

CRONBACH-ALFA: VELE VAGY NÉLKÜLE?

T. KÁRÁSZ JUDIT^{1,2,3*} – NAGYBÁNYAI NAGY OLIVÉR⁴ –
SZÉLL KRISZTIÁN² – TAKÁCS SZABOLCS⁴

¹ELTE Eötvös Loránd Tudományegyetem Neveléstudományi Doktori Iskola,
Budapest, Magyarország

²ELTE Eötvös Loránd Tudományegyetem Neveléstudományi Intézet, Budapest, Magyarország

³Oktatási Hivatal, Budapest, Magyarország

⁴Károli Gáspár Református Egyetem Bölcsész- és Társadalomtudományi Kar,
Pszichológiai Intézet, Budapest, Magyarország

E-mail: t.karasz.judit@ppk.elte.hu

Béérkezett: 2020. december 14. – Elfogadva: 2021. október 1.

Háttér és célkitűzések: *Tanulmányunkban arra vállalkozunk, hogy egy olyan eljárást mutassunk be, melynek segítségével megalapozottabban használható az eddig is széles körben alkalmazott Cronbach-alfa mutató.*

Módszer: *Elemzésünkben egy olyan, egyszerűen elkészíthető szimulációs eljárás alkalmazását javasoljuk, melynek segítségével az adatainkhoz igazodva egy, az adott helyzetben elvárt együttes konzisztenciaszint (átlagos korrelációs szint) eléréséhez szükséges Cronbach-alfa viszonyítási pontot adhatunk.*

Eredmények: *Elemzésünk ugyanis rámutat arra, hogy nemcsak a kérdőívben használt ítemek számától, hanem a kérdőívben alkalmazott Likert-skála értékeinek számától is függővé kellene tenni a Cronbach-alfa mutató viszonyítási pontját. Szintén eredménye az elemzésünknek, hogy az alkalmazott formulával nem a hagyományosan meghatározott 0,7-es értéket kapjuk, hanem mind a Likert-skálától, mind pedig az ítemek számától, illetve az elvárt, ítemek közötti kapcsolati erősségtől függővé lehet tenni az elvárt Cronbach-alfa értéket.*

Következtetések: *A javasolt algoritmus segítségével a Cronbach-alfa mutatónak egy olyan alkalmazását lehet megvalósítani, mely alkalmas arra, hogy a kérdőívünkhöz (a válaszadási Likert-skálához) igazodva, illetve a skála ítemszáma alapján referenciapontot adhassunk e megbízhatósági mutatóhoz.*

Kulcsszavak: *Cronbach-alfa, szimuláció, küszöbérték*

* Levelező szerző

BEVEZETÉS

A Cronbach-alfa (Cronbach, 1951) mutató elterjedtségét jól mutatja, hogy skálák alkotásakor a kutatók, elemzők általában kiszámolják és értelmezik annak értékét. Az értelmezés során a legtöbb esetben egy általánosan elfogadott küszöbértéket is szoktak mellé illeszteni: a Cronbach-alfa mutató esetében elvártnak tekintjük, hogy ha egy kérdőívben skálákat, indexeket, faktorokat alkotunk, akkor az így létrehozott összevont mutatónak nagyjából 0,7-es alfa értéket kell mutatnia (pl. Peterson, 1994; George és Mallery, 2019). Ez némileg leegyszerűsített megközelítés, hiszen számos tényezőn múlik, mekkora értéket várhatunk el, ilyen a mért dimenzióknak tartalma (hétköznapi értelemben, hogy a mérendő konstruktum mennyire képlékeny vagy stabil jellegzetességgel bír), vagy éppen a mérésre szolgáló dimenzió itemkészletének nagysága, vagyis a skála hossza is befolyásolhatja azt, sőt még az elemzett minta mérete is (pl. Churchill és Peter, 1984; Ponterotto és Ruckdeschel, 2007). Ráadásul nem csak pontszerűen megállapított értékről beszélhetünk, hiszen a Cronbach-alfára konfidenciaintervallum is megadható (Fan és Thompson, 2001; Duhachek és Lacobucci, 2004).

A Cronbach-alfa alsó határértékének meghatározását mindenképpen ki kell egészíteni azzal, hogy ha igen magas az értéke (0,9–0,95 feletti), akkor általánosságban egyfajta redundanciát jelez a mutató (pl. Nagybányai Nagy, 2006). E magas érték esetében feltételezhető, hogy olyan kérdéseket vontunk össze, melyek kvázi duplikált kérdések, ezzel pedig mesterségesen megemeltük a Cronbach-alfa értéket (Ponterotto és Ruckdeschel, 2007). Ez azért jelent problémát, mert a Cronbach-alfa mutató általánosságban a kérdőívünkben alkalmazott skálánk, indexünk, összevont mutatónk konzisztenciájának/reliabilitásának mutatója, egyfajta „átlagos korrelációs szintnek” tekinthető, azaz a mutató értékéből még nem tudhatjuk, hogy a kérdőív mit mér, csak annyit, hogy azt legalább konzisztens módon teszi. Vagyis, ha mesterségesen (néhány kérdés kvázi duplikálásával, bizonyos kérdések túl erős összefüggését alkalmazva) felduzzasztjuk ezt a mutatót, akkor elfedheti olyan kérdések jelenlétét, melyek egyáltalán nem illeszkednek a mérni kívánt jelenséghez. Természetesen ezt egyéb eljárások alkalmazásával (például feltáró vagy megerősítő faktorelemzések segítségével) orvosolni kell, azaz egy megalapozott elemzés végére az ilyen itemek a legritkább esetben maradhatnak a kérdőívben.

Eljutva így a faktorelemzéshez, érdemes megemlíteni egy elvi szempontból analóg mutatót, a Kaiser–Meyer–Olkin (KMO) mutatót, mely általában kéz a kézben jár a Cronbach-alfa mutató alkalmazásával – de meglátásunk szerint bizonyos anyagokban a megközelítés árnyaltabb képet kellene, hogy kapjon (Napitupulu, Kadar és Jati, 2017; Cho és Kim, 2015). Hasonlóan a Cronbach-alfa mutatóhoz, a KMO érték szerepét, amely szintén egyfajta átlagos korrelációs mutatónak tekinthető, általánosságban nem minden esetben közelítik jó irányból a faktorelemzésnél. Azt szoktuk meg, hogy ha a mutatók magasak, akkor a vizsgált jelenségünk megfelelő – sőt, a KMO érték esetében még táblázatokkal is szoktunk találkozni (Glen, 2016), hogy mennyire jó a mutató értéke, mennyire lesz megfelelő a faktorelemzés számunkra. Sokan a Cronbach-alfa esetében a 0,7-es küszöbértéket is hasonló elv szerint használják, aminek részben az is az oka, hogy ilyen eredményekkel tűzdelt publikációkkal találkozunk legtöbbször.

A téma egyik metaanalitikus áttekintése során különféle folyóiratok több évtizedet felölelő cikkeinek összképe 0,7–0,8 körüli átlag Cronbach-alfákkal volt jellemezhető (Peterson, 1994). Tegyük hozzá, hogy a kiadatlan kéziratok egy részét valószínűleg éppen ezen küszöbérték miatt utasították el korábban.

Matematikai nézőpontból azonban e megfogalmazásnak épp a fordítottját érdemes használni: ha egy itemegyüttes esetében akár a Cronbach-alfa érték, akár a KMO érték nem éri el a megfelelő szintet, akkor szakmai szempontból még nem lesznek jók, nem lesznek stabilak az elemzéseink. Arra viszont semmilyen garancia nincs, hogy akár matematikai, akár szakmai oldalról a magasabb értékek megbízható eredményeket hoznának. Tehát a Cronbach-alfa esetében – hiszen tanulmányunknak ez a fókusza – nem helyes az az értelmezés, hogy a 0,7-es vagy annál magasabb értékek esetében a skálánk, indexünk, összevont mutatónk megfelelő megbízhatóságú lenne. Ellenben: ha alacsony szinten van a Cronbach-alfa, akkor azt feltételezhetjük, hogy legalábbis egy-egy item kilóg a sorból, további munkánk lesz a megbízhatóság elérésével. Összességében tehát elmondható, hogy a Cronbach-alfa magas értéke annyit jelez, hogy jó úton járhatunk, ugyanakkor tudni kell, hogy a mutató gond nélkül tud magas értéket mutatni úgy is, hogy egymáshoz csak igen lazán kapcsolódó itemeket alkalmazunk.

Fontos megemlíteni, hogy a Cronbach-alfa mutató és a faktorelemzések egy területen való említése már önmagában is némi ellentmondást vet fel. A Cronbach-alfa mutató ugyanis eleve olyan esetekben használható és használandó, amikor a vizsgálandó struktúra egyetlen faktort jelez, egyetlen komponens mérésére szolgál (Webb, Shavelson és Haertel, 2006). Azaz akkor van létjogosultsága a Cronbach-alfa (vagy ómega vagy egyéb, egyfaktoros megbízhatóságot ellenőrző mutatók) kiszámításának, amikor a faktorelemzés egyetlen látens faktor jelenlétét jelzi. Ez már csak azért is fontos, mert általánosságban igaz az, hogy e két módszert nem is ugyanazon elvek mentén kell használnunk, nem ugyanazon esetekben van létjogosultságuk.

Tanulmányunkban arra vállalkozunk, hogy rámutassunk a Cronbach-alfa mutató és értékeinek kellő kritikával történő alkalmazására, alkalmazhatóságára. Ennek érdekében egyfajta algoritmust adunk ahhoz, hogy a vizsgálatok során használt Cronbach-alfa mutatókat milyen standard értékekhez érdemes viszonyítani. Elemzésünkben a konzisztencia általános szintjeként tehát nem a Cronbach-alfa egy adott referenciaértékét fogjuk vizsgálni, hanem az itemek közötti átlagos korrelációs szintet fogjuk rögzítettnek tekinteni. Ezzel azt is javasoljuk, hogy adott tudományterületeken vagy adott mérőeszközök esetében bátran vállaljuk fel azt a döntést, hogy egy-egy kérdőív adott skáláján mi magunk mondjuk ki az általunk elvárt konzisztenciaszintet. Például területtől függően előre fixáljuk, hogy az itemek közötti átlagos korrelációs szint legyen legalább 0,5 (vagy ennél magasabb/alacsonyabb). A mérés tartalmi jellegetől, vagyis a mérendő konstruktum jellegzetességeitől is függ, mennyire könnyű azt mérhetővé tenni. Míg például az intelligenciamérésnél, ahol az egyik legszilárdabb pszichológiai entitással van dolgunk (sőt egyesek – pl. Neisser és munkatársai [1996] – inkább biológiai konstruktumként tárgyalják), érthető módon a 0,8–0,9 érték egyáltalán nem ritka, ugyanakkor a személyiségvonásoknál már 0,5–0,7 körüli, míg az attitűdskáláknál pedig jó esetben 0,5 feletti a szakmai alapon elvárható Cronbach-alfa szint (Horváth, 1997, 192).

Cronbach-alfa körüli viták: pró és kontra érvek

A Cronbach-alfa körüli viták nem új keletűek. Mindenképpen a mutató alkalmazása mellett szól, hogy egyszerű számítani, gyors, ráadásul minden lényeges matematikai statisztikai programcsomagban (még az ingyenes változatokban is) elérhető. Utóbbi azért is fontos szempont, mivel a vizsgálatok esetében komoly hátrány lehet, ha rendkívül korlátozott a megfelelő, letesztelt és megbízható megvalósításokhoz való hozzáférés.

Szintén a mutató mellett szól az, hogy több általánosítása és korrekciója is ismert (Cho, 2016): ómega, glb, tau, KR-20 stb. Ezek az alternatívák számos esetben szintén széles körben elérhetők, illetve különböző kérdőívtípusokra optimalizáltak. Ezen általánosított mutatók (pl. KR_20; Eklou és Quainoo, 2019) ad absurdum adaptív tesztekhez is használhatók. A jóval Cronbach előtt publikált ún. KR-20 formula (Kuder és Richardson, 1937) – amely az alfa mutatóval lényegében megegyezik – volt az az évtizedekig várt áttörés, amely felváltva a Brown (1910) és Spearman (1910) által felállított tesztfelezési technikát, innovatív megközelítése révén a megbízhatósági együtthatók új korszakát nyitotta meg.

Számos tanulmány vizsgálta, kritizálta és módosította az eredeti Cronbach-alfa metódust, illetve annak hibás alkalmazását (Bentler, 2008; Cho és Kim, 2015; Cortina, 1993; Green, Lissitz és Mulaik, 1977; Green és Yang, 2009; Hunt és Bentler, 2015; McDonald, 1981; Osburn, 2000; Revelle és Zinbarg, 2008; Sijtsma, 2008, 2015; Van der Ark, Van der Palm és Sijtsma, 2011; Yang és Green, 2011). A Cronbach-alfa mutató téves értelmezései közül az egyik legjellemzőbb típust a mérési dimenzió homogenitásáról levont fals következtetések képezik. Fontos tudatosítani, hogy önmagában a Cronbach-alfa mutató által jelzett magas belső konzisztencia nem teremt biztos alapot arra, hogy a skálát alkotó itemeket homogén halmazként interpretáljuk – ahogy erre több tanulmány is rámutat (pl. Schmitt, 1996). Például egy 6 tételes skálában az itemek közötti 0,5-es interkorreláció éppúgy vezethet 0,8-es Cronbach-alfához, mint ha két kupacban hármassával összeálló itemblokkon belüli 0,8-es korrelációk mellett egyébként csak 0,3-es itemek közötti együtt járások adódnának. Tehát a Cronbach-alfát homogenitásiindex helyett sokkal inkább belső konzisztenciamutatónak célszerű tekinteni.

Ez nem véletlen, hiszen a Cronbach-alfa alapvetően olyan egydimenziós konstruktmok (skálák, indexek, faktorok stb.) vizsgálatát teszi lehetővé, amelyek minden eleme azonos (faktor)súllyal (tau-ekvivalencia) rendelkezik, és ahol korrelálatlanok a mérési hibák, vagyis torzít minden olyan esetben, amikor ezek a feltételek nem állnak fenn. Beláthatjuk, hogy a legtöbb konstruktm esetében ezek a feltételek nem valósulnak meg, vagyis a Cronbach-alfa nagy eséllyel pontatlanul becsli meg a konstruktm belső megbízhatóságát. Ezért a reliabilitás tesztelésére több szerző is jobb alternatívának tartja a McDonald (1999) által a faktoranalízis keretrendszeréből kiindulva javasolt ómega (ω) együtthatót, amely alapvetően azt fejezi ki, hogy az általános és specifikus faktorok együttesen mennyire megbízhatóan mérik a kapott összpontszám alapján a konstruktmot, vagyis hogy mekkora az a hányad, amely a konstruktm összpontszámának varianciájából a közös faktorokkal megmagyarázható (McDonald, 1999; Dunn, Baguley és Brunson, 2014). Természetesen a mutatóval az egyes aldimenziók meg-

bízhatósága is tesztelhető, mely esetben az aldimenziók pontszámának varianciáját és a közös faktorok által lefedett varianciát vetik össze egymással.

Az ómega tulajdonképpen korrigálja a Cronbach-alfa torzítását abban az esetben, ha a tau-ekvivalencia feltételezése – miszerint a konstruktumot alkotó változók faktorsúlyai azonosak – sérül (Dunn, Baguley és Brunsten, 2014). Amennyiben a mérési hibák korrelálnak, vagy egynél több látens dimenzió van az adatokban, akkor egyes (akár általános, akár specifikus) faktorok, illetve dimenziók a teljes variancia magyarázatához való hozzájárulását – vagyis azt, hogy ezek mennyire megbízhatóan mérik a konstruktumot – az úgynevezett hierarchikus ómega (ω_h) mutatja meg, mely mutató lehetővé teszi a többdimenziós adatokon számolt Cronbach-alfa torzításának kijavítását (Zinbarg, Revelle, Yovel és Li, 2005; Revelle és Zinbarg, 2008; Trizano-Hermosilla és Alvarado, 2016). Az ómega és a hierarchikus ómega együttthatók egyenértékűek az egyszimulációs adatokban (Trizano-Hermosilla és Alvarado, 2016).

Számos elemzés szerint az ómega pontosabb becslést ad a több változóból álló konstruktumok megbízhatóságának tesztelésére (Zinbarg, Revelle, Yovel és Li, 2005; Revelle és Zinbarg, 2008; Trizano-Hermosilla és Alvarado, 2016; Hayes és Coutts, 2020). Hazai elemzés is rámutatott már arra, hogy a belső konzisztencia becslésére széles körben alkalmazott Cronbach-alfa mutató komoly módszertani dilemmákat vet fel, ezért mindenképpen javasolt a mutató korlátainak hangsúlyozása és alternatív mutatók alkalmazása (pl. hierarchikus ómega együtttható) (Rózsa és mtsai, 2019). Peters (2014) is arra a következtetésre jut, hogy a skáladiagnosztikák javítása érdekében célszerű többféle módon, többféle mutató segítségével ellenőrizni a skálák megbízhatóságát.

Érdekes ezen a ponton megjegyezni, hogy maga Cronbach már 1943-ban kritikával illette az addig használt reliabilitási mutatókat, hiszen már a 40-es évekre tudott volt, hogy a reliabilitási mutatók alulbecsülték a valódi reliabilitás mértékét. Vélhetően ennek hatására is írta meg 1951-es dolgozatát (Cronbach, 1951), melyben mind a KR-20, mind a split-half, mind egyéb korrekciós módszerekkel precíz matematikai alapokon hasonlítja össze a mutatóját, vagyis alapvetően ő maga is az 1937-es KR-20-as mutatóból indult ki, és ő maga javasolta e mutató alfa néven való alkalmazását, módosítását. Cronbach ebben az írásában számos táblázatot közöl, melyben különböző átlagos itemek közötti korrelációkkal, válaszadók számával és itemszámokkal számolva mutatja be az alfa mutató működését – összehasonlítva például a split-half és a KR-20-as, általánosításnak tekinthető mutatóval. Ugyanakkor sehol nem lelhető fel olyan megjegyzés, ahol Cronbach azt javasolná, hogy 0,7-es vágópontot használjunk az alfa mutatóhoz. Cronbach már ebben a cikkében is számos kritikával illette az alfa mutatót, utolsó anyagában (Cronbach és Shavelson, 2004) pedig egyértelműsíti, hogy a Cronbach-alfa mutatót csak mint egy teljes elemzés része érdemes vizsgálni, hiszen azóta számos egyéb reliabilitásvizsgálati módszer látott napvilágot.

A kérdés tehát az, hogy ha már maga Cronbach is kellő kritikával szemlélte az alfa mutatót, akkor miként lehetséges, hogy e mutató esetében az általa használt számos táblázat és formula helyett, 50–70 év elteltével (és számos új eljárás, mutató megjelenése után) sok cikkben még mindig az a hivatkozási alap, hogy „ha a Cronbach-alfa értéke nagyobb, mint 0,7, akkor minden rendben van”? Természetesen most szándékosan sarkítjuk a kérdés felvetését, hiszen egy-egy kérdőív vizsgálatánál ma már szinte alapelvárásnak tekinthető akár egy feltáró, de még inkább egy megerősítő faktor-

elemzés alkalmazása és bemutatása. Sőt, ezek mellett megjelenik minden esetben a szakmai kritika, kérdésvetetés is. Érdeemes kiemelni azonban, hogy már Cronbach is javasolt és alkalmazott olyan metódusokat (hozzátéve, hogy a számítógépek kapacitása jelentősen javult az 1951-es vizsgálatokhoz képest), melyeket jelen elemzésünkben részletesen bemutatunk és értelmezünk.

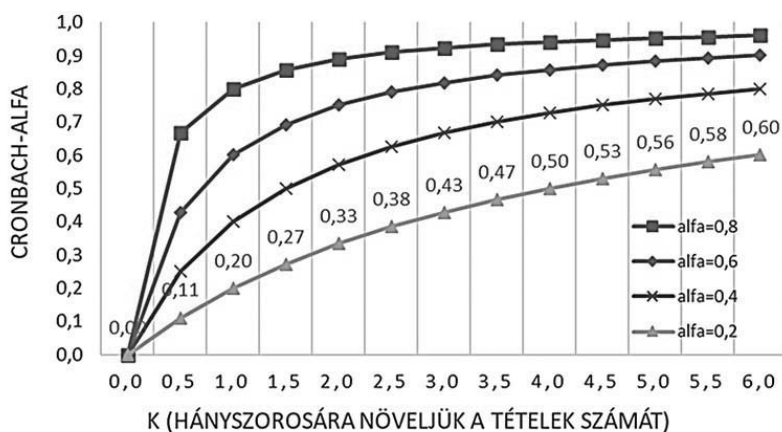
A formula és ami mögötte van

Ahhoz, hogy a Cronbach (1951) által javasolt formulák alapján számított eredményeket érdemben vizsgálni tudjuk, érdemes magára a formulára is vetni egy pillantást. Tegyük fel, hogy az összevontó itemek száma az adott, összevont mutatóban k . Legyen továbbá s_i^2 az egyes itemek varianciája, míg s^2 a teljes variancia. Ekkor a formula:

$$\rho = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k s_i^2}{s^2} \right).$$

Felhívjuk a figyelmet arra, hogy Cronbach már többször citált 1951-es cikke számos, ennél sokkal komolyabb formulát tartalmaz. Amit azonban most szeretnénk kicsit jobban górcső alá venni, az a fenti formula.

Tegyük fel, hogy elég sok itemet szeretnénk összevonni egy skálában. Ekkor a formulában használt első tört $\frac{k}{k-1}$ értéke igen közel lesz 1-hez. Tulajdonképpen ez felfogható úgy is, mint egy „általános” korrekciós tényező (a korábban jelzett alulbecslés „felfelé irányuló korrekciója miatt, hiszen ez a tényező mindig nagyobb, mint 1”): minél kevesebb itemmel dolgozunk, annál jobban „korrigál” az érték, tehát annál inkább terel el minket 1-től (20 item esetében $20 / 19 = 1,05$, 11 item esetében $11 / 10 = 1,1$, míg 3 item esetében $3 / 2 = 1,5$). Az 1. ábrán látható, miként növekszik meg a teszt hosszának többszörözésével a reliabilitás értéke. Ez azt is jelenti, hogy a sok kérdést tartalma-



1. ábra. Az itemszám-többszörözések vonalai, melyek a kezdeti reliabilitáshoz (alfa) társuló növekvő reliabilitási értékeket (R) eredményezik (Gulliksen [2013, 78] alapján)

zó skálák esetében ez a fajta alulbecslés már nem nyom annyit a latban, kisebb lesz a korrekciós mérték.

A szorzat második tagjának értelmezéséhez tegyük fel, hogy a kérdőívünk itemei között mérhető varianciával rendelkezünk. Ez azt jelenti, hogy a kérdőív válaszai nem „konstans” válaszok, hanem a válaszadók egymástól is különböző válaszokat adnak, valamint egy-egy válaszoló a különböző itemekre is eltérő válaszokat ad (ezek nem túlságosan erős feltételezések, ha kérdőívekkel dolgozunk). Másként fogalmazva:

1. azokat az itemeket, amelyekre mindenki ugyanazt válaszolja, töröljük;
2. azokat a válaszadókat, akik mindenre ugyanazt az értéket (pl. 3-ast) adják, szintén töröljük.

Ezen korrekció után a válaszainkon belül kétfajta „variabilitás” jelenik meg: egyik oldalról az alanyaink egymáshoz képesti válaszainak eltérései, másik oldalról pedig a válaszadók saját válaszai közötti különbségek. Az 1. táblázatban ez úgy értelmezhető, hogy az itemek varianciája az oszloponkénti variancia, míg a soronkénti különbségek mutatják a válaszadók válaszaiban lévő eltéréseket.

1. táblázat. Adatok hagyományos elrendezése egy adatállományban

Válaszadók/válaszok	item1	item2	...	item(k)
Első válaszadó				
Második válaszadó				
...				
(N) válaszadó				

A fenti képlet egy kis átalakításával jól interpretálható, hogy mit is mond a formula:

$$\rho = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k s_i^2}{s^2} \right) = \frac{k}{k-1} \left(\frac{s^2 - \sum_{i=1}^k s_i^2}{s^2} \right).$$

Ha ismerjük a teljes varianciát (azaz tudjuk, hogy összességében az összes válasz mennyire variábilis), és ebből elvesszük azt a mennyiséget, amely az itemek varianciája (ez éppen a szummáció), megkapjuk azt a mennyiséget, mely ettől független: a válaszadók varianciáját, tehát a válaszokban lévő konzisztenciát. Csak akkor beszélhetünk megbízható, erős belső konzisztenciával bíró tesztről, ha az egyes itemek ugyanabba az irányba mérnek, tehát ugyanannak a látens konstruktnak a leképeződései. Mivel ugyanarra a „dologra” kérdezzük rá, csak más-más aspektusból, így elvárható, hogy az egyéni válaszok is nagyjából konzisztensen alakuljanak, egy irányba mutassanak. Ha nincs meg ez a konzisztencia a tételek között, tehát valamelyik item kilóg a többi közül, akkor az vészesen leronthatja a Cronbach-alfa értékét.

Az 1. táblázat esetében ez annyit mond, hogy a tört számlálója nem más, mint a soronkénti varianciák megjelenése, azaz hogy a vizsgálati alanyaink válaszai soronként mennyire konzisztensek, mennyire válaszolják „ugyanazt” a különböző itemekre. Ezt mérjük hozzá a teljes varianciához.

Alapvetően azt szeretnénk, hogy ez a mutató magas legyen, kérdés tehát, hogy mikor lesz ez a mutató közel az 1-hez. A válasz tulajdonképpen egyszerű: az érték akkor lesz közel az 1-hez, ha a mutatónk számlálója és nevezője közel megegyezik. Azaz $s^2 - \sum_{i=1}^k s_i^2 \approx s^2$ mértékű, ami egyben azt is jelenti, hogy

$$\sum_{i=1}^k s_i^2 \approx 0.$$

Azaz az itemek önmagukon belül – a teljes varianciához képest – lényegében keveset szórnak, és a valódi szórást csak a vizsgálati alanyok közötti különbségek hordozzák. Összefoglalva:

1. Egy-egy vizsgálati alany válaszai között nincsen nagy különbség (az 1. táblázatban soronként kicsi a szóródás).

2. Egy-egy item önmagán belül keveset változik a teljes varianciához képest (az 1. táblázatban oszloponként sincsenek nagy különbségek).

3. A vizsgálati alanyok egymáshoz képest nagyobb különbségeket mutatnak.

Cronbach (1947) megfogalmazása szerint a reliabilitás azt fejezi ki, hogy mennyire alkalmas az itemek együttese az individuális különbségek jól értelmezhető megállapítására. Nyilvánvaló ez alapján, hogy egy olyan mérőeszköz használhatatlan lesz, amely nem tud különbségeket felmutatni a vizsgált személyek között. Tegyük hozzá, hogy ez nemcsak a teszten múlik, hanem a felmért mintán is: nagyon homogén mintákon, ahol az egyes személyek lényegében azonos pozíciót foglalnak el a mérendő tulajdonság tekintetében, lényegében nulla értékű Cronbach-alfát kapnánk – még az egyébként szakmailag alaposan megszerkesztett tesztünkre is.

Akkor, amikor konzisztens skálákat alkotunk, tulajdonképpen azt szeretnénk elérni, hogy az itemek legyenek stabilak: (1) legyenek a vizsgálati alanyaink válaszai hasonlóak a saját válaszaikhoz (ha egy adott kérdésre 1-es értéket adott, és hasonló dolgot kérdeztünk, akkor a hasonló kérdésre adott válasza 1-es vagy 2-es érték körül legyen, ne a skála másik végén – hacsak nem fordított a tétel); (2) a vizsgálati alanyaink legyenek változatosak, hogy a kérdőívünket kellően széles merítésen tudjuk alkalmazni – azaz a teljes kép legyen annyira színes, amennyire csak lehetséges. Összességében tehát azt mondhatjuk, hogy képletszinten a Cronbach-alfa mutató valóban azt csinálja, amit egy konzisztenciamutatótól elvárhatunk. Fontos leszögezni, hogy a Cronbach-alfa lényegében egydimenziós mutató, azaz egyetlen értéket ad vissza – ezért a kérdőívünk strukturális sajátosságait nem veszi figyelembe. Így, mint azt majd látni fogjuk a következő részben bemutatott szimulációk eredményeiből, a 0,7-es érték mint abszolút határérték, nem megfelelő. Valójában nem a formulával van a gond, hanem annak felhasználásával, amikor egy merev 0,7-es értéket használunk – függetlenül attól, hogy mennyi itemmel és milyen struktúrával állunk szemben.

Bár szerepel a formulában egy itemszámra vonatkozó korrekciós tényező, az itemek minőségét, azok viselkedését nem veszi figyelembe a mutató. Képzeljünk el egy extrém esetet, amikor is egy kérdőívben sokszor ismételve pontosan ugyanúgy szerepeltetünk egyetlen itemet. Ebben az esetben a válaszok konzisztenciájával nyilván nem lesz probléma – ugyanarra a kérdésre egy-egy személy vélhetően stabilan ugyanazt a

választ fogja adni (ha meg nem unja a kérdéseinket). De vajon ettől egy jól használható mérőeszközünk lett? Nyilván nem – és ez már nem a reliabilitás, inkább a tartalmi validitás kérdését veti fel (kellően körbelőttük-e a mérendő konstruktumunkat, azaz azt mérjük-e, amit valóban mérni szeretnénk), ami szükséges feltétele egy megfelelő minőségű teszt kialakításának.

MÓDSZER

Annak szemléltetésére, hogy ne egy általános, például 0,7-es értéket alkalmazzunk a döntéshozatalhoz, egy szimulációs eljárást készítettünk (a szimulációt mindenki maga is bármikor elvégezheti a saját adataihoz igazítva).

A szimulációt az alábbi formában valósítottuk meg.

1. Normális eloszlású változókat vettünk, melyek között rögzítettük az általános korrelációs mértéket. Ezt 0,1 és 0,9 között, tizedesugrásonként definiáltuk, tehát 0,1 és 0,9 között minden tizedes esetre készítettünk egy véletlen változószettet. A véletlen szimuláció nagyságát 10 ezres elemszámban határoztuk meg. Összefoglalva: vettünk egy többdimenziós normális eloszlású változószettet, melyben a változók között első esetben 0,1, második esetben 0,2 – és így tovább 0,9-ig – fixált korrelációs együtthatókat állítottunk be.

2. Ezek után a normális eloszlású változókat a szórás mértéke alapján kategóriákra osztottuk (1–7 osztópontot definiálva). Ezen azt értjük, hogy például az átlag alatt és felett vágva 2 értékű változókat kapunk újrakódolással (ahol az eredeti változóink közötti korreláció 0,1 és 0,9 között rögzített). 2 vágópont esetében az átlag és a szórás mértékében (1 szórás negatív, 1 szórás pozitív irányba) lesz 3 kategóriánk. 7 osztópont esetében 8 kategóriát definiáltunk. Vagyis az átlagtól távolodva, a szórás mértékét alapul véve 0,5; 1; 1,5; 2; 2,5 szórástávolságokat ugorva felosztottuk 11 sávra a normális eloszlású változóinkat.

3. Természetesen ilyenkor nem egyenletes az eloszlás a különböző csoportokban, de ez nem is volt cél, hiszen az eredeti struktúra is normális eloszlást feltételez. Egy másik megközelítés lehet az, ha diszkrét egyenletes eloszlás szerint generált véletlen változókkal dolgozunk – ebben az esetben viszont a Pearson-féle korrelációnak a jelentése is megváltozik, hiszen azt csak normális eloszlású változók között tudjuk érvényesen értelmezni.

Fontos kiemelni, hogy a kategóriák megalkotásakor nem volt cél az egyenletes eloszlás szerinti kategóriák létrehozása (bináris esetben az egyenletes eloszlás szerinti szétosztás történik szükségszerűen, de egyéb esetekben nem). Nem törekedtünk erre már csak azért sem, mert az itemek közötti elvárt kapcsolati erősséget a Pearson-féle korrelációs együttható alapján definiáltuk, ez azonban a normális eloszlásnak mint háttéreloszlásnak a választása esetében adekvát.

Ezzel az eljárással tehát szimulálni tudjuk azt, hogy adott Likert-skálák (illetve igen-nem típusú dichotóm kérdések) esetén – ahol elvileg az itemek között 0,1 és 0,9 közötti korrelációkat feltételezünk – adott itemszámok mellett milyen Cronbach-alfa értékek adódnak a szimulált, véletlen adatokon.

EREDMÉNYEK

Vegyük példaként az alábbi mátrixot (2. táblázat). Ebben látható, hogy ha 2 értékű változókkal dolgozunk, akkor 2 változó esetében, ha a közöttük lévő korreláció 0,8 vagy 0,9, a Cronbach-alfa érték eléri az általános küszöbként használt 0,7-es értéket. Ezzel szemben – és ez talán érdekesebb –, ha 2 értékű változót használunk, és a kérdőívünkben van 10 darab item, akkor 0,3-es közöttük lévő korreláció mellett is elérhető a 0,7-es Cronbach-alfa érték. Azaz: 0,3-es szinten korrelálnak csak egymással a kérdőívünk itemjei (alig-alig mérik ugyanazt a jelenséget), de 10 kérdést feltéve, egymással igen laza összefüggésben lévő kérdések esetében is elérhető magas Cronbach-alfa érték.

2. táblázat. 2 értékű (bináris, dichotóm) itemek használatakor, adott itemek közötti korrelációs szint melletti szimulált Cronbach-alfa értékek táblázata

Kódolt értékek száma	Változók száma	Itemek közötti korreláció								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
2	3	0,158	0,288	0,403	0,506	0,593	0,672	0,744	0,813	0,882
2	4	0,194	0,346	0,472	0,575	0,659	0,729	0,794	0,851	0,907
2	5	0,238	0,405	0,530	0,631	0,709	0,771	0,828	0,878	0,924
2	6	0,273	0,449	0,577	0,671	0,745	0,801	0,852	0,896	0,936
2	7	0,309	0,489	0,615	0,705	0,774	0,825	0,871	0,909	0,945
2	8	0,340	0,527	0,650	0,734	0,796	0,843	0,884	0,919	0,951
2	9	0,370	0,558	0,677	0,757	0,815	0,858	0,896	0,927	0,956
2	10	0,397	0,586	0,700	0,776	0,830	0,871	0,906	0,934	0,960
2	11	0,423	0,611	0,720	0,793	0,844	0,881	0,914	0,940	0,964

A mi javaslatunk tehát az, hogy ne a Cronbach-alfa értékből induljunk ki, hanem az itemek közötti átlagos korrelációs elvárásunk legyen rögzítve. Például egy kérdőívet akkor tekintünk az adott környezetben elfogadhatónak, ha az itemek közötti korrelációs szint legalább 0,5. Természetesen ez tudományterületenként újradefiniálható – például lehetünk ennél szigorúbbak (0,7 vagy 0,8 is elfogadható akár), de lehetünk kevésbé szigorúak is (0,4-es korrelációs szinttel).

A 2. táblázatban látható, hogy ha az adott kérdőívben az azt használni kívánók 7 itemet vonnak össze egy skálában, és 0,5-es szintet írnak elő, akkor igen-nem típusú kérdésekkel 0,77-os Cronbach-alfa értéket érdemes elvárni konzisztenciaként.

Az alábbi táblázatokban látható, hogy a vizsgált kódolt értékek (a Likert-skála értékeinek száma), illetve a változók száma (a kérdőív itemeinek száma) függvényében milyen Cronbach-alfa értékek adódtak (3–8. táblázatok).

3. táblázat. 3 értékű itemek használatakor, adott itemek közötti korrelációs szint mellett szimulált Cronbach-alfa értékek táblázata

Kódolt értékek	Változók száma	Itemek között elvárt korreláció								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	3	0,191	0,344	0,466	0,557	0,638	0,706	0,769	0,827	0,889
3	4	0,239	0,406	0,535	0,625	0,703	0,763	0,814	0,865	0,915
3	5	0,278	0,462	0,585	0,672	0,745	0,799	0,845	0,888	0,930
3	6	0,320	0,510	0,630	0,712	0,778	0,827	0,868	0,906	0,942
3	7	0,356	0,550	0,666	0,744	0,805	0,849	0,885	0,918	0,950
3	8	0,390	0,582	0,695	0,769	0,825	0,865	0,898	0,927	0,956
3	9	0,419	0,610	0,718	0,789	0,842	0,879	0,909	0,935	0,961
3	10	0,444	0,635	0,739	0,807	0,855	0,890	0,917	0,941	0,964
3	11	0,468	0,655	0,757	0,821	0,867	0,898	0,924	0,946	0,967

4. táblázat. 4 értékű itemek használatakor, adott itemek közötti korrelációs szint mellett szimulált Cronbach-alfa értékek táblázata

Kódolt értékek	Változók száma	Itemek között elvárt korreláció								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
4	3	0,211	0,371	0,501	0,604	0,689	0,761	0,824	0,877	0,925
4	4	0,257	0,433	0,569	0,668	0,747	0,807	0,861	0,905	0,942
4	5	0,303	0,491	0,622	0,714	0,787	0,840	0,885	0,923	0,953
4	6	0,345	0,538	0,665	0,750	0,815	0,862	0,902	0,934	0,961
4	7	0,383	0,578	0,699	0,779	0,838	0,880	0,915	0,943	0,966
4	8	0,418	0,613	0,728	0,802	0,855	0,893	0,925	0,950	0,970
4	9	0,448	0,641	0,750	0,820	0,869	0,904	0,933	0,955	0,973
4	10	0,474	0,666	0,770	0,835	0,881	0,913	0,939	0,959	0,976
4	11	0,500	0,687	0,786	0,848	0,891	0,920	0,944	0,963	0,978

5. táblázat. 5 értékű itemek használatakor, adott itemek közötti korrelációs szint melletti szimulált Cronbach-alfa értékek táblázata

Kódolt értékek	Változók száma	Itemek közötti elvárt korreláció								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
5	3	0,206	0,362	0,483	0,578	0,663	0,731	0,796	0,853	0,908
5	4	0,262	0,430	0,554	0,647	0,726	0,786	0,837	0,886	0,930
5	5	0,299	0,485	0,604	0,694	0,766	0,820	0,865	0,906	0,943
5	6	0,340	0,530	0,647	0,732	0,797	0,845	0,886	0,921	0,952
5	7	0,374	0,568	0,682	0,761	0,821	0,865	0,900	0,931	0,959
5	8	0,406	0,599	0,710	0,784	0,840	0,880	0,912	0,939	0,964
5	9	0,438	0,628	0,733	0,804	0,855	0,892	0,921	0,946	0,968
5	10	0,462	0,651	0,754	0,820	0,868	0,902	0,929	0,951	0,971
5	11	0,486	0,672	0,771	0,834	0,878	0,910	0,935	0,955	0,974

6. táblázat. 6 értékű itemek használatakor, adott itemek közötti korrelációs szint melletti szimulált Cronbach-alfa értékek táblázata

Kódolt értékek	Változók száma	Itemek közötti elvárt korreláció								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
6	3	0,222	0,386	0,518	0,622	0,706	0,775	0,835	0,886	0,930
6	4	0,273	0,453	0,587	0,685	0,762	0,820	0,870	0,911	0,946
6	5	0,321	0,511	0,639	0,730	0,800	0,850	0,893	0,928	0,957
6	6	0,361	0,556	0,681	0,764	0,827	0,872	0,909	0,939	0,964
6	7	0,399	0,595	0,714	0,791	0,849	0,888	0,921	0,947	0,969
6	8	0,433	0,629	0,741	0,813	0,865	0,901	0,930	0,953	0,973
6	9	0,464	0,657	0,763	0,830	0,878	0,911	0,937	0,958	0,975
6	10	0,491	0,681	0,782	0,845	0,889	0,919	0,944	0,962	0,978
6	11	0,516	0,702	0,798	0,857	0,898	0,926	0,948	0,966	0,980

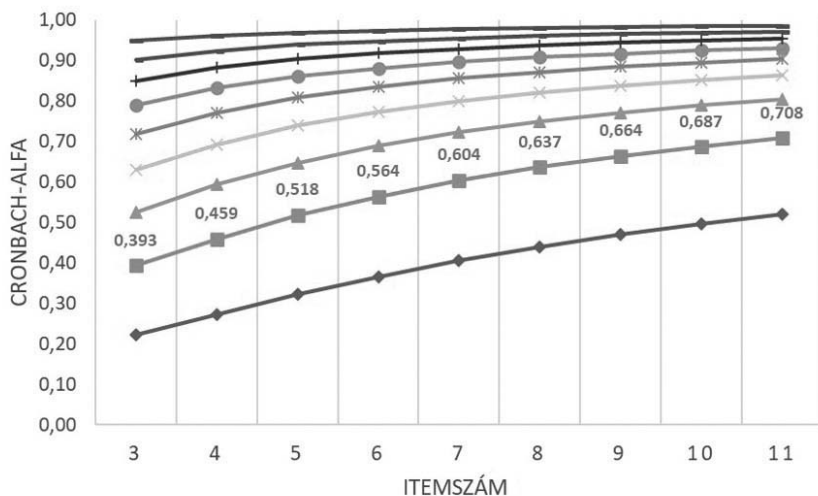
7. táblázat. 7 értékű itemek használatakor, adott itemek közötti korrelációs szint mellett szimulált Cronbach-alfa értékek táblázata

Kódolt értékek	Változók száma	Itemek közötti elvárt korreláció								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
7	3	0,235	0,411	0,542	0,643	0,727	0,796	0,852	0,901	0,945
7	4	0,291	0,480	0,611	0,706	0,780	0,839	0,885	0,924	0,958
7	5	0,339	0,536	0,661	0,750	0,815	0,866	0,905	0,938	0,966
7	6	0,383	0,583	0,701	0,783	0,841	0,885	0,919	0,948	0,972
7	7	0,421	0,620	0,734	0,808	0,861	0,900	0,930	0,955	0,976
7	8	0,454	0,650	0,759	0,827	0,876	0,911	0,938	0,960	0,979
7	9	0,484	0,677	0,780	0,844	0,888	0,920	0,945	0,965	0,981
7	10	0,510	0,699	0,797	0,857	0,898	0,928	0,950	0,968	0,983
7	11	0,534	0,719	0,812	0,868	0,907	0,934	0,954	0,971	0,984

8. táblázat. 8 értékű itemek használatakor, adott itemek közötti korrelációs szint mellett szimulált Cronbach-alfa értékek táblázata

Kódolt értékek	Változók száma	Itemek közötti elvárt korreláció								
		0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
8	3	0,223	0,393	0,526	0,631	0,717	0,789	0,850	0,902	0,949
8	4	0,272	0,459	0,595	0,693	0,771	0,832	0,883	0,924	0,961
8	5	0,322	0,518	0,647	0,739	0,808	0,861	0,903	0,939	0,968
8	6	0,365	0,564	0,689	0,773	0,835	0,881	0,918	0,948	0,973
8	7	0,405	0,604	0,722	0,799	0,856	0,896	0,929	0,955	0,977
8	8	0,440	0,637	0,749	0,820	0,871	0,908	0,937	0,961	0,980
8	9	0,469	0,664	0,771	0,837	0,884	0,917	0,944	0,965	0,982
8	10	0,496	0,687	0,789	0,851	0,894	0,925	0,949	0,968	0,984
8	11	0,521	0,708	0,804	0,863	0,903	0,931	0,953	0,971	0,985

Példaként a 2. ábráról mutatja a tendenciákat, ahol a 8 értékű válaszkálák esetében ábrázoltuk a szimulált adatokon kapott Cronbach-alfa értéket, vagyis vonalakkal összekötvük tüntettük fel az egyes korrelációs szintekhez tartozó változószámmal növekvő Cronbach-alfa értékeket.



2. ábra. Egy 8 értékű válaszkálát használó teszt szimulált adatain kapott Cronbach-alfa értékek (y-tengely) az itemszám (x-tengely), illetve az inter-item korrelációk 0,1–0,9 közötti tízedes ugrásonkénti szintjei (vonalak) alapján. (A konkrét számértékek példaként a 0,2-es korreláció esetén kapott alfákat mutatják.)

A fenti táblázatok alapján az alábbi következtetések vonhatók le:

1. Minél több értékű Likert-skálát alkalmazunk, várhatóan annál kevesebb itemszám már elegendő lesz az általános küszöbként használt 0,7-es Cronbach-alfa eléréséhez. Példaként mutatnánk, hogy az egymással 0,2-es szinten korreláló itemekből létrehozott, 11 itemet tartalmazó skála (ez még nem olyan nagyon sok kérdés, hogy alkalmazása a valóságtól elrugaszkodott feltételezés lenne) esetén elérhető az eddigi 0,7-es szint. Tegyük hozzá, hogy ennél alacsonyabb, 0,1 körüli korrelációkat már szakmai értelemben irrelevánsnak tekintünk, annyira csekély magyarázóerővel bírnak, hatásmértékük (effect-size) elenyésző.

2. A 0,95 feletti értékek igen ritkán fordulnak elő (egymással 0,8 vagy a feletti korrelációs értékek esetében jellemző), ami valóban redundáns információk megjelenését jelentheti a gyakorlatban, tehát vélelmezhető ilyen esetekben, hogy a kérdőív jelentősen rövidíthető, modernizálható.

A Cronbach-alfa alkalmazásakor tehát a következő algoritmus alapján végzett szimulációt javasoljuk:

1. Generáljunk véletlen normális eloszlású többdimenziós változószettet úgy, hogy a változók közötti korrelációt az általunk elvárt szinthez állítjuk be.

2. Ezek után kódoljuk újra a szórás függvényében annyi értékre a változóinkat, amennyi a kérdőívben a Likert-skálán szerepel.

3. A generátum itemjein számítsuk ki a Cronbach-alfa értéket.

Megjegyezzük, hogy a fenti algoritmus független attól, hogy fordított vagy azonos irányú itemek vannak, hiszen a fordított itemeket úgyszólván meg kell fordítani a Cronbach-alfa kiszámításához.

DISZKUSSZIÓ

Cho (2016) megfogalmazása szerint a Cronbach-alfa mutató a közfelfogásban az 1951-es szinten megállt, mint egy törött óra. Ugyanakkor egy széles körben használt mutatóról van szó, így fontos további kapaszkodókat keresni a szakmailag megalapozott alkalmazhatóságára. Tanulmányunkban ezért arra kerestük a választ, hogy skálák, indexek alkotásakor mennyiben lehet általános Cronbach-alfa küszöbértéket megadni, alkalmazni. (Sok elemzésben, publikációban 0,7-es küszöbérték jelenik meg.) Másként fogalmazva: vajon létezik-e olyan általánosan érvényes küszöbértéke a mutatónak, amely minden esetben kielégítő választ ad egy kérdőíves skála, index megbízhatóságával (és érvényességével) kapcsolatban, függetlenül attól, hogy hány kérdést tettünk fel, és milyen előzetes minőségi elvárásaink voltak a kérdőív kérdései közötti összefüggéseket illetően?

Elemzésünkben megvizsgáltuk a mutató formuláját, illetve kitértünk annak 1951-es intuitív bevezetésére is (Cronbach, 1951). Eredményeink alapján látható, hogy a mutatót a kérdőív itemeinek számától semmiképpen sem lehetne függetleníteni, továbbá a kérdőívben használt Likert-skála fokozataitól is érdemes függővé tenni a Cronbach-alfa elvárt értékét. Elemzésünk során azt is hangsúlyoztuk, hogy célszerű a Cronbach-alfa mutatót úgy értelmezni, amelynek ha nem ér el egy megfelelő szintet az értéke, akkor nagy az esélye annak, hogy szakmai problémák is felmerülnek a kérdőívünkkel kapcsolatosan. Ugyanakkor ez a szint eltérő lehet annak függvényében, hogy hány kérdéssel és milyen skálán mérünk egy konstruktumot. Szimulációnk eredményei jól szemléltetik, hogy ha adott vizsgálatban egy itemek közötti elvárt korrelációs szintet rögzítünk, akkor az adott kérdőív kódolása (hány válaszerőket adunk a Likert-skála kódolásakor) és az itemek számának függvényében megadható, hogy melyik az a Cronbach-alfa mutató, amelyet alsó értékként elvárhatunk egy adott szinten konzisztens skálától.

Fontos kiemelni – mint ahogy Cronbach és Shavelson (2004) is hangsúlyozták munkájukban –, hogy a Cronbach-alfa mutató csak az első, apróbb lépés lehet egy kérdőíves skála, index struktúrájának feltárása érdekében. Ehhez számos további kiegészítő vizsgálatra van szükség, azonban mindenképpen érdemes lehet azon elgondolkodni, hogy egy-egy tudományterületen milyen általános, átlagos szinteket tartunk elfogadhatónak az itemek közötti kapcsolat leírására. Innentől pedig ezen elváráshoz lehet mérni – figyelembe véve az itemek számát, illetve a Likert-skála felosztását is – az elvárt Cronbach-alfa mutató értékét (szemben az eddigi, 0,7-es általános gyakorlattal). Jelen elemzésünk alapján ellenőrizhető, hogy ha adott egy leendő mérőeszközünk, ahol adott fokszámmal Likert-skálákat alkalmazunk, akkor milyen minimális Cronbach-alfa értékeket tarthatunk elfogadhatónak.

IRODALOM

- Bentler, P. M. (2008). Alpha, Dimension-Free, and Model-Based Internal Consistency Reliability. *Psychometrika*, 74(1), 137. DOI: <https://doi.org/10.1007/s11336-008-9100-1>
- Brown, W. (1910). Some Experimental Results in the Correlation of Mental Abilities1. *British Journal of Psychology*, 1904–1920, 3(3), 296–322. DOI: <https://doi.org/10.1111/j.2044-8295.1910.tb00207.x>
- Cho, E. (2016). Making Reliability Reliable: A Systematic Approach to Reliability Coefficients. *Organizational Research Methods*, 19(4), 651–682. DOI: <https://doi.org/10.1177/1094428116656239>
- Cho, E., & Kim, S. (2015). Cronbach's Coefficient Alpha: Well Known but Poorly Understood. *Organizational Research Methods*, 18(2), 207–230. DOI: <https://doi.org/10.1177/1094428114555994>
- Churchill, G. A., & Peter, J. P. (1984). Research Design Effects on the Reliability of Rating Scales: A Meta-Analysis. *Journal of Marketing Research*, 21(4), 360–375. DOI: <https://doi.org/10.2307/3151463>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. DOI: <https://doi.org/10.1037/0021-9010.78.1.98>
- Cronbach, L. J. (1947). Test "reliability": Its meaning and determination. *Psychometrika*, 12(1), 1–16. DOI: <https://doi.org/10.1007/BF02289289>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. DOI: <https://doi.org/10.1007/BF02310555>
- Cronbach, L. J., & Shavelson, R. J. (2004). My Current Thoughts on Coefficient Alpha and Successor Procedures. *Educational and Psychological Measurement*, 64(3), 391–418. DOI: <https://doi.org/10.1177/0013164404266386>
- Duhachek, A., & Lacobucci, D. (2004). Alpha's standard error (ASE): An accurate and precise confidence interval estimate. *The Journal of Applied Psychology*, 89(5), 792–808. DOI: <https://doi.org/10.1037/0021-9010.89.5.792>
- Dunn, T. J., Baguley, T., & Brunnsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. DOI: <https://doi.org/10.1111/bjop.12046>
- Ekolu, S. O., & Quainoo, H. (2019). Reliability of assessments in engineering education using Cronbach's alpha, KR and split-half methods. *Global Journal of Engineering Education*, 21(1), 24–29.
- Fan, X., & Thompson, B. (2001). Confidence Intervals for Effect Sizes: Confidence Intervals about Score Reliability Coefficients, Please: An EPM Guidelines Editorial. *Educational and Psychological Measurement*, 61(4), 517–531. DOI: <http://dx.doi.org/10.1177/0013164401614001>
- George, D., & Mallery, P. (2019). *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference*. New York: Routledge.
- Glen, S. (2016). *Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy*. Statistics How To. Letöltve 2021. 04. 30-án: <https://www.statisticshowto.com/kaiser-meyer-olkin/>
- Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of Coefficient Alpha as an Index of Test Unidimensionality1. *Educational and Psychological Measurement*, 37(4), 827–838. DOI: <https://doi.org/10.1177/001316447703700403>
- Green, S. B., & Yang, Y. (2009). Commentary on Coefficient Alpha: A Cautionary Tale. *Psychometrika*, 74(1), 121–135. DOI: <https://doi.org/10.1007/s11336-008-9098-4>
- Gulliksen, H. (2013). *Theory of Mental Tests*. New York: Routledge.

- Hayes, A. F., & Coutts, J. J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But... *Communication Methods and Measures*, 14(1), 1–24. DOI: <https://doi.org/10.1080/19312458.2020.1718629>
- Horváth, G. (1997). *A modern tesztmodellek alkalmazása*. Illyés, S. (Ed.). Budapest: Akadémiai.
- Hunt, T. D., & Bentler, P. M. (2015). Quantile Lower Bounds to Reliability Based on Locally Optimal Splits. *Psychometrika*, 80(1), 182–195. DOI: <https://doi.org/10.1007/s11336-013-9393-6>
- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151–160. DOI: <https://doi.org/10.1007/BF02288391>
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34(1), 100–117. DOI: <https://doi.org/10.1111/j.2044-8317.1981.tb00621.x>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Hillsdale: Lawrence Erlbaum Associates Publishers.
- Nagybányai Nagy, O. (2006). A pszichológiai tesztek reliabilitása. In Rózsa, S., Nagybányai Nagy, O., & Oláh, A. (Eds), *A pszichológiai mérés alapjai: Elmélet, módszer és gyakorlati alkalmazás* (pp. 103–116). Bölcsész Konzorcium.
- Napitupulu, D., Kadar, J. A., & Jati, R. K. (2017). Validity Testing of Technology Acceptance Model Based on Factor Analysis Approach. *Indonesian Journal of Electrical Engineering and Computer Science*, 5(3), 697–704. DOI: <http://doi.org/10.11591/ijeecs.v5.i3.pp697-704>
- Neisser, U., Boodoo, G., Bouchard Jr., T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77–101. DOI: <https://doi.org/10.1037/0003-066X.51.2.77>
- Osburn, H. G. (2000). Coefficient alpha and related internal consistency reliability coefficients. *Psychological methods*, 5(3), 343–355. DOI: <https://doi.org/10.1037/1082-989X.5.3.343>
- Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*. *The European Health Psychologist*, 16(2), 56–69.
- Peterson, R. A. (1994). A Meta-Analysis of Cronbach's Coefficient Alpha. *Journal of Consumer Research*, 21(2), 381–391.
- Ponterotto, J. G., & Ruckdeschel, D. E. (2007). An Overview of Coefficient Alpha and a Reliability Matrix for Estimating Adequacy of Internal Consistency Coefficients with Psychological Research Measures. *Perceptual and Motor Skills*, 105(3), 997–1014. DOI: <https://doi.org/10.2466/pms.105.3.997-1014>
- Revelle, W., & Zinbarg, R. E. (2008). Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145. DOI: <https://doi.org/10.1007/s11336-008-9102-z>
- Rózsa, S., Hupuczi, E., Martin, L., Birkás, B., Hartung, I., Hargitai, R., et al. (2019). A Tellegen Abszorpciós Skála részletes pszichometriai elemzése. *Mentálhigiéné és Pszichoszomatika*, 20(1), 35–77. DOI: <https://doi.org/10.1556/0406.20.2019.003>
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350–353. DOI: <https://doi.org/10.1037/1040-3590.8.4.350>
- Sijtsma, K. (2008). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*, 74(1), 107. DOI: <https://doi.org/10.1007/s11336-008-9101-0>
- Sijtsma, K. (2015). Delimiting Coefficient α from Internal Consistency and Unidimensionality. *Educational Measurement: Issues and Practice*, 34(4), 10–13. DOI: <https://doi.org/10.1111/emip.12099>
- Spearman, C. (1910). Correlation Calculated from Faulty Data. *British Journal of Psychology*, 1904–1920, 3(3), 271–295. DOI: <https://doi.org/10.1111/j.2044-8295.1910.tb00206.x>

- Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in Psychology*, 7(769). DOI: <https://doi.org/10.3389/fpsyg.2016.00769>
- Van der Ark, L. A., Van der Palm, D. W., & Sijtsma, K. (2011). A Latent Class Approach to Estimating Test-Score Reliability. *Applied Psychological Measurement*, 35(5), 380–392. DOI: <https://doi.org/10.1177/0146621610392911>
- Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2006). 4 Reliability Coefficients and Generalizability Theory. In C. R. Rao & S. Sinharay (Eds), *Handbook of Statistics* (Vol. 26, pp. 81–124). Elsevier. DOI: [https://doi.org/10.1016/S0169-7161\(06\)26004-8](https://doi.org/10.1016/S0169-7161(06)26004-8)
- Yang, Y., & Green, S. B. (2011). Coefficient Alpha: A Reliability Coefficient for the 21st Century? *Journal of Psychoeducational Assessment*, 29(4), 377–392. DOI: <https://doi.org/10.1177/0734282911406668>
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α Revelle's β and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133. DOI: <https://doi.org/10.1007/s11336-003-0974-7>

CRONBACH'S ALFA: WITH OR WITHOUT IT?

T. KÁRÁSZ, JUDIT – NAGYBÁNYAI NAGY, OLIVÉR – SZÉLL, KRISZTIÁN –
TAKÁCS, SZABOLCS

Background and purpose: *In our study, we undertake to present a procedure that allows a more well-founded use of the Cronbach's alpha index, which has been widely used so far.*

Method: *In our analysis, we propose the use of a simple simulation procedure, which can be used in a given situation to provide a Cronbach's alpha reference point to achieve the expected overall consistency level (average correlation level).*

Results: *This is because our analysis points out that the Cronbach's alpha index should be made dependent not only on the number of items, but also on the number of Likert scale values used in the questionnaire. It is also the result of our analysis that the applied formula does not give the traditionally determined value of 0.7, but the expected Cronbach's alpha value dependent on both the Likert scale and the number of items, as well as the expected relationship strength between items.*

Conclusion: *With the help of the proposed algorithm, an application of the Cronbach's alpha index can be implemented, which is suitable to add a reference point to this reliability index based on our questionnaire (the answer Likert scale) and the number of items on the scale.*

Keywords: *Cronbach's alfa, simulation, threshold*

A cikk a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>) feltételei szerint publikált Open Access közlemény, melynek szellemében a cikk bármilyen médiumban szabadon felhasználható, megosztható és újraközölhető, feltéve, hogy az eredeti szerző és a közlés helye, illetve a CC License linkje és az esetlegesen végrehajtott módosítások feltüntetésre kerülnek. (SID_1)