kaiser, 1968). Pogrešno poimanje standardiziranog α koeficijenta kao mjere saturacije prvim faktorom impliciralo je njegovu povezanost s rangom matrice interkorelacija i, stoga, dimenzionalnošću (Hattie, 1985). Međutim, to je u suprotnosti i s onim što je navedeno u Cronbachovu izvornom članku i opovrgnuto je u istraživanjima koja su uslijedila (pogledati Cortina, 1993).

$$\alpha_{xx'} = \frac{J \bar{r}_{j,j-1}}{1 + (J-1) \bar{r}_{j,j-1}} \quad (17)$$

McDonald (1978, 1999) predlaže ω koeficijent (još poznat pod nazivom ω total; ω_T), na temelju faktorsko-analičkog pristupa uz pretpostavku da je ciljanu domenu moguće objasniti pomoću faktora u podlozi kovariranja čestica, u kojemu je suma faktorskih saturacija čestica ($\lambda_{j,k}$) podijeljena sumom sume faktorskih saturacija čestica j i komunaliteta čestica j . Temelji se na dekompoziciji varijance testa u četiri dijela: 1) generalni faktor g s varijancom zajedničkom svim varijablama, 2) grupne faktore k , 3) specifičnosti i 4) pogreške mjerenja (Revelle, 2016). Moguće ga je računati koristeći vrijednosti dobivene eksploratornom i konfirmatornom faktorskom analizom (Revelle, 2019; Zinbarg, Yovel, Revelle i McDonald, 2006). Ovim je pristupom ispravljena pristranost u procjeni pouzdanosti α koeficijenta, koja nastaje kada su narušene pretpostavke o τ -ekvivalentnosti i unidimensionalnosti.

$$\omega = \frac{\left(\sum_{j=1}^J \lambda_{j,k} \right)^2}{\left[\left(\sum_{j=1}^J \lambda_{j,k} \right)^2 + \left(\sum_{j=1}^J (1 - \lambda_{j,k}^2) \right) \right]} \quad (18)$$

Kada su pogreške korelirane ili podaci multidimenzionalni, procijenjen je doprinos svake dimenzije ukupnoj varijanci, pa je u tu svrhu razvijen tzv. hijerarhijski koeficijent ω_H , poseban oblik α koeficijenta (McDonald, 1985, 1999), koji je otporan na pristranosti kojima podliježe α koeficijent te koji omogućava uvid u dimenzionalnost skale ako je korišten u kombinaciji s ω koeficijentom (pogledati Revelle i Zinbarg, 2009; Zinbarg, Revelle i Yovel, 2007; Zinbarg, Revelle, Yovel i Li, 2005). Ovaj se koeficijent razlikuje od prethodnog ω koeficijenta utoliko što u brojniku koristi sumu saturacija na generalnom faktoru g ($\lambda_{j,g}$). Kada su podaci unidimenzionalni, vrijedi $\omega = \omega_H$.

$$\omega_H = \frac{\left(\sum_{j=1}^J \lambda_{j,g} \right)^2}{\left[\left(\sum_{j=1}^J \lambda_{j,k} \right)^2 + \left(\sum_{j=1}^J (1 - \lambda_{j,k}^2) \right) \right]} \quad (19)$$

Najveća donja granica (ρ_{NDG} ; Jackson i Agunwamba, 1977; Woodhouse i Jackson, 1977) izvedena je na temelju glavne pretpostavke KTT-a u $\text{kov}_X = \text{kov}_T + \text{kov}_E$, odnosno matrice kovarijanci među česticama koju je moguće razdvojiti u sumu kovarijanci čestica za prave rezultate i matricu kovarijanci čestica za pogrešku (ten Berge i Sočan, 2004). ρ_{NDG} ne zahtijeva unidimensionalnost, ali

zahtijeva nekorelirane pogreške (Green i Yang, 2009a). Predstavlja najnižu moguću pouzdanost s obzirom na matricu opaženih rezultata uz ograničenje da je suma varijanci pogrešaka maksimalizirana za pogreške koje nisu korelirane s ostalim česticama. Od ukupne varijance oduzima se sumu varijanci pogrešaka čestica, procijenjenu na temelju određenih algoritama (pogledati Bentler i Woodward, 1980; ten Berge, Snijders i Zegers, 1981).

$$\rho_{NDG} = 1 - \frac{\sum_{j=1}^J (\text{var}_{e_j})}{\text{var}_X} \quad (20)$$

Ten Berge i Zegers (1978) razvili su niz koeficijenata koji se odnose na procjene donjih granica pouzdanosti: μ_0 , μ_1 , μ_2 , μ_3 , itd. μ_0 se odnosi na α , a μ_1 na λ_2 koeficijent. Drugim riječima, pokazali su da su α i λ_2 koeficijenti elementi niza procjena donjih granica pouzdanosti, koje također zahtijevaju pretpostavku o τ -ekvivalentnosti. Predstavljaju beskonačan skup sukcesivnih poboljšanja koeficijenata, a poboljšanje je zanemarivo nakon prva tri koraka (μ_2), kada se vrijednosti koeficijenata prestaju povećavati.

$$\mu_2 = \left(\sum_{j=1}^{J-1} \sum_{m=j+1}^J \text{kov}_{j,m} + \frac{\sqrt{\sum_{j=1}^{J-1} \sum_{m=j+1}^J \text{kov}_{j,m}^2 + \frac{J}{J-1} \sum_{j=1}^{J-1} \sum_{m=j+1}^J \text{kov}_{j,m}^4}}{\text{var}_X} \right) \quad (21)$$

Revelleov β koeficijent (1979) ili *najlošija split-half procjena pouzdanosti* temelji se na hijerarhijskoj klaster analizi i kombinira psihometrijske principe s procedurama klasteriranja. Provođi se u sljedećim koracima: 1) utvrditi matricu sličnosti/korelacija među česticama, 2) pronaći najsličnije parove čestica/varijabli iz matrice, 3) kombinirati parove varijabli u novu kompozitnu varijablu, 4) izračunati sličnost nove kompozitne varijable s preostalim varijablama, 5) ponoviti korake 2 - 4 uzimajući u obzir početne varijable i kompozitne varijable te 6) prekinuti postupak kada više nije moguće kombinirati varijable ili kada je dosegnut postavljeni kriterij. Za razliku od α koeficijenta β koeficijent moguće je odrediti podjelom testa na dva dijela (duljina J_1 i J_2) na način da je kovarijanca između polovica minimalizirana ili da je suma varijanci unutar testova maksimalizirana. β koeficijent nije osjetljiv na duljinu testa i omogućava klasteriranje subtestova te stoga daje manje pristranu procjenu homogenosti testa od α koeficijenta. Uz to, moguće ga je koristiti kao pravilo za odluku o kombiniranju dvaju subtestova u jedan test višeg reda. Također, kako je α koeficijent prosjek *split-halfova*, vrijedi $\beta \leq \alpha$.

$$\beta = \frac{(J_1 + J_2)^2 \overline{\text{kov}}_{1,2}}{\text{var}_X} \quad (22)$$

Konačno, pouzdanost skale moguće je procijeniti i koristeći estimacije parametara strukturalnih modela (ρ_{SM} ⁸; Bentler, 2009; Green i Hershberger, 2000; Green i Yang, 2009a, 2009b; Raykov, 1997a, 1998; Raykov i Shrout, 2002) u slučaju konceptualno smislenog modela koji pristaje podacima (Green i Yang, 2009a), pri čemu je korištena definicija pouzdanosti kao paralelnih formi u razvoju koeficijenta pouzdanosti. Simulacijske su studije pokazale da su procjene stabilne ukoliko su indeksi pristajanja zadovoljavajući (Yang i Green, 2010). Točnije, pouzdanost iz okvira KTT-a moguće je odrediti i pomoću strukturalnih modela, u kojima je pravi rezultat suma procjena rezultata faktora (F_{ik}), odnosno procjena rezultata sudionika i na faktoru k , ponderiranih njihovim saturacijama ($\lambda_{V_j F_k}$) i reziduala (R_{ij}), pri čemu se V_{ij} odnosi na rezultat sudionika i na čestici j . Drugim riječima, pristup pouzdanosti putem strukturalnih modela razdvaja varijancu pravih rezultata u različite komponente varijance, a istraživač treba odlučiti koje komponente varijance pridonose pouzdanosti rezultata (Bentler, 2009). Kako bi se procijenilo ρ_{SM} , prvo je potrebno odrediti model koji pristaje podacima. Na temelju su dobivenog modela vrijednosti dobivene estimacijom parametara uzorka uvrštene kao vrijednosti u brojniku, a reproducirane varijance i kovarijance čestica uvrštene su kao vrijednosti parametara u nazivniku (Green i Yang, 2009b), što rezultira procjenom pouzdanosti.

$$\rho_{SM_L} = \frac{\sum_{j=1}^J \sum_{j'=1}^J \sum_{k=1}^K \sum_{k'=1}^K \lambda_{V_j F_k} \lambda_{V_{j'} F_{k'}} r_{F_k F_{k'}}}{\sum_{j=1}^J \sum_{j'=1}^J \sigma_{V_j V_{j'}}} \quad (23)$$

Green i Yang (2009b) prezentiraju metodu nelinearnog strukturalnog modeliranja za izračun pouzdanosti skala koje se sastoje od čestica na ordinalnoj skali, također uz definiciju pouzdanosti kao paralelnih formi. U svrhu procjene pouzdanosti potrebno je izgraditi nelinearni strukturalni model koji pristaje podacima. Na temelju su takva modela estimacije uzorka (pragovi⁹, faktorske saturacije, faktorske korelacije i polihoričke korelacije) uvrštene kao vrijednosti potrebne za izračun procjene pouzdanosti. U formuli se C odnosi na broj kategorija, dok su τ vrijednosti pragovi za kategorije c primijenjeni na kontinuirane čestice V_j^* kako bi se produciralo ordinalne rezultate na manifestnim česticama V_j . Kumulativna

⁸ Važno je napomenuti kako je u brojnim istraživanjima korišten termin ω za različite koeficijente pouzdanosti dobivene putem strukturalnog modeliranja i kada se ne radi o McDonaldovu koeficijentu te bi trebalo obratiti pozornost na formulu u podlozi navedenih. Također, za koeficijente pouzdanosti dobivene strukturalnim modeliranjem korišten je i naziv kompozitna pouzdanost. Za detaljnije informacije pogledati Cho (2016).

⁹ Pragovi se odnose na vrijednosti koje razgraničavaju kategorije kod dihotomnih ili politomnih indikatora u modelu, a kojima su u podlozi pretpostavljene kontinuirane normalno distribuirane latentne varijable. Za detaljnije objašnjenje pojma preporučeno je pogledati Muthén (1983).

vjerojatnost τ_{VjC} ili ekvivalenta paralelne forme τ_{VjC} je ϕ_1 , uz uvjet standardne normalne kumulativne distribucije, a ϕ_2 je združena kumulativna vjerojatnost τ_{VjC} i $\tau_{Vj'C}$ uz uvjet bivarijatne standardne normalne kumulativne distribucije s korelacijom/pouzdanošću $\rho_{VjVj'}$.

$$\rho_{SM_N} = \frac{\sum_{j=1}^J \sum_{j'=1}^J \left[\sum_{c=1}^{C-1} \sum_{c'=1}^{C-1} \phi_2 \left(\tau_{V_j c}, \tau_{V_j' c'}, \sum_{k=1}^K \sum_{k'=1}^K \lambda_{V_j^* F_k} \lambda_{V_j'^* F_{k'}} r_{F_k F_{k'}} \right) - \left(\sum_{c=1}^{C-1} \phi_1(\tau_{V_j c}) \right) \left(\sum_{c'=1}^{C-1} \phi_1(\tau_{V_j' c'}) \right) \right]}{\sum_{j=1}^J \sum_{j'=1}^J \left[\sum_{c=1}^{C-1} \sum_{c'=1}^{C-1} \phi_2(\tau_{V_j c}, \tau_{V_j' c'}, r_{V_j V_j'}) - \left(\sum_{c=1}^{C-1} \phi_1(\tau_{V_j c}) \right) \left(\sum_{c'=1}^{C-1} \phi_1(\tau_{V_j' c'}) \right) \right]} \quad (24)$$

Miskoncepcije i primjene α koeficijenta unutarnje konzistencije

Unatoč postojanju različitih metoda procjene pouzdanosti, istraživači kontinuirano ignoriraju psihometrijske spoznaje i ne integriraju ih u svoj metodološki inventar, što je odraz udaljenosti psihometrije i ostalih grana psihologije, koja se povećala od 1950-ih¹⁰. Isto tako, trenutna konvencija za procjenu pouzdanosti je nesistematična, što istraživačima otežava razumijevanje problematike i navodi na bezuvjetnu primjenu Cronbachova α koeficijenta (pogledati Cho, 2016). Vjerojatno nijedan koeficijent nije toliko često navođen u istraživanjima kao indikator kvalitete rezultata na testu kao α koeficijent, ali isto tako nijedan koeficijent nije toliko pogrešno shvaćen i upotrebljavan bez provjere pretpostavki (Sijtsma, 2009a). Mnogi su autori upozorili na pogrešnu interpretaciju α koeficijenta, odnosno zadovoljavajuće ili visoke vrijednosti koeficijenta uporno su korištene kao indikacija unidimenzionalnosti skale (Cortina, 1993; Schmitt, 1996; Sijtsma, 2009a). Osim toga, često je i pogrešno shvaćanje koje se odnosi na tretiranje unutarnje konzistencije i homogenosti kao dvaju istoznačnih pojmova u literaturi (pogledati Cortina, 1993), što je pridonijelo pogrešnoj upotrebi α koeficijenta. Međutim, α koeficijent je funkcija stupnja u kojemu čestice testa imaju visoke komunalitete i niske unikvitete, kao i funkcija njihove međusobne povezanosti (Cortina, 1993), što ne implicira unidimenzionalnost, odnosno homogenost u tom kontekstu (pogledati Green, Lissitz i Mulaik, 1977). McDonald (1981) te Yang i Green (2011) navode kako skala može biti multidimenzionalna s visokim α koeficijentom i obrnuto. Često je opravdanje da α koeficijent može biti indikator unidimenzionalnosti jer bi mu vrijednost bila veća da je skala unidimenzionalna, ali to nije slučaj (Trizano-Hermosilla i Alvarado, 2016). Postoje još dvije česte miskoncepcije o α koeficijentu: 1) jednom kada je određena vrijednost koeficijenta, poznata je pouzdanost testa u svim kontekstima, dok se pouzdanost zapravo odnosi na rezultate uzorka i o njoj uvijek treba izvještavati kao o pouzdanosti konkretnih rezultata određenog uzorka te 2) više vrijednosti uvijek su poželjnije, iako vrijednosti

¹⁰ Za detaljnije informacije i raspravu preporučeno je pogledati Borsboom (2006a, 2006b), Clark (2006) i Sijtsma (2006).

koeficijenta iznad .90 mogu upućivati na previsoke korelacije između čestica i suvišnost nekih od njih (Appelbaum i sur., 2018; Streiner, 2003).

Meta-analize vrijednosti α koeficijenata¹¹ sugeriraju da je prosječna vrijednost koeficijenta iznosila .77, a 75 % ih je veće od .70, dok ih je 14 % prelazilo .90 (Peterson, 1994). Što se tiče interpretacije, istraživači su koristili širok raspon kvalitativnih opisa vrijednosti sa svrhom interpretacije α koeficijenta (Taber, 2018): izvrsno (.93 - .94), snažno (.91 - .93), pouzdano (.84 - .90), robusno (.81), prilično visoko (.76 - .95), visoko (.73 - .95), dobro (.71 - .91), relativno visoko (.70 - .77), blago nisko (.68), razumno (.67 - .87), adekvatno (.64 - .85), umjereno (.61 - .65), zadovoljavajuće (.58 - .97), prihvatljivo (.45 - .98), dovoljno (.45 - .96), nezadovoljavajuće (.40 - .55) i nisko (.11). U prvoj verziji knjige Nunnally (1967) predlaže vrijednosti od .50 i .60 kao prihvatljive u ranim fazama istraživanja, .80 za temeljna istraživanja te između .90 i .95 u slučajevima u kojima se donosi odluke za pojedinca, primjerice u kliničkom kontekstu. Međutim, visoke vrijednosti α koeficijenta treba interpretirati uz oprez jer mogu upućivati na velik broj suvišnih čestica (Boyle, 1991; Streiner, 2003). U kasnijoj verziji knjige (Nunnally, 1978) predlaže vrijednost .70 kao najnižu zadovoljavajuću, što je postalo općeprihvaćeno (Schmitt, 1996), iako neopravdano jer navedena preporuka nije rezultat empirijskog istraživanja ili logičkog rezoniranja, već Nunnallyjeve intuicije (Cho i Kim, 2015). Neki istraživači izvještavali su o niskim vrijednostima (npr. $\alpha = .65$) kao prihvatljivima, a kao razlog navodili utjecaj broja čestica na koeficijent (Schmitt, 1996). Međutim, bez obzira na ovisnost α koeficijenta o broju čestica nije moguće donositi precizne zaključke o sudionicima na temelju skala s niskim pouzdanostima, neovisno o njihovoj duljini (Yang i Green, 2011). Bez obzira na razilaženja u trenutnoj literaturi, preporučeno je držati se određenih smjernica. Primjerice, arbitrarno određena granica od .70 preniska je, posebno zbog pristranosti koeficijenata, i poželjnija je granica od barem .75 ili .80, i za istraživački i za praktični kontekst. S druge strane, koeficijenti iznad .95 indikator su previše specifične skale, a ne zadovoljavajuće unutarnje konzistencije. Ako je unutarnja konzistencija između .90 i .95, nije preporučeno tretirati je kao previše specifičnu skalu, ali u tom slučaju potrebno je dodatnu pozornost pridodati mogućim kolinearnostima čestica. Pri tome također nije poželjno koristiti Cronbachov α koeficijent kao indikator unutarnje konzistencije, osim ako su zadovoljene sve pretpostavke za njegovu upotrebu, već bi prednost trebalo dati λ_2 koeficijentu ili, ako su zadovoljene pretpostavke za provjeru latentne strukture, ω koeficijentu. Nadalje, što je mjerni model kompleksniji, to je utjecaj akumulirane pogreške mjerenja izraženiji i zato bi uz porast broja konstrukata i indikatora kriterij prihvatljivih unutarnjih konzistencija trebao biti nešto stroži, pogotovo ako model sadrži

¹¹ U svrhu meta-analize koeficijenata unutarnje konzistencije razvijena je metoda generalizacije pouzdanosti. Za detaljnije informacije preporučeno je pogledati Vacha-Haase (1998), Vacha-Haase, Henson i Caruso (2002) te Bonett (2010).

višestruke medijacijske i multiplikativne odnose, kao što je slučaj kod moderacijskih efekata (pogledati npr. Cole i Preacher, 2014; Ledgerwood i Shrout, 2011).

Prema tome, postoje različite preporuke za ispravno korištenje α koeficijenta. Cortina (1993) i Schmitt (1996) navode tri smjernice. Kao prvo, treba razlučiti korištenje sirovog naspram standardiziranog koeficijenta. Standardizirana verzija primjerena je kada su korišteni standardizirani rezultati čestica za formiranje rezultata na skalama, dok je sirova primjerena kada je korišten sirovi rezultat¹². Kao drugo, α koeficijent je koristan za procjenu pouzdanosti kada je istraživač zainteresiran za unikvitete čestica, pa visoka vrijednost koeficijenta upućuje na to da je moguće velik udio varijabiliteta pripisati generalnim faktorima. S obzirom na to da visoka vrijednost koeficijenta ne sugerira unidimenzionalnost, prvi bi korak u utvrđivanju navedene bila primjena eksploratorne faktorske analize i utvrđivanje broja faktora metodom paralelne analize (Horn, 1965) ili primjena konfirmatorne faktorske analize kako bi se provjerilo unidimenzionalnost¹³ skale, koja je utvrđena ako su indeksi pristajanja zadovoljavajući i koja je smatrana superiornom eksploratornoj faktorskoj analizi zbog uvida u višestruke indekse pristajanja (Green i Yang, 2009a, 2015). Neki autori (Raykov, 1998) preporučuju konfirmatornu faktorsku analizu za provjeru unidimenzionalnosti jer omogućuje istovremenu provjeru koreliranosti pogrešaka i τ -ekvivalentnosti, dok eksploratorna faktorska analiza ne razlikuje slučajnu od sistematske pogreške i stoga nije opravdana. Treće (Cortina, 1993; Fan i Thompson, 2001; Schmitt, 1996), preporučeno je koristiti i standardnu pogrešku α koeficijenta, koja upućuje na raspršenje interkorelacija čestica te, ako je visoka, simptom je multidimenzionalnosti i velike pogreške uzorkovanja. U skladu s navedenim, Iacobucci i Duhachek (2003) predlažu primjenu samoizvlačenja¹⁴ u svrhu procjene intervala sigurnosti α koeficijenta, a simulacijskom su studijom pokazali prednosti njihova pristupa u odnosu na prethodne (pogledati Duhacek i Iacobucci, 2004). Naposljetku, uz prezentiranje α koeficijenta i njegove standardne pogreške, preporučeno je izvijestiti i o prosječnoj interkorelaciji čestica, koja može uputiti na multikolinearnost i bez koje nije moguće interpretirati vrijednost koeficijenta kao zadovoljavajuću ili nezadovoljavajuću. Vezano uz same vrijednosti interkorelacija čestica Kline (1979) navodi da bi čestice s interkorelacijama iznad .70 trebale biti izbačene jer skalu čine previše specifičnom, odnosno podižu pouzdanost nauštrb valjanosti. S druge strane, čestice s interkorelacijama do .30 upućuju na potencijalnu multidimenzionalnost. Što se tiče pretpostavke o normalnim distribucijama pravih rezultata i pogrešaka, rezultati su

¹² Sirovi rezultat odnosi se na slučajeve u kojima je ukupni rezultat neke skale zbroj rezultata pojedinih čestica navedene skale, uzimajući u obzir i potencijalnu prisutnost rekodiranih čestica.

¹³ U ovom je radu pitanje dimenzionalnosti pojednostavljeno. Za detaljniju obradu problematike preporučeno je pogledati McDonald (1981), Hambleton i Rovinelli (1986) te Mellenbergh (2019, str. 135-154).

¹⁴ Engleski termin je *bootstrapping*.

dvoznačni, a utjecaj asimetričnosti na procjenu pouzdanosti slabo je istražen (Trizano-Hermosilla i Alvarado, 2016), iako se istraživači gotovo redovito susreću s asimetričnim skalama (Micceri, 1989). Sheng i Sheng (2012) izvješćuju da leptokurtične distribucije dovode do pristranosti u smjeru smanjenja procjene, a platikurtične do pristranosti u smjeru povećanja procjene pouzdanosti. Obje pristranosti smanjivale su se povećanjem uzorka i prave pouzdanosti. S druge strane, Zimmerman, Zumbo i Lalonde (1993) zaključuju da je α koeficijent otporan na devijacije od normaliteta. Osim toga, pretpostavka o normalitetu često je narušena jer su uglavnom korištene diskretne skale i stoga je procjena pouzdanosti pristrana te Gadermann, Guhn i Zumbo (2012) i McNeish (2018) preporučuju korištenje polihoričkih kovarijanci ili korelacija prilikom izračuna α koeficijenta. Međutim, za razliku od narušene τ -ekvivalentnosti narušena pretpostavka o koreliranim pogreškama neće nužno dovesti do pristranosti α koeficijenta u predvidljivom smjeru (McNeish, 2018), a pristranost u smjeru povećanja može biti nezanemariva (npr. Green i Yang, 2009a; Komaroff, 1997; Maxwell, 1968; Miller, 1995; Raykov, 1998; Zimmerman i sur., 1993). Točnije, koeficijent može biti veći čak i za .38 u odnosu na pravu pouzdanost (Gu, Little i Kingston, 2013).

Nadalje, Sočan (2000) navodi određene kontekste u kojima je opravdano koristiti α koeficijent: 1) ograničenost alata za izračun pouzdanosti, 2) uzorak je malen¹⁵, 3) indeksi pristajanja τ -ekvivalentnog modela zadovoljavajući su ili su kovarijance među česticama podjednake. Također, α koeficijent može biti indikator viška čestica i previše specifične skale (Boyle, 1991). Osim toga, α koeficijent daje korisne informacije u nacrtima s heterogenim uzorcima i multikomponentnim instrumentima prilikom provjere njegove invarijantnosti u latentnim klasama i praćenja promjene njegove vrijednosti tijekom revizije instrumenta (Raykov i Marcoulides, 2015b, 2019; Raykov, Marcoulides, Harrison i Menold, 2019; Raykov, West i Traynor, 2015).

Kada i kako primijeniti ostale metode procjene unutarnje konzistencije?

U većini testova τ -ekvivalentnost nije zadovoljena, već se radi o kongeneričkim¹⁶ testovima (Geldhof, Preacher i Zyphur, 2014), odnosno mjere isti konstrukt, ali u različitom stupnju (Raykov, 1997a), što ne znači da bi čestice s nižim saturacijama trebalo izbaciti, nego da je narušena jedna od pretpostavki za korištenje α koeficijenta i stoga bi on bio pristran i preniske vrijednosti (Miller, 1995). To je slučaj čak i ako je pretpostavka narušena samo za jednu česticu (Raykov, 1997a) i

¹⁵ Iako u radu nije eksplicitno navedeno što je malen uzorak, pretpostavka je da se autor referirao na uvjete u kojima nisu zadovoljene pretpostavke za provjeru latentne strukture, odnosno kada je $N < 200$.

¹⁶ Za detaljnije informacije o tri vrste mjernih modela: paralelnim, τ -ekvivalentnim i kongeneričnim te njihovoj primjeni prilikom procjene pouzdanosti preporučeno je pogledati Graham (2006).

procjena zbog toga može biti 19 % niža (Green i Yang, 2009a), a pristranost je najizraženija kada se test sastoji od 10 ili manje čestica (Feldt i Qualls, 1996). U prijašnjim istraživanjima postoji konsenzus da bi trebalo promijeniti praksu korištenja α koeficijenta kao univerzalne procjene pouzdanosti, ali nije zaključeno koja bi ga alternativna metoda trebala zamijeniti (Bentler, 2009; Green i Yang, 2009b; Revelle i Zinbarg, 2009; Sijtsma, 2009a).

Primjerice, umjesto α koeficijenta može biti korišten λ_2 koeficijent (Sijtsma, 2009a). Podrazumijeva procjenu donje granice pouzdanosti i jednak je pravoj pouzdanosti ako je zadovoljena pretpostavka o esencijalnoj τ -ekvivalentnosti. Precizan je barem koliko i α koeficijent, ali otporan na multidimenzionalnost (Osburn, 2000).

Jedna od mjera koja također može biti primjerenija od α koeficijenta je λ_4 koeficijent. Međutim, u slučaju malog uzorka i velikog broja čestica λ_4 koeficijent izrazito je pristran u smjeru povećanja procjene pouzdanosti (Benton, 2015; Osburn, 2000; Shapiro i ten Berge, 2000). Callender i Osburn (1979) usporedili su λ_4 i α koeficijent te utvrdili da λ_4 redovito ima višu vrijednost, a razlika se kretala od .052 do .143 uz medijan od .080. Osburn (2000) je usporedio λ_4 s još devet koeficijenata i zaključio da je λ_4 najkonzistentniji, ali je to učinio na razini populacije i bez uvažavanja pogreške uzorkovanja (pogledati Thompson, Green i Yang, 2010). Benton (2015) je utvrdio da je pristranost ovog koeficijenta manje vjerojatna ako je vrijednost iznad $\lambda_4 = .85$, a uzorak veći od $N = 3000$, dok uzorak veličine $N = 1000$ može biti dovoljan ako je vrijednost koeficijenta barem $\lambda_4 = .90$.

Nadalje, ako je korišten α koeficijent, Sijtsma (2009a) preporuča uz njega navođenje ρ_{NDG} koeficijenta kako bi se na temelju njihove usporedbe moglo zaključiti u kojoj je mjeri α koeficijent pristran, pogotovo jer se smatra da je ρ_{NDG} najbliži pravoj pouzdanosti (Sijtsma, 2009b; ten Berge i Sočan, 2004). Postoje empirijski nalazi da je ρ_{NDG} manje pristran od α koeficijenta (Bentler i Woodward, 1980; ten Berge i Sočan, 2004), dok drugi sugeriraju da je pristraniji od ω koeficijenta (pogledati Revelle i Zinbarg, 2009). Sočan (2000) preporuča korištenje ρ_{NDG} -a ako: 1) je uzorak velik i potrebna je maksimalno precizna procjena pouzdanosti te je 2) planirano provesti korekciju za atenuaciju. Ako je uzorak malen, ρ_{NDG} koeficijent izrazito je pristran u smjeru povećanja procjene pouzdanosti (Shapiro i ten Berge, 2000), a neopravdano ga je koristiti ako je broj subtestova velik. Osim toga, pristran je u smjeru povećanja procjene pouzdanosti i ako je zadovoljen normalitet distribucija (Shapiro i ten Berge, 2000; Sijtsma, 2009a; ten Berge i Sočan, 2004). Prema tome, ovaj je koeficijent preporučeno koristiti kada postoji blago ili srednje odudaranje od normaliteta rezultata testa te kada je proporcija asimetričnih čestica velika jer je u takvim uvjetima neopravdano koristiti α i ω koeficijente (Trizano-Hermosilla i Alvarado, 2016).

Ako je korišten instrument s malim brojem čestica ($j \leq 4$) i širokim rasponom interkorelacija čestica, preporučeno je koristiti μ_2 koeficijent umjesto λ_2 , α ili ρ_{NDG} koeficijenta, jer je tada manje pristran (ten Berge i Zegers, 1978). Ipak, u praksi su

takvi slučajevi iznimno rijetki, a i μ_2 koeficijent je u simulacijskim studijama gotovo u potpunosti neistražen.

Što se tiče β koeficijenta, on može biti korišten prilikom konstrukcije testova u odluci o kombiniranju većeg broja subtestova u jedan test, odnosno ako mu kombinacija subtestova povećava vrijednost, tada je njihovo kombiniranje opravdano. Također, moguće je odrediti željeni kriterij vrijednosti β koeficijenta za prestanak kombiniranja. Revelle (1979) je demonstrirao kako α koeficijent može iznositi i preko .90 kod multidimenzionalnih konstrukata, dok kod β koeficijenta to nije slučaj. Ipak, ovaj koeficijent nije preporučeno koristiti samostalno kao indikator pouzdanosti jer koristi minimaliziranu kovarijancu polovica i zapravo je *split-half* koeficijent pouzdanosti. Ako bi se kombiniralo subtestove, preporučeno je provjeriti njihove unutarnje konzistencije nekim od drugih koeficijenata, čiji bi odabir ovisio o karakteristikama subtestova u skladu s prethodno navedenim pretpostavkama, a uz to provjeriti i vrijednost β koeficijenta. Također, primjena hijerarhijske klaster analize i, posebno, uvid u dendrogram mogu pružiti korisne informacije o grupiranju čestica i istraživaču olakšati postavljanje hipoteza o faktorima prije provedbe faktorske analize. Unatoč činjenici da rezultati hijerarhijske analize klastera i faktorske analize na istim česticama u konačnici mogu biti slični, preporučeno je preferirati konfirmatornu faktorsku analizu zbog uvida u indekse pristajanja i prilikom donošenja odluke o potencijalnom kombiniranju subtestova informacije dobivene na temelju dendrograma i β koeficijenta koristiti kao dodatne, a ne ključne.

Jedan od koeficijenata koji je u skladu s kongeneričkim modelom je ω koeficijent, koji je smatran senzibilnijim pokazateljem unutarnje konzistencije u odnosu na α koeficijent, ali i ostale (Green i Yang, 2009a; Raykov, 1997a; Revelle i Zinbarg, 2009; Zinbarg i sur., 2005, 2006, 2007). Dunn, Baguley i Brunsten (2014) nabrojali su glavne prednosti ω koeficijenta u odnosu na α koeficijent: 1) zahtijeva manji broj pretpostavki, koje su ujedno realističnije, 2) vjerojatnost inflacije i atenuacije procjene unutarnje konzistencije znatno je manja, 3) provjera vrijednosti ω koeficijenta ako se ukloni pojedinu česticu točnije će odražavati procjenu prave pouzdanosti u populaciji te 4) interval sigurnosti pouzdanosti ω koeficijenta preciznije odražava varijabilitet u postupku procjene pouzdanosti. Ako nije zadovoljena pretpostavka o τ -ekvivalentnosti, ω predstavlja manje pristranu procjenu pouzdanosti uz nižu pogrešku prognoze¹⁷ i postotak pristranosti (Trizano-Hermosilla i Alvarado, 2016), a u slučaju zadovoljene pretpostavke o normalitetu distribucija, ω je najmanje pristran od svih koeficijenata (Revelle i Zinbarg, 2009). Osim toga, Deng i Chan (2017) pokazali su da je u otprilike 80 % slučajeva ω koeficijent više vrijednosti od α koeficijenta i tada ga je opravdanije koristiti. Kako

¹⁷ Pogreška prognoze ili predikcije (engl. *Root mean square of error*; RMSE), predstavlja standardnu devijaciju reziduala, odnosno raspršenje vrijednosti u kriterijskoj varijabli za pojedinu vrijednost prediktorske varijable. U simulacijskim studijama uobičajeni su veliki simulirani uzorci kako bi se provjerilo uzorak procjena pouzdanosti u specificiranim uvjetima u odnosu na „kontrolne“ uvjete, pri čemu su za evaluaciju korišteni indikatori poput RMSE.

navode Green i Yang (2015), ω pruža detaljnije informacije o pouzdanosti od α koeficijenta, a uz to je vjerojatnost pogrešne interpretacije manja zbog transparentnog odnosa koeficijenta i faktorskog modela. Uz to, moguće je i izračunati ω_H , koji omogućava uvid u doprinos generalnih i grupnih faktora skali. Ako je utvrđeno da je skala multidimenzionalna, tada je preporučeno razdvojiti je u subskale i izračunati ω koeficijent za svaku subskalu (Sočan, 2000). S druge strane, Green i Yang (2015) upozoravaju na ograničenja ω koeficijenta, poput činjenice da se radi o koeficijentu unutarnje konzistencije i, time, neosjetljivost na varijancu pogreške koja proizlazi iz mjerenja u dvije vremenske točke. Ograničenje je i što se koeficijent temelji na faktorsko-analitičkom pristupu, što zahtijeva velike uzorke. Naposljetku, α i ω koeficijente treba smatrati gotovo identičnima ako su pretpostavke za primjenu α koeficijenta zadovoljene (Raykov i Marcoulides, 2015a).

Naposljetku, procjena pouzdanosti putem strukturalnog modeliranja opravdana je u slučaju kompleksnih mjernih modela i modela s velikim brojem fiksiranih parametara (Green i Yang, 2009a). Ovakav pristup omogućava provjeru pretpostavke o τ -ekvivalentnosti na način da se fiksira faktorske saturacije na jednake vrijednosti i promatra razliku u χ^2 -indeksu u odnosu na razliku u stupnjevima slobode. U uvjetima zadovoljene pretpostavke o τ -ekvivalentnosti i identičnih distribucija kategorija ordinalnih skala procjene pouzdanosti putem linearnih i nelinearnih strukturalnih modela međusobno se ne razlikuju, a procjena im se ne razlikuje ni od α koeficijenta. Razlika ne postoji ni kada linearni i nelinearni modeli imaju jednake reproducirane matrice kovarijanci čestica na ordinalnoj skali. S druge strane, kada su unidimenzionalne čestice s jednakim saturacijama pozitivno i negativno asimetrične distribuirane, procjene se pouzdanosti linearnih i nelinearnih strukturalnih modela razlikuju (Green i Yang, 2009b). Green i Yang (2015) navode i primjenu bifaktorskih modela, koji omogućavaju zadržavanje jednog konstrukta, ali istovremeno uvažavajući njegovu multidimenzionalnost te koji omogućavaju procjenu jesu li čestice unidimenzionalne u tolikoj mjeri da su njihovi rezultati interpretabilni (pogledati Reise, 2012). Neki preferiraju navoditi ω_H koeficijent uz ρ_{SM} , koji služi kao indikator proporcije varijance skale povezan s generalnim faktorom (Zinbarg i sur., 2005, 2006, 2007). Međutim, s obzirom na to da generalni faktor reprezentira konstrukt određene skale, ω_H u tom slučaju čini diskontinuitet pouzdanosti i valjanosti manje jasnom (Green i Yang, 2009a). Također, neki sugeriraju da je razlika procjene pouzdanosti u okviru strukturalnog modeliranja i α koeficijenta zanemariva (Peterson i Kim, 2013). Osim toga, izgradnja modela uključuje subjektivnu procjenu i stoga različiti istraživači mogu doći do različitih procjena pouzdanosti na istim podacima. Isto tako, ako je odabrani model podspecificiran, ρ_{SM} može biti izrazito pristran. Jedno je od rješenja prespecifikacija modela, ali to može dovesti do teškoća s estimacijom. Naposljetku, metode estimacije temelje se na pretpostavkama o distribucijama i veličini uzorka i stoga ako su mjerni modeli kompleksni, a faktori slabije definirani, ρ_{SM} zahtijeva velike uzorke ($N \geq 300$) (Raykov, 1997a; Yang i Green, 2010).

Zaključak i implikacije

Na temelju pregleda literature moguće je zaključiti kako koncept pouzdanosti mjerenja ima opsežnu povijest razvoja i stoga više potencijalnih pristupa procjeni. U početku je pouzdanost procjenjivana putem test-retest korelacije ili paralelnih formi, a kasnije najčešće koeficijentima unutarnje konzistencije. U tu je svrhu uglavnom korišten Cronbachov α koeficijent, iako uz pogrešnu interpretaciju i bez provjere pretpostavki koje zahtijeva. Stoga su navedene preporuke za ispravno izvještavanje i korištenje α koeficijenta i nekih njegovih varijanti, poput izvještavanja o njegovoj standardnoj pogrešci i prosječnoj interkorelaciji čestica, kao i uvjeti u kojima je opravdano koristiti druge koeficijente. Primjerice, Guttmanov λ_2 koeficijent uvijek je primjereniji od α koeficijenta u slučaju multidimenzionalnosti, a Guttmanov λ_4 koeficijent najprimjereniji je od svih koeficijenata u slučaju procjene pouzdanosti na rezultatima populacije, bez pretpostavki o unidimenzionalnosti i τ -ekvivalentnosti. Nadalje, ρ_{NDG} moguće je koristiti uz α koeficijent u svrhu uvida koliko je potonji pristran, ali i samostalno, kod asimetričnih ukupnih rezultata i velike proporcije asimetričnih čestica. U odnosu na ostale ω koeficijent je najsenzibilniji pokazatelj unutarnje konzistencije i gotovo ga je uvijek preporučeno koristiti umjesto α koeficijenta, a kada je korišten uz ω_H , pruža dodatne informacije. Ako je mjerni model kompleksan, opravdano je koristiti ρ_{SM} , pogotovo jer se pomoću strukturalnih modela može provjeriti τ -ekvivalentnost, koreliranost pogrešaka i unidimenzionalnost. Konačno, procjena pouzdanosti konstrukta vrlo je sofisticirana i treba joj pristupiti s posebnim oprezom te se pri tome oslanjati na više različitih metoda. Točnije, preporučeno je provjeravati latentnu strukturu uvijek kada su zadovoljene pretpostavke za faktorsku analizu i davati prednost ω koeficijentu, ali i provjeriti vrijednosti većeg broja koeficijenata unutarnje konzistencije, kao i njihove standardne pogreške i interkorelacije čestica.

Također, pregled literature upućuje na plodna područja za daljnje istraživanje pouzdanosti, ali i određene konceptualne probleme. Primjerice, pristranosti koeficijenata u slučaju različitog broja dimenzija, distribucija ukupnih rezultata i čestica, njihova pristranost u slučaju različitog stupnja koreliranosti pogrešaka te pristranost u slučaju različitih proporcija aberantnih vrijednosti i različitih vrsta i proporcija nedostajućih vrijednosti nisu dovoljno istraženi. Također, pojedini koeficijenti nisu istraženi u uvjetima različitog broja kategorija odgovora. Moguće je da je upotreba nekih koeficijenata opravdanija kod primjerice pet, a drugih kod sedam kategorija odgovora. Konsenzus o pretežnom korištenju skale s pet kategorija odgovora proizlazi iz istraživanja u kojima je unutarnja konzistencija procijenjena Cronbachovim α koeficijentom (npr. Lissitz i Green, 1975; Preston i Colman, 2000; Weng, 2004). Bilo bi korisno i utvrditi koje je koeficijente opravdano koristiti u uzorcima kod kojih su zadovoljene pretpostavke za provjeru latentne strukture i u uzorcima u kojima nisu, uzimajući u obzir i ostale uvjete. Osim toga, simulacijske studije nisu brojne i u njima je većinom istovremeno uspoređen mali broj koeficijenata unutarnje konzistencije, što otežava akumuliranje znanja o

opravdanosti izbora pojedinog koeficijenta u određenim uvjetima. Uz to, može se istaknuti kako su i nacrti u simulacijskim studijama relativno jednostavni i zaključni na simuliranim uvjetima često nisu prenosivi na praktična istraživanja. Prema tome, složeniji nacrti poput primjerice $5 \times 3 \times 4 \times 3 \times 3 \times 4 \times 5 \times 4$ (veličina uzorka od $N = 100$, $N = 200$, $N = 400$, $N = 700$ i $N = 1000$ \times pet, sedam i devet kategorija odgovora \times pet, 10, 15 ili 20 čestica \times jedna, dvije ili tri dimenzije \times niska, srednja i visoka interkorelacija čestica unutar dimenzije \times nekorelirane, blago, umjereno i visoko korelirane dimenzije \times normalna distribucija, izrazito asimetrična, umjereno asimetrična, platikurtična, leptokurtična \times udio asimetričnih čestica od 0, .2, .4 i .7) uz usporedbu α , λ_2 , λ_4 , ω i ω_H , ρ_{NDG} , μ_2 i ρ_{SM} koeficijenata rezultirali bi korisnijim informacijama o funkcioniranju pojedinog koeficijenta od većine dosadašnjih simulacijskih studija. Bitno je napomenuti i kako je pitanje odnosa indeksa pristajanja strukturalnih modela i pouzdanosti neistraženo¹⁸, a utvrđivanje navedenoga moglo bi biti koristan dodatan indikator prilikom evaluacije pristajanja strukturalnih modela. Nadalje, iako su pouzdanost ponovljenog mjerenja i pouzdanost unutarnje konzistencije jasno razlučeni, utvrđivanje njihova odnosa gotovo je u potpunosti zanemareno, a pouzdanost ponovljenog mjerenja vrlo je rijetko korištena. Čak su i u meta-analizama generalizacije pouzdanosti kao zavisne varijable uglavnom korišteni koeficijenti unutarnje konzistencije, i to isključivo Cronbachov α koeficijent. Prema tome, istraživanje pouzdanosti pogrešno je usmjereno utoliko što nastavlja lošu praksu bezuvjetne primjene α koeficijenta i na razini meta-analiza, koje bi u budućnosti trebalo provoditi uz primjenu i drugih koeficijenata unutarnje konzistencije kao zavisnim varijablama. Neistraženost odnosa pouzdanosti unutarnje konzistencije i pouzdanosti ponovljenog mjerenja dovodi u pitanje nalaze velikog broja istraživanja, pogotovo kada se uzme u obzir da je pouzdanost unutarnje konzistencije „zamjena“ za pouzdanost ponovljenog mjerenja i kao takva samo procjena donje granice pouzdanosti. Ovo je nezanemariv konceptualni problem jer obje pouzdanosti predstavljaju proporciju prave varijance, ali zahvaćaju različite varijance pogreške, što su strani istraživači uvažili. Međutim, u istraživanjima su pretežno korištena mjerenja u jednoj vremenskoj točki i interpretirana kao procjena donje granice pouzdanosti, ignorirajući da su opažene visoke i niske vrijednosti pouzdanosti unutarnje konzistencije u kombinaciji s visokim i niskim vrijednostima pouzdanosti ponovljenog mjerenja. Nadalje, termin 'donja granica procjene pouzdanosti' dodatno je upitan u slučaju većine koeficijenata jer su uočene pristranosti i u smjeru povećanja i u smjeru smanjenja. Također, ponovno se referirajući na simulacijske studije – u njima je, primjerice, za provjeru pristranosti koeficijenata u slučaju različitih oblika distribucija kao standard za usporedbu korištena normalna distribucija, što je i pretpostavka za neke od koeficijenata unutarnje konzistencije. Ovakva pretpostavka potencijalno navodi istraživače na pogrešne zaključke jer nisu nužno svi konstrukti normalno distribuirani, što treba uvažiti prilikom procjene njihove unutarnje konzistencije.

¹⁸ Preporučeno je pogledati raspravu u Stanley i Edwards (2016).

Primjerice, ako bi nalazi većeg broja istraživanja sugerirali asimetričnu distribuiranost konstrukta, bilo bi opravdanije njegovu unutarnju konzistenciju procjenjivati pomoću koeficijenta koji ne zahtijeva pretpostavku o normalitetu distribucije. Točnije, moguće je zaključiti kako su prilikom razvoja koeficijenata procjene pouzdanosti ignorirani empirijski podaci, što bi u budućnosti trebalo promijeniti. S druge strane, meta-analize generalizacije pouzdanosti imaju cilj utvrditi empirijsku distribuciju procjena pouzdanosti, što bi moglo rezultirati korisnim informacijama za donošenje odluke o graničnoj vrijednosti prihvatljive pouzdanosti. Osim toga, logičnije je pretpostaviti da je „prava“ pouzdanost varijabla, a ne fiksna vrijednost, što u tom slučaju vrijedi i za procjenu donje granice pouzdanosti. Implikacija je ovoga da informacije o pouzdanosti neke mjere iz prijašnjih vremenskih točaka zapravo uopće nisu poznate te da su procjene pouzdanosti većinom nepotpune. Prema tome, primjena Bayesova teorema, odnosno derivacija bayesijanske procjene pouzdanosti mogao bi biti idući korak u razvoju pouzdanosti mjerenja.

Literatura

- Allen, M. J. i Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Brooks/Cole Publishing Company.
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M. i Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3-25.
- Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, 74, 137-143. doi:10.1007/s11336-008-9100-1
- Bentler, P. M. i Woodward, J. (1980). Inequalities among lower bounds to reliability: With applications to test construction and factor analysis. *Psychometrika*, 45(2), 249-267. doi:10.1007/bf02294079
- Benton, T. (2015). An empirical assessment of Guttman's Lambda 4 reliability coefficient. *Springer Proceedings in Mathematics & Statistics*, 301-310. doi:10.1007/978-3-319-07503-7_19
- Bonett, D. G. (2010). Varying coefficient meta-analytic methods for alpha reliability. *Psychological Methods*, 15(4), 368-385. doi:10.1037/a0020142
- Borsboom, D. (2006a). The attack of the psychometricians. *Psychometrika*, 71(3), 425-440. doi:10.1007/s11336-006-1447-6
- Borsboom, D. (2006b). Can we bring about a velvet revolution in psychological measurement? A rejoinder to commentaries. *Psychometrika*, 71(3), 463-467. doi:10.1007/s11336-006-1502-3

- Boyle, G. J. (1991). Does item homogeneity indicate internal consistency or item redundancy in psychometric scales? *Personality and Individual Differences*, 12(3), 291-294. doi:10.1016/0191-8869(91)90115-r
- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3(3), 296-322. doi:10.1111/j.2044-8295.1910.tb00207.x
- Callender, J. C. i Osburn, H. G. (1979). An empirical comparison of coefficient alpha, Guttman's lambda-2, and MSPLIT maximized split-half reliability estimates. *Journal of Educational Measurement*, 16(2), 89-99.
- Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients. *Organizational Research Methods*, 19(4), 651-682. doi:10.1177/1094428116656239
- Cho, E. i Kim, S. (2015). Cronbach's coefficient alpha well known but poorly understood. *Organizational Research Methods*, 18(2), 207-230. doi:10.1177/1094428114555994
- Clark, L. A. (2006). When a psychometric advance falls in the forest. *Psychometrika*, 71(3), 447-450. doi:10.1007/s11336-006-1500-5
- Cohen, R. J. i Swerdlik, M. J. (2018). *Psychological testing and assessment* (9. izdanje). Boston: McGraw-Hill.
- Cole, D. A. i Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, 19(2), 300-315. doi:10.1037/a0033805
- Cortina, J. M. (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
- Crocker, L. i Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt Brace.
- Cronbach, L. J. (1943). On estimates of test reliability. *Journal of Educational Psychology*, 34(8), 485-494. doi:10.1037/h0058608
- Cronbach, L. J. (1947). Test „reliability“: Its meaning and determination. *Psychometrika*, 12, 1-16. doi:10.1007/bf02289289
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. doi:10.4135/9781412961288.n54
- Cronbach, L. J. i Furby, L. (1970). How we should measure “change”: Or should we? *Psychological Bulletin*, 74, 68-80. doi:10.1037/h0029382
- Cronbach, L. J., Gleser, G. C., Nanda, H. i Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Deng, L. i Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and Psychological Measurement*, 77(2), 185-203. doi:10.1177/0013164416658325

-
- Duhachek, A. i Iacobucci, D. (2004). Alpha's standard error (ASE): An accurate and precise confidence interval estimate. *Journal of Applied Psychology*, 89(5), 792-808. doi:10.1037/0021-9010.89.5.792
- Dunn, T. J., Baguley, T. i Brunsten, 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
- Fan, X. i Thompson, B. (2001). Confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement*, 61(4), 517-531. doi:10.1177/0013164401614001
- Feldt, L. S. (1969). A test of the hypothesis that Cronbach's alpha or Kuder-Richardson coefficient twenty is the same for two tests. *Psychometrika*, 34(3), 363-373. doi:10.1007/BF02289364
- Feldt, L. S. i Qualls, A. L. (1996). Bias in coefficient alpha arising from heterogeneity of test content. *Applied Measurement in Education*, 9(3), 277-286. doi:10.1207/s15324818ame0903_5
- Flanagan, J. C. (1937). A proposed procedure for increasing the efficiency of objective tests. *Journal of Educational Psychology*, 28, 17-21. doi:10.1037/h0057430
- Gadermann, A. M., Guhn, M. i Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research and Evaluation*, 17, 1-13.
- Geldhof, G. J., Preacher, K. J. i Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19, 72-91. doi:10.1037/e698612011-001
- Graham, J. M. (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. B. i Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling*, 7(2), 251-270. doi:10.1207/s15328007sem0702_6
- Green, S. B., Lissitz, R. W. i Mulaik, S. A. (1977). Limitation of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37(4), 827-838. doi:10.1177/001316447703700403
- Green, S. B. i Yang, Y. (2009a). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74, 121-135. doi:10.1007/s11336-008-9098-4
- Green, S. B. i Yang, Y. (2009b). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, 74, 155-167.
- Green, S. B. i 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

- Gu, F., Little, T. D. i Kingston, N. M. (2013). Misestimation of reliability using coefficient alpha and structural equation modeling when assumptions of tau-equivalence and uncorrelated errors are violated. *Methodology*, 9, 30-40. doi:10.1027/1614-2241/a000052
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255-282. doi:10.1007/bf02288892
- Hambleton, R. K. i Rovinelli, R. J. (1986). Assessing the dimensionality of a set of test items. *Applied Psychological Measurement*, 10(3), 287-302. doi:10.1177/014662168601000307
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139-164. doi:10.1177/014662168500900204
- Hogan, T. P., Benjamin, A. i Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, 60(4), 523-531. doi:10.1177/00131640021970691
- Horn, J. L. (1965). A rationale and a test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185. doi:10.1007/bf02289447
- Iacobucci, D. i Duhachek, A. (2003). Advancing alpha: Measuring reliability with confidence. *Journal of Consumer Psychology*, 13(4), 478-487. http://dx.doi.org/10.1207/S15327663JCP1304_14
- Jackson, P. H. i Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items I: Algebraic lower bounds. *Psychometrika*, 42(4), 567-578. doi:10.1007/bf02295979
- Kaiser, H. F. (1968). A measure of the average intercorrelation. *Educational and Psychological Measurement*, 28(2), 245-247. doi:10.1177/001316447103100303
- Kelley, T. L. (1921). The reliability of test scores. *The Journal of Educational Research*, 3(5), 370-379. doi:10.1080/00220671.1921.10879169
- Kline, P. (1979). *Psychometrics and psychology*. London, UK: Academic Press.
- Komaroff, E. (1997). Effect of simultaneous violations of essential τ -equivalence and uncorrelated error on coefficient α . *Applied Psychological Measurement*, 21(4), 337-348. doi:10.1177/01466216970214004
- Kuder, F. (1991). Comments concerning the appropriate use of formulas for estimating the internal-consistency reliability of tests. *Educational and Psychological Measurement*, 51(4), 873-874. doi:10.1177/001316449105100407
- Kuder, G. F. i Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151-160. doi:10.1007/bf02288391

-
- Ledgerwood, A. i Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology*, *101*(6), 1174-1188. doi:10.1037/a0024776
- Lissitz, R. W. i Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, *60*(1), 10-13. doi:10.1037/h0076268
- Loevinger, J. (Ur.). (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs*, *61*(4). doi:10.1037/h0093565
- Lord, F. M. (1955a). Estimating test reliability. *Educational and Psychological Measurement*, *15*(4), 325-336. doi:10.1177/001316445501500401
- Lord, F. M. (1955b). Sampling fluctuations resulting from the sampling of test items. *Psychometrika*, *20*, 1-22. doi:10.1007/bf02288956
- Lord, F. M. (1958). Some relation between Guttman's principal components of scale analysis and other psychometric theory. *Psychometrika*, *23*(4), 291-296. doi:10.1007/bf02289779
- Lord, F. M. (1959a). An approach to mental test theory. *Psychometrika*, *24*(4), 283-302. doi:10.1007/bf02289812
- Lord, F. M. (1959b). Statistical inferences about true scores. *Psychometrika*, *24*, 1-17. doi:10.1007/bf02289759
- Lord, F. M. (1959c). Test norms and sampling theory. *Journal of Experimental Education*, *27*(4), 247-263. doi:10.1080/00220973.1959.11010632
- Lord, F. M. (1984). Standard errors of measurement at different ability levels. *Journal of Educational Measurement*, *21*(3), 239-243. doi:10.1111/j.1745-3984.1984.tb01031.x
- Lord, F. M. i Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Maxwell, A. E. (1968). The effect of correlated errors on estimates of reliability coefficients. *Educational and Psychological Measurement*, *28*(3), 803-811. doi:10.1177/001316446802800309
- McDonald, R. P. (1978). Generalizability in factorable domains: Domain validity and generalizability. *Educational and Psychological Measurement*, *38*, 75-79. doi:10.1177/001316447803800111
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*, 100-117. doi:10.1177/001316447803800111
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum. doi:10.4324/9781315802510
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.Y.: L. Erlbaum Associates.

- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412-433. doi:10.1037/met0000144
- Mellenbergh, G. J. (2019). *Counteracting methodological errors in behavioral research*. Cham: Springer. doi:10.1007/978-3-030-12272-0
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, 105, 156-166. doi:10.1037/0033-2909.105.1.156
- Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling*, 2(3), 255-273. doi:10.1080/10705519509540013
- Muthén, B. O. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, 22(1-2), 43-65. doi:10.1016/0304-4076(83)90093-3
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3, 1-18. doi:10.1016/0022-2496(66)90002-2
- Novick, M. R. i Lewis, C. L. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 39, 1-13. doi:10.1007/bf02289400
- Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.
- Nunnally, J. C. (1978). *Psychometric theory* (2. izdanje). New York: McGraw-Hill.
- Nunnally, J. C. i Bernstein, I. H. (1994). *Psychometric theory* (3. izdanje). New York: McGraw-Hill.
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological Methods*, 5(3), 343-355. doi:10.1037/1082-989x.5.3.343
- Peterson, R. A. (1994). A meta-analysis of Cronbach's coefficient alpha. *Journal of Consumer Research*, 21(2), 381-391. doi:10.1086/209405
- Peterson, R. A. i Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology*, 98, 194-198. doi:10.1037/a0030767
- Preston, C. C. i Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1-15.
- Raykov, T. (1997a). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21(2), 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(4), 329-353. doi:10.1207/s15327906mbr3204_2
- Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement*, 22(4), 375-385. doi:10.1177/014662169802200407

-
- Raykov, T. i Marcoulides, G. A. (2015a). A direct latent variable modeling based method for point and interval estimation of coefficient alpha. *Educational and Psychological Measurement*, 75, 146-156. doi:10.1177/0013164414526039
- Raykov, T. i Marcoulides, G. A. (2015b). Scale reliability evaluation with heterogeneous populations. *Educational and Psychological Measurement*, 75(5), 875-892. doi:10.1177/0013164414558587
- Raykov, T. i Marcoulides, G. A. (2019). Thanks coefficient alpha, we still need you! *Educational and Psychological Measurement*, 79, 200-210. doi:10.1177/0013164417725127
- Raykov, T., Marcoulides, G. A., Harrison, M. i Menold, N. (2019). Multiple-component measurement instruments in heterogeneous populations: Is there a single coefficient alpha? *Educational and Psychological Measurement*, 79(2), 399-412. doi:10.1177/0013164417733305
- Raykov, T. i Shrout, P. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural Equation Modeling*, 9(2), 195-212. doi:10.1207/s15328007sem0902_3
- Raykov, T., West, B. T. i Traynor, A. (2015). Evaluation of coefficient alpha for multiple-component measuring instruments in complex sample designs. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(3), 429-438. doi:10.1080/10705511.2014.936081
- Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, 47(5), 667-696. doi:10.1080/00273171.2012.715555
- Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, 14, 57-74. doi:10.1207/s15327906mbr1401_4
- Revelle, W. (2016). *An introduction to psychometric theory with applications in R*. Preuzeto s [http:// personality-project.org/r/book/](http://personality-project.org/r/book/)
- Revelle, W. (2019). *Using R and the psych package to find ω* . Preuzeto s <http://personality-project.org/r/psych/HowTo/omega.pdf>
- Revelle, W. i Zinbarg, R. E. (2009). Coefficients alpha, beta, omega and the glb: Comments on Sijtsma. *Psychometrika*, 74, 145-154. doi:10.1007/s11336-008-9102-z
- Rulon, P. J. (1939). A simplified procedure for determining the reliability of a test by split-halves. *Harvard Educational Review*, 9, 99-103.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350-353. doi:10.1037/1040-3590.8.4.350
- Shapiro, A. i ten Berge, J. M. F. (2000). The asymptotic bias of Minimum Trace Factor Analysis, with applications to the greatest lower bound to reliability. *Psychometrika*, 65(3), 413-425. doi:10.1007/bf02296154
- Sheng, Y. i Sheng, Z. (2012). Is coefficient alpha robust to non-normal data? *Frontiers in Psychology*, 3, 1-13. doi:10.3389/fpsyg.2012.00034

- Sijtsma, K. (2006). Psychometrics in psychological research: Role model or partner in science? *Psychometrika*, 71(3), 451-455. doi:10.1007/s11336-006-1497-9
- Sijtsma, K. (2009a). On the use, misuse, and the very limited usefulness of Cronbach's Alpha. *Psychometrika*, 74, 107-120. doi:10.1007/s11336-008-9101-0
- Sijtsma, K. (2009b). Reliability beyond theory and into practice. *Psychometrika*, 74, 169-173. doi:10.1007/s11336-008-9103-y
- Sočan, G. (2000). Assessment of reliability when test items are not essentially τ -equivalent. *Advances in Methodology and Statistics*, 15, 23-35. Preuzeto s: <http://ams.sisplet.org/uploadi/editor/mz15socan.pdf>
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15, 72-101. doi:10.2307/1412159
- Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3(3), 271-295. doi:10.1111/j.2044-8295.1910.tb00206.x
- Stanley, L. M. i Edwards, M. C. (2016). Reliability and model fit. *Educational and Psychological Measurement*, 76(6), 976-985.
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80, 99-103. doi:10.1207/S15327752JPA8001_18
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273-1296. doi:10.1007/s11165-016-9602-2
- Ten Berge, J. M., Snijders, T. A. B. i Zegers, F. E. (1981). Computational aspects of the greatest lower bound to the reliability and constrained minimum trace factor analysis. *Psychometrika*, 46(2), 201-213. doi:10.1007/bf02293900
- Ten Berge, J. M. i Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69(4), 613-625. doi:10.1007/bf02289858
- Ten Berge, J. M. i Zegers, F. E. (1978). A series of lower bounds to the reliability of a test. *Psychometrika*, 43(4), 575-579. doi:10.1007/bf02293815
- Thompson, B. L., Green, S. B. i Yang, Y. (2010). Assessment of the maximal split-half coefficient to estimate reliability. *Educational and Psychological Measurement*, 70(2), 232-251. doi:10.1177/0013164409355688
- Thurstone, L. L. (1931). *The reliability and validity of tests*. Ann Arbor: Edwards.
- Trizano-Hermosilla, I. i Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and asymmetrical measures. *Frontiers in Psychology*, 7(769), 1-8. doi:10.3389/fpsyg.2016.00769
- Tryon, R. C. (1957). Reliability and behavior domain validity: Reformulation and historical critique. *Psychological Bulletin*, 54(3), 229-249. doi:10.1037/h0047980

-
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, 58, 6-20. doi:10.1177/0013164498058001002
- Vacha-Haase, T., Henson, R. K. i Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement*, 62(4), 562-562. doi:10.1177/001316402128775012
- Webb, N. M., Shavelson, R. J. i Haertel, E. H. (2006). Reliability coefficients and generalizability theory. *Handbook of Statistics*, 26, 81-124.
- Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64(6), 956-972.
- Woodhouse, B. i Jackson, P. H. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogenous items: II: A search procedure to locate the greatest lower bound. *Psychometrika*, 42(4), 579-591. doi:10.1007/bf02295980
- Yang, Y. i Green, S. B. (2010). A note on structural equation modeling estimates of reliability. *Structural Equation Modeling*, 17, 66-81. doi:10.1080/10705510903438963
- Yang, Y. i Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century?. *Journal of Psychoeducational Assessment*, 29(4), 377-392. doi:10.1177/0734282911406668
- Zimmerman, D. W., Zumbo, B. D. i 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. E., Revelle, W. i Yovel, I. (2007). Estimating ω_H for structures containing two group factors: Perils and prospects. *Applied Psychological Measurement*, 31(2), 135-157. doi:10.1177/0146621606291558
- Zinbarg, R. E., Revelle, W., Yovel, I. 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, 123-133. doi: 10.1007/s11336-003-0974-7
- Zinbarg, R., Yovel, I., Revelle, W. i McDonald, R. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_H . *Applied Psychological Measurement*, 30(2), 121-144. doi:10.1177/0146621605278814

Measurement Reliability in Psychology: Method Development, Infatuation with Cronbach's Alpha Coefficient, and Recommendations for Appropriate Reliability Estimation

Abstract

The purpose of this review article is to indicate the abundant history of the concept of measurement reliability and steep improvement in reliability estimation since the 1940s, which produced a large number of approaches to its estimation. At the beginning of the article, a brief historical review of the emergence of the concept of measurement reliability together with methods of reliability estimation is given. Afterwards, different coefficients of internal consistency are described in chronological order. Application of Cronbach's α coefficient without verifying necessary assumptions is criticized, and common misconceptions about α coefficient are elucidated. Necessary assumptions for use of Cronbach's α coefficient, its biases in case of unfulfilled assumptions, best practice of reporting, and potential applications of the coefficient beside reliability estimation are outlined. However, necessary assumptions extremely rarely hold in practice and application of some other coefficients is often more appropriate, such as λ_2 , λ_4 , ρ_{GLB} or ω , the least biased one in most cases. Guidelines for use of mentioned coefficients are offered. The most recent approach to reliability estimation is by using structural equation modelling due to its ability to conveniently assess aforementioned assumptions and its synergy with ω and ω_H coefficients, as well as its appropriateness in case of complex measurement models. Finally, conceptual issues on measurement reliability are discussed and guidelines for further research of measurement reliability are suggested.

Keywords: psychometrics, measurement reliability, Cronbach's alpha, McDonald's omega, test theory, structural equation modelling

Primljeno: 9.1.2020.