

## Abstract

*Cronbach alpha* coefficient is still commonly used in research dedicated to the development of psychological tests. However, there is a certain lack of understanding of the significance of real and, especially, has its limits. The article presents the fundamental postulates of classical test theory and analyzes, in this context, the main issues affecting the calculation and interpretation of *Cronbach alpha* coefficient: unidimensionality, internal consistency, item number, characteristics of the data, the sampling error. Finally, are summarized recommendations of good practices on the use and reporting *Cronbach's alpha*.

**Keywords:** psychological tests, reliability, internal consistency

## Résumé

Le coefficient alpha de Cronbach est fréquemment utilisé dans la recherche dédiée au développement des tests psychologiques. Cependant, on constate un certain manque de compréhension de sa signification réelle mais aussi de ses limites. L'article présente les postulats fondamentaux de la théorie classique du test et fait l'analyse dans ce contexte des principaux aspects qui affectent le calcul et l'interprétation du coefficient alpha de Cronbach : l'unidimensionnalité, la cohérence interne, le nombre d'items, les caractéristiques des données, erreur d'échantillonnage. Enfin, on fait une synthèse des recommandations de bonnes pratiques concernant l'utilisation et le rapport de l'alpha de Cronbach.

**Mots clés :** tests psychologiques, fidelite, cohérence interne

## Rezumat

Coeficientul *Cronbach alfa* este încă frecvent utilizat în cercetarea dedicată dezvoltării testelor psihologice. Cu toate acestea, se constată o anumită lipsă de înțelegere a semnificației lui reale și, mai ales, a limitelor sale. Articolul prezintă postulatele fundamentale ale teoriei clasice a testului și analizează în acest context principalele aspecte care afectează calcularea și interpretarea coeficientului *Cronbach alfa*: unidimensionalitatea, consistența internă, numărul de itemi, caracteristicile datelor, eroarea de eșantionare. În final, sunt sintetizate o serie de recomandări de bune practici cu privire la utilizarea și raportarea lui *Cronbach alfa*.

**Cuvinte-cheie:** teste psihologice, fidelitate, consistență internă

Cunoașterea științifică în psihologie depinde în mod decisiv de dezvoltarea unor instrumente de măsură (teste, chestionare) destinate măsurării diverselor caracteristici psihice. În acest context, calcularea coeficientului *Cronbach alfa* încă reprezintă o practică larg răspândită, în ciuda unor dezvoltări moderne, cum ar fi teoria răspunsului la itemi sau modelarea ecuației de

structură (ten Holt, van Duijn, & Boomsma, 2010). Deși *Cronbach alfa* se bucură de o mare popularitate, persistă totuși numeroase confuzii cu privire la semnificația sa reală și interpretarea lui adecvată.

De cele mai multe ori, calcularea coeficientului *Cronbach alfa* este abordată ca un ritual statistic mecanic, ce decurge dintr-o obligație

---

<sup>1</sup> Universitatea din București.

Adresa de corespondență: popamarian@gmail.com.

formală, care se consideră încheiată odată ce a fost atins pragul „critic“ de 0,70, general acceptat. În cele ce urmează, ne propunem să aducem în discuție câteva aspecte care sunt destul de frecvent ignorate, precum și modalitățile prin care putem ameliora interpretarea analizei de fidelitate, cu referire specială la coeficientul *Cronbach alfa*.

## 1. Teoria clasică a testului și fidelitatea măsurării

Psihologia poate aspira la atributul de știință numai în limitele capacității de a reflecta în mod adecvat realitatea psihică. Dacă este adevărat faptul că știința începe cu observarea/măsurarea realității, atunci și problemele științei încep tot din acest punct. Și aceasta deoarece „*orice măsurare este afectată de eroare*“ (McNemar, 1946). *Teoria clasică a testului* (TCT), cunoscută și sub denumirea de „teoria scorului adevărat“, a fost prefigurată de Spearman (1904) și formulată sistematic de Gulliksen (1950). Presupunerea fundamentală pe care se întemeiază TCT este aceea că variația răspunsurilor la un test ar trebui să reflecte exclusiv variația abilităților/trăsăturilor vizate de testul respectiv. În acest context, celelalte surse de variație a scorurilor reprezintă *erori*, care pot fi de două feluri: *sistematice* sau *aleatorii*. Erorile sistematice au un efect constant asupra tuturor scorurilor obținute prin măsurare, ca rezultat al standardizării testului și al procedurii de aplicare. Din acest motiv, ele mai sunt denumite și „erori de metodă“. În ce privește erorile aleatorii, acestea au un caracter variabil de la un scor la altul. Ele afectează fiecare va-

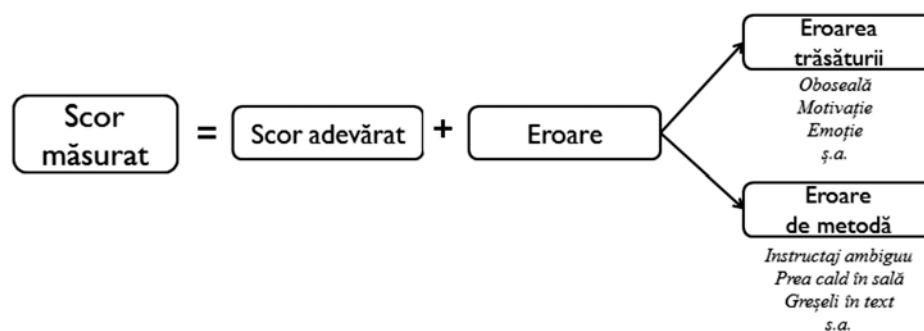
loare obținută prin măsurare, fie crescător, fie descrescător, dar natura lor aleatorie face ca erorile pozitive (supraestimarea valorii adevărate) să fie neutralizate de erorile negative (subestimarea valorii adevărate). Ca urmare, pe o distribuție în care fiecare scor este afectat de o eroare aleatoare, media distribuției nu este afectată. Erorile aleatorii provin din interacțiunea caracteristicii supuse măsurării cu alte caracteristici individuale, motiv pentru care mai sunt denumite și „erori ale trăsăturii“. O prezentare detaliată a surselor de eroare în măsurarea psihologică nu intră în obiectivul nostru aici, dar recomandăm celor interesați analiza sistematică a erorilor de testare realizată de Pitariu și Albu (1996).

Modelul măsurării promovat de TCT se interesează cu precădere de erorile aleatorii, problema erorilor sistematice rămânând în mare parte nerezolvată (Kline, 2005). Din această perspectivă, orice valoare obținută prin măsurare ( $X$ ) este compusă dintr-o parte „adevărată“ ( $T$ ), care este expresia caracteristicii pe care dorim să o măsurăm, la care se adaugă eroarea aleatorie ( $E$ ), care indică abaterea a valorii măsurate de la valoarea reală:

$$X = T + E \quad (1)$$

Principala provocare pentru măsurarea în psihologie constă în limitarea componentei aleatorii a măsurării. Cu cât aceasta este mai mică, cu atât valoarea măsurată ( $X$ ) descrie mai fidel valoarea adevărată ( $T$ ). Dacă am aplica aceeași persoană un test de inteligență de  $n$  ori, ar trebui să ne așteptăm să obținem  $n$  scoruri diferite, uneori mai mari, alteori mai mici decât valoarea reală a inteligenței, în funcție de dispoziția de moment, de starea de odihnă, de impactul unor

Figura 1. Teoria clasică a testului



evenimente personale etc. Într-o asemenea situație, media scorurilor la testul de inteligență ar reprezenta cea mai bună estimare a nivelului adevărat de inteligență al persoanei evaluate. În acest context, în conformitate cu TCT, având în vedere natura aleatorie a erorilor de măsurare de-a lungul evaluărilor repetate, ne așteptăm să fie întrunite trei condiții fundamentale:

- media erorilor de măsurare tinde spre 0, dacă numărul măsurărilor crește la infinit;
- distribuția erorilor are o formă normală, erorile mici alcătuind partea centrală a distribuției;
- erorile sunt independente una de alta, altfel spus, nu există nici un model sistematic de fluctuație a erorilor, de la o măsurare la alta;
- erorile sunt independente de valoarea adevărată, ceea ce înseamnă că variația erorii nu are nici o legătură cu variația scorului adevărat.

Aplicarea repetată a unui test aceleiași persoane este, evident, un lucru nepractic. Acest exercițiu de imaginație însă, ne permite transferul condițiilor fundamentale ale TCT, de la eșantionul de măsurări pe o singură persoană, la distribuția măsurărilor obținute prin aplicarea testului pe un eșantion de persoane diferite. În această nouă situație, modelul TCT poate fi formalizat cu expresia:

$$VAR(X)=VAR(T)+VAR(E) \quad (2)$$

unde  $VAR$  se referă la varianța fiecăruia dintre termenii ecuației la nivelul eșantionului de valori.

În acest context se devine operant conceptul de *fidelitate a măsurării*, care exprimă, în termeni generali, cât de consistentă este o măsurare aplicată în mod repetat. Desigur, nu putem ști niciodată care este valoarea adevărată a caracteristicii supuse măsurării, dar putem presupune că este „stabilă“ de la un moment la altul. Ca urmare, considerăm fidelă o măsurare care produce valori identice (sau măcar apropiate) în momente diferite. În același timp, o măsurare este fidelă dacă, efectuată pe mai mulți subiecți, este capabilă să reproducă corect diferențele reale dintre aceștia. Cu alte cuvinte, fidelitatea se referă la precizia măsurării, care este cu atât mai mare, cu cât eroarea de măsurare este mai mică.

Una dintre formalizările posibile ale fidelității ( $R$ ) poate fi exprimată ca raport între variabilitatea scorului măsurat și variabilitatea scorului adevărat:

$$R = \frac{VAR(T)}{VAR(X)} \quad (3)$$

Din expresia (3) rezultă că fidelitatea măsurării poate lua valoarea minimă 0, atunci când variabilitatea scorului adevărat este 0. În conformitate cu expresia (2), aceasta ar însemna că toată variabilitatea scorului măsurat este datorată variabilității erorii. Fidelitatea maximă posibilă este 1 și este atinsă atunci când variabilitatea scorului adevărat este identică cu variabilitatea scorului măsurat. În conformitate cu expresia (2), acest lucru ar însemna măsurare fără eroare. Cum în practică nu există măsurare liberă de eroare, rezultă că fidelitatea oricărei măsurări va fi întotdeauna mai mică de 1. Drept urmare, încercând să cuantificăm fidelitatea nu facem altceva decât să cuantificăm eroarea de măsurare. Altfel spus, dacă măsurăm cu o fidelitate  $R=0,70$ , aceasta înseamnă că doar proporție de 70% din variația scorurilor măsurate ( $X$ ) poate fi pusă pe seama variației valorilor adevărate ( $T$ ), restul de 30% ( $1-0,70$ ) exprimând variație determinată de eroare ( $E$ ). Pe măsură ce nivelul fidelității de apropiere de 0, valorile obținute prin măsurare variază aleatoriu, fără nici o legătură cu ceva anume, ceea ce este echivalent cu a spune că nu măsoară, de fapt, nimic.

Pe parcursul discuției de mai sus am avut în vedere doar scorul global obținut la un test, dar testele psihologice sunt, de regulă, compuse din itemi (întrebări), fiecare dintre ei fiind expresia mai mult sau mai puțin exactă a constructului vizat de testul respectiv. Fidelitatea poate fi discutată deci, și din perspectiva modului în care itemii unui test descriu constructul psihologic supus măsurării.

În contextul TCT, fidelitatea nu poate fi calculată în mod direct, deoarece scorul adevărat ( $T$ ) este inaccesibil măsurării rămâne necunoscut, la fel ca și eroarea de măsurare ( $E$ ). În acest scop, abordarea fidelității este inclusă în contextul mai larg al *analizei de itemi*. În linii mari, aceasta este compusă din două mari categorii de proceduri: unele care vizează itemii luați separat (indicele de dificultate, indicele de discriminare,

corelația item-scor total ș.a.) și altele care vizează toți itemii la un loc. Procedurile din a doua categorie se referă la: *fidelitatea test-retest* (gradul de concordanță între două măsurări repetate, cu același instrument, al acelorași persoane); *fidelitatea formelor paralele* (gradul de concordanță dintre măsurările acelorași persoane cu două teste diferite, considerate echivalente), *fidelitatea split-half* (gradul de concordanță între scorurile parțiale, obținute prin împărțirea în două a itemilor testului) și, ultima pe listă, dar nu cea din urmă, *consistența internă* (gradul de concordanță între răspunsurile la itemii testului). Obiectul discuției noastre în continuare îl va face doar ultima dintre acestea și anume, analiza de consistență internă.

## 2. Coeficientul *Cronbach alfa*

Coeficientul *alfa*, denumit în mod curent *Cronbach alfa* ( $\alpha$ ), a fost propus de reputatul psihometrician american Lee J. Cronbach (1951), ca o generalizare a coeficientului Kuder-Richardson (*KR20*), pentru itemi cu răspuns dichotomic. Deși, așa cum vom vedea mai departe, nu este singurul indicator al fidelității, *Cronbach alfa* este, de departe, cel mai utilizat și cel mai frecvent raportat în literatura de specialitate (Christmann & Van Aelst, 2006; Ercan, Yazici, Sigirli, Ediz, & Kan, 2007; ten Holt, et al., 2010). Explicația acestei situații rezidă, pe de o parte, într-o aparentă accesibilitate teoretică și, pe de altă parte, în faptul că poate fi ușor calculat cu ajutorul programelor statistice uzuale, cum este, de exemplu, *SPSS* (Borsboom, 2006).

Formula (4) este una dintre expresiile de calcul pentru *Cronbach alfa*.

$$\alpha = \frac{N * r_m}{1 + (N - 1) * r_m} \quad (4)$$

unde:

$N$ =numărul itemilor

$r_m$ =media coeficienților de corelație dintre itemi

Teoretic,  $\alpha$  poate lua valori între 0 și 1, dar nu este exclus, în cazuri extreme, să obținem

uneori și valori mai mari de 1 sau negative. *Cronbach alfa* poate lua valoare negativă atunci când suma corelațiilor negative dintre itemi este mai mare decât suma corelațiilor pozitive (Field, 2009). Se poate ajunge într-o astfel de situație atunci când datele conțin erori, eșanționul este foarte mic sau numărul itemilor este foarte redus.

În mod normal, *Cronbach alfa* se calculează pe baza covarianței dintre itemi. Se poate obține și o valoare standardizată, pe baza corelației inter-itemi. Prima formă este recomandată atunci când scorul global este obținut prin însumarea valorilor la fiecare item. Se va apela la varianta standardizată atunci când itemii sunt măsurăți cu unități de măsură diferite, fiind necesară standardizarea lor prealabilă (programul *SPSS* calculează implicit ambele forme).

Nu există un standard absolut cu privire la mărimea pe care ar trebui să o aibă un coeficient *Cronbach alfa* pentru a indica o fidelitate corespunzătoare. În general însă, valorile în jur de 0,90 sunt considerate „excelente“, în jur de 0,80, „foarte bune“, iar cele în jur de 0,70, „adecvate“ (Kline, 2005). Există însă și autori care acceptă și o valoare de 0,60, dar numai în studii cu caracter exploratoriu (Garson, 2010)\*.

## 3. Impactul fidelității asupra validității

Importanța mărimii coeficientului *Cronbach alfa* poate fi înțeleasă mai bine dacă o plasăm pe terenul relației dintre validitate și fidelitate, deoarece „*Chiar și acei cercetători care consideră fidelitatea drept o umbră palidă a mult mai importante probleme a validității, nu pot ignora fidelitatea măsurărilor lor. Nici un coeficient de validitate și nici o analiză factorială nu pot fi interpretate în afara unei estimări corespunzătoare a mărimii erorii de măsurare*“ (Cronbach, 1951).

O măsurare poate fi fidelă, fără a fi în mod necesar validă. *Validitatea* se referă la corectitudinea măsurării, altfel spus, la capacitatea acesteia de a reflecta în mod adecvat ceea ce dorim să măsurăm și nu altceva. Intuitiv, ne putem re-

\* Pentru precizia argumentației statistice, pe parcursul acestui articol vom prezenta uneori valorile lui *alfa* cu trei zecimale, dar în practică raportarea cu doar două zecimale este suficientă.

**Tabelul 1.** Fidelitatea și validitatea măsurărilor în raport cu valoarea adevărată ( $T$ )

	$T=40$					
fidelitate și validitate	40	40	40	40	40	40
fidelitate fără validitate	30	30	30	30	30	30
validitate fără fidelitate	39	41	39	40	42	38

prezenta diferența dintre fidelitate și validitate cu ajutorul valorilor din tabelul 1, care cuprinde trei serii de măsurări repetate, cu același instrument, ale unei caracteristici a cărei valoare adevărată o presupunem cunoscută ( $T=40$ ).

Dacă înțelegem validitatea ca fiind intensitatea corelației dintre un predictor și criteriul său, atunci trebuie să ne așteptăm ca nivelul acesteia să fie influențat de orice aspect care afectează îndeobște corelația. Coeficienții de corelație pot fi mai mici decât în realitate din multe motive, iar unul dintre cele mai importante motive este eroarea de măsurare.

Când valorile predictorului și ale criteriului variază „haotic“ (aleatoriu), atunci o anumită cantitate de variație a unuia nu va putea avea nici o relație cu variația celui de-al doilea. Având în vedere că nu există măsurare fără eroare, înseamnă că orice coeficient de validitate este mai mic decât ar putea fi în realitate. În conformitate cu teoria clasică a fidelității, limita maximă a unui coeficient de corelație nu poate fi mai mare decât radicalul coeficientului de fidelitate al oricăreia dintre cele două variabile. Acest fenomen, denumit *atenuarea corelației*, a fost observat de Spearman (1904), care a propus o formulă de corecție a atenuării, pe care o reproducem mai jos, cu notațiile originale:

$$r_{pq} = \frac{r_{p'q'}}{\sqrt{r_{p'p'} * r_{q'q'}}} \quad (5)$$

unde

$r_{pq}$  = coeficientul de corelație corectat dintre  $p$  și  $q$

$r_{p'q'}$  = coeficientul de corelație necorectat dintre  $p$  și  $q$

$r_{p'p'}$  = coeficientul de fidelitate al variabilei  $p$

$r_{q'q'}$  = coeficientul de fidelitate al variabilei  $q$

Rezultatul corecției reprezintă o estimare a corelației adevărate dintre cele două variabile, dacă fidelitatea măsurării ar fi perfectă. Să luăm

ca exemplu constatarea că variabilele psihologice, mai ales cele de personalitate, au coeficienți de validitate care depășesc rareori valoarea 0,50 (Crețu, 2005) și să presupunem un nivel „acceptabil“ de fidelitate atât pentru predictor cât și pentru criteriu (0,70). În aceste condiții, valoarea adevărată a validității predictorilor de personalitate ar putea fi estimată astfel:

$$r_{pq} = \frac{0,50}{\sqrt{0,70 * 0,70}} = 0,71$$

În exemplul de mai sus am presupus cunoscut nivelul de fidelitate al ambelor variabile. Există și posibilitatea corecției parțiale de atenuare, în care numitorul conține doar coeficientul de fidelitate al uneia dintre cele două variabile. Mai mult, corecția poate fi efectuată chiar și atunci când nu cunoaștem coeficienții de fidelitate, prin asumarea unei anumite valori implicite pentru aceștia. O soluție în acest caz este asumarea unui nivel „acceptabil“ de fidelitate dar, pentru a evita o corecție „excesivă“, se recomandă varianta mai conservatoare a asumării unui nivel mare pentru fidelitatea implicită (de ex., 0,90), care produce o corecție mai mică a corelației (Kline, 2005).

Pentru a ne face o imagine cu privire la impactul deficitului de fidelitate asupra nivelului validității, este suficient să consemnăm rezultatele analizei efectuate de Carreta și Ree (2001) în legătură cu validitatea bateriei de teste AFQT (*Air Force Qualification Test*). Dacă înainte de corecție, coeficienții de validitate ai celor 16 teste variau între 0,026 și 0,13, după corecție, nivelul validității s-a plasat pe o plajă cuprinsă între 0,24-0,57. Deficitul de fidelitate nu afectează doar nivelul validității, ci și precizia predicției în ecuațiile de regresie ori estimarea încărcării în analiza factorială.

În general însă, așa cum constată Kanyongo et al. (2007) un nivel mai ridicat al erorii de

măsurare se transpune într-o diminuare a puterii statistice pentru majoritatea testelor parametrice și neparametrice. La rândul lor, Ree și Carreta (2006), au adus dovezi în sprijinul ideii că fidelitatea scăzută afectează multe dintre procedurile statistice cele mai frecvent folosite: corelația, corelația parțială, analiza de varianță, analiza factorială ș.a. Este de remarcat și faptul că, în contextul utilizării modelării ecuațiilor structurale, includerea în model a erorii de măsurare oferă posibilitatea unei estimări mai adecvate a corelației adevărate dintre variabile (Bedeian, Day, & Kelloway, 1997; Kline, 2005)

Corecția de atenuare reprezintă o modalitate practică prin care eroarea de măsurare este luată în considerare în estimarea validității. Totuși în ciuda utilității evidente, această procedură ridică o serie de probleme și controverse (Muchinsky, 1996; Schmitt, 1996; Zimmerman & Williams, 1997):

- corecția de atenuare nu poate reprezenta o alternativă la lipsa de fidelitate a măsurării;
- coeficienții de corelație corecți nu mai prezintă aceeași distribuție de eșantionare ca și varianta necorectată, iar ca urmare, ei nu pot utilizați pentru testarea ipotezelor;
- se recomandă calcularea limitelor de încredere pentru  $r$  corectat pe baza limitelor de încredere pentru  $r$  necorectat;
- combinația unor coeficienți reduși de fidelitate, cu o variabilitate ridicată și un eșantion mic, poate conduce la o supracorecție, cu valori chiar mai mari decât 1, situație în care valoarea corectată va fi limitată artificial la 1;
- diferența dintre  $r$  corectat și  $r$  necorectat poate fi utilizată ca informație utilă pentru aprecierea contribuției erorii de măsurare la nivelul obținut al corelației;
- în general, corecția de atenuare pare a fi utilă atunci când coeficienții de fidelitate ai am-

belor variabile sunt relativ ridicați, iar eșantionul este mare;

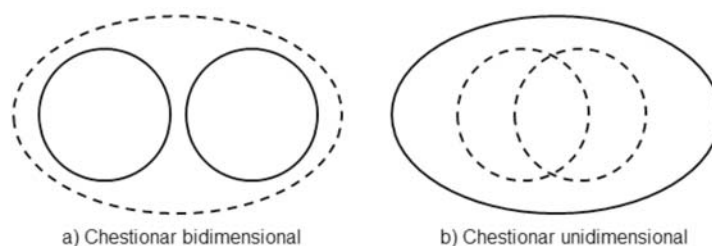
- corecția de atenuare produce supraestimarea corelației adevărate, atunci când testele supuse corelației nu sunt unidimensionale, ci multidimensionale.

#### 4. *Cronbach alfa*, indicator imprecis al unidimensionalității

Coeficientul *Cronbach alfa* este prezentat cel mai adesea ca fiind un indicator al existenței unui construct unic, care înglobează cea mai mare parte a variabilității răspunsurilor la itemii analizați. În limbajul modelului factorial, variabilitatea scorului măsurat este compusă din trei componente: varianța datorată factorului comun; varianța factorilor specifici (unicitatea itemilor) și varianța reziduală (eroarea). Din această perspectivă, *Cronbach alfa* ne spune cât de mare este comunalitatea itemilor unui test, cât de redusă este „unicitatea“ lor și, de asemenea, cât de mare este legătura dintre itemi. Aceste caracteristici însă nu implică faptul că itemii ar descrie un singur factor (Cortina, 1993; Schmitt, 1996; Tan, 2009). De altfel, Cronbach însuși precizează că *alfa* estimează cu precădere „concentrația primului factor“ (Cronbach, 1951, p. 321). Din acest motiv, pe măsură ce numărul itemilor încărcăți cu primul factor crește, valoarea coeficientului *Cronbach alfa* crește și ea. Altfel spus, *alfa* reprezintă o măsură a saturației primului factor (Cortina, 1993).

Aici trebuie remarcată confuzia frecventă care se face între omogenitate și consistența internă. În timp ce omogenitatea se referă la unidimensionalitatea itemilor, consistența internă se referă la intensitatea corelației dintre aceștia (Green, Lissitz, & Mulaik, 1977). Consistența internă este o condiție necesară a omogenității,

Figura 2. Unidimensionalitatea și indicele *Cronbach alfa*



dar nu și suficientă. Astfel, o valoare mică a lui *alfa* s-ar putea datora faptului că itemii descriu nu unul, ci, să zicem, două constructe divergente (fig. 2a). Dacă însă am calcula coeficientii fidelitate pentru itemii care compun fiecare dintre cele două constructe, am putea obține două valori ridicate ale lui *alfa*. Pe de altă parte, putem obține un coeficient *alfa* acceptabil atunci când itemii chestionarului ar descrie nu unul, ci doi sau mai mulți factori, care nu corelează între ei, sau chiar dacă sunt corelați (de exemplu, două componente ale agresivității, „verbală“ și „fizică“directă“) (fig. 2b).

Calcularea coeficientului *Cronbach alfa* se bazează pe presupunerea că există un factor unic care explică variabilitatea itemilor, iar valoarea lui este afectată atunci când această presupunere nu se susține. Totuși, o valoare acceptabilă a lui *alfa* nu reprezintă o dovadă a unidimensionalității, ci doar indicatorul unei dimensiuni unice despre care nu putem ști dacă este cel mai adecvat model de a descrie datele respective. Dacă dorim o dovadă a unidimensionalității, atunci aceasta trebuie căutată cu alte mijloace (de ex., analiza factorială de confirmare).

Calcularea lui *Cronbach alfa* nu are sens atunci când presupunerea unidimensionalității nu există. Acesta este cazul, de exemplu, chestionarelor cu întrebări factuale, care nu sunt interpretate prin adăugarea răspunsurilor, lucru care este îndeobște cunoscut. Mai puțin cunoscut este faptul că *alfa* nu are sens nici în cazul chestionarelor de cunoștințe. Constructele vizate de chestionarele psihologice sunt presupuse a avea o distribuție naturală, aleatorie, pe când distribuția performanței la testele de cunoștințe este influențată de procesul de învățare. Dacă, de exemplu, aplicăm un test de cunoștințe înainte de parcurgerea unei materii de învățământ, vom obține un procentaj foarte scăzut de răspunsuri corecte și o corelație item-total aproape de zero, fapt care reflectă necunoașterea de către elevi a materiei respective (răspunsuri greșite, răspunsuri la întâmplare etc.). În acest caz soluția nu constă în eliminarea sau refacerea itemilor, ci în instruirea elevilor. Dacă aplicăm testul respectiv după parcurgerea perioadei de instruire, vom obține procentaje mari de răspunsuri corecte, ceea ce ar fi interpretabil drept o redusă capacitate discriminativă a itemilor. În realitate,

eliminarea acestor itemi ar fi de asemenea o eroare, deoarece am elimina tocmai itemii care pun în evidență succesul învățării, ceea ce este exact obiectivul măsurării. Din aceste motive, în astfel de situații utilizarea indicelui de consistență internă *Cronbach alfa* este inadecvată. Cu toate acestea, indicii de dificultate și de discriminare pot fi utilizați ca modalități de identificare a itemilor greșit formulați, a celor nerelevanți sau, pur și simplu, pentru a scoate în evidență erorile tipice pe care le fac subiecții la anumite întrebări.

## 5. *Cronbach alfa*, indicator nesigur al consistenței interne

Având un test format din mai mulți itemi, coeficientul *Cronbach alfa* cuantifică proporția varianței comune tuturor itemilor care se regăsește în scorul total (Cronbach, 1951). Presupunerea fundamentală este aceea că fiecare item reprezintă, în felul lui, un retest al celorlalți itemi. Altfel spus, de exemplu, dacă avem răspunsurile la 10 itemi ai unui chestionar, îi considerăm ca și cum am avea 10 răspunsuri la un item unic. Dacă această presupunere este corectă, atunci trebuie să ne așteptăm să avem o corelație medie ridicată între răspunsuri. Bazându-se pe corelația medie dintre itemi, *Cronbach alfa* tinde să fie cu atât mai mare, cu cât corelația medie dintre itemi este mai mare sau cu cât numărul corelațiilor egale dintre itemi este mai mare. Pe măsură ce itemii corelează mai puternic este nevoie de mai puțini subiecți pentru a se atinge niveluri ridicate ale coeficientului *Cronbach alfa*. Itemii care dau corelații cu alți itemi de 0,4 sau mai mari, contribuie într-o măsură mai importantă la creșterea valorii lui *alfa*.

În conformitate cu caracteristicile de mai sus, *Cronbach alfa* este un indicator al consistenței interne. Numai că, dacă ar fi așa, valoarea lui *alfa* nu ar trebui să fie afectată de numărul itemilor, ceea ce în realitate se întâmplă. Dacă o scală are mai mult de 20 de itemi, poate avea un coeficient *alfa* de 0,70, chiar atunci când corelația dintre itemi este foarte mică (Cortina, 1993). Chiar și pentru o corelație medie inter-itemi de 0,1, valoarea lui *alfa* poate să crească de la 0,37

**Tabelul 2.** Valoarea indicelui *Cronbach Alfa* în funcție de numărul de itemi și de corelația medie dintre aceștia (Zeller, 2001)

<i>r</i> mediu	Număr de itemi				
	5	10	20	30	50
0.1	.37	.53	.69	.77	.850
0.2	.56	.71	.83	.88	.930
0.3	.68	.81	.90	.93	.960
0.4	.77	.87	.93	.95	.970
0.5	.83	.91	.95	.97	.980
0.6	.88	.94	.97	.98	.990
0.7	.92	.96	.98	.99	.990
0.8	.95	.98	.987	.992	.995
0.9	.98	.99	.994	.996	.998

(nesatisfăcător), pentru o scală cu cinci itemi, la 0,96 (ridicat), pe o scală cu 50 de itemi (tab. 2).

Din moment de *Cronbach alfa* crește odată cu numărul itemilor, am putea fi tentați să utilizăm chestionare cu un număr mai mare de întrebări. Acest lucru însă, poate genera cel puțin două probleme pe care trebuie să le luăm în considerare:

- din motive practice, testele mai scurte sunt de dorit, dată fiind reducerea costurilor de realizare și a duratei de aplicare;

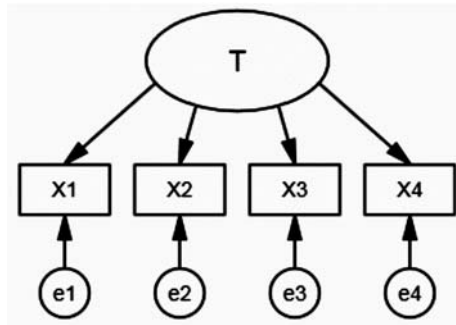
- în cazul testelor cu număr mare de itemi pot exista întrebări care, chiar dacă au o corelație pozitivă cu scala, contribuie la reducerea coeficientului *alfa*, sau au o contribuție minoră la consistența generală a scalei.

Valoarea lui *Cronbach alfa* nu este influențată doar de numărul itemilor, ci și de volumul eșantionului. Cu cât acesta crește, *alfa* tinde să crească la rândul lui. Astfel, dacă ne propunem un anumit nivel de fidelitate, îl putem atinge în mod artificial, prin creșterea volumului eșantionului. Desigur, cu cât corelația dintre itemii testului este mai redusă, cu atât volumul eșantionului necesar trebuie să fie mai mare. De exemplu, în cazul unei corelații medii interitemi de 0,1, putem atinge o valoare a lui *alfa* de 0,7, cu 21 de subiecți, pe care o putem ridica la 0,9, dacă utilizăm 81 de subiecți (Tabelul 3). În general, a fost dovedit faptul că acuratețea estimării fidelității crește odată cu creșterea volumului eșantionului, care nu ar trebui să fie mai mic de 50 de subiecți (Ercan, et al., 2007).

**Tabelul 3.** Volumul eșantionului în funcție de diferite valori *Cronbach Alfa* și media coeficienților de corelație *r* (Zeller, 2001)

<i>r</i> mediu	<i>Cronbach alfa</i>		
	0.7	0.8	0.9
0.1	21	36	81
0.2	10	16	36
0.3	6	10	21
0.4	4	6	14
0.5	3	4	9
0.6	2	3	6
0.7	1	2	4
0.8	1	1	3

Figura 3. Model SEM de principiu, pentru fidelitatea unui chestionar unidimensional cu 4 itemi



### 6. Cronbach alfa, indicator limitat al fidelității

Fidelitatea este o realitate complexă, pe care Cronbach alfa nu o descrie nici corect și nici complet. Din perspectiva corectitudinii, coeficientul Cronbach alfa reprezintă doar o estimare a gradului de acuratețe cu care itemii chestionarului ( $X_1 \dots X_n$ ), descriu împreună constructul latent ( $T$ ). Nivelul de acuratețe al estimării nu poate fi cunoscut, dar toți autorii sunt de acord că alfa subestimează nivelul real al fidelității fiind, mai exact, limita ei inferioară (Cronbach, 1951). Acest lucru se datorează încălcării condițiilor asumate prin modelul de măsurare  $\tau$  (*tau*)-echivalent esențial (Graham, 2006; Sijtsma, 2009). Pentru a înțelege mai ușor acest lucru să ne imaginăm un chestionar care vizează un construct unic, format din patru itemi. Fiecare item are o anumită încărcătură cu constructul vizat, precum și o anumită valoare reziduală (unicitate) în raport cu acesta. Expresia grafică a acestei situații, în spiritul modelului ecuațiilor structurale, este cea ilustrată de figura 3.

Modalitatea de estimare a fidelității cu care variabilele observate  $X_1 \dots X_4$  compun valoarea adevărată ( $T$ ), depinde condițiile ni le asumăm cu privire la cei patru indicatori și la erorile asociate lor. În funcție de configurația specifică a acestor condiții, ne asumăm unul din variatele modele de măsurare posibile (Raykov, 1997). Cel mai restrictiv model, *modelul paralel*, se bazează pe condiția fundamentală a unidimensionalității (toți itemii testului măsoară același construct). La această condiție de bază se adaugă echivalența strictă a itemilor, ceea ce înseamnă că itemii măsoară scorul adevărat ( $T$ ) pe aceeași

scală, cu același nivel de precizie și au aceeași cantitate de eroare. Spre deosebire de acest model, modelul  $\tau$ -echivalent esențial presupune că fiecare item măsoară același construct, pe aceeași scală, dar cu un nivel de precizie diferit, erorile fiind identice. Măsurarea pe aceeași scală înseamnă că toți itemii au aceeași varianță. Precizia diferită se referă la faptul că există itemi ale căror valori se grupează, în timp ce valorile altor itemi pot fi mai împrăștiate una față de alta. De exemplu, itemii a căror încărcare cu scorul adevărat este moderată pot corela mai bine între ei, decât itemii cu încărcare mare, al căror răspuns corect este asumat mai greu de subiecți.

Încălcarea condițiilor modelului  $\tau$ -echivalent esențial este responsabilă de subestimarea fidelității de către coeficientul Cronbach alfa, iar subestimarea este cu atât mai mare cu cât abaterea de la condițiile modelului este mai mare (Graham, 2006). În acest sens, o atenție specială trebuie acordată unidimensionalității, deoarece aceasta este cea mai frecvent încălcată condiție a modelului de măsurare (Huysamen, 2006; Widhiarso, 2007).

O estimare mai adecvată a fidelității se poate obține prin utilizarea altor indicatori ai fidelității decât Cronbach alfa: coeficientul de fidelitate *ro* al lui Raykov, *teta* al lui Armor, *ro* al lui Spearman (altul decât corelația ordinală) sau coeficientul  $\lambda_2$  ( $\lambda_2$ ) din seria celor 6 coeficienți *lambda* ai lui Gutmann (Garson, 2010; Sijtsma, 2009). Dintre aceștia, mai accesibil este coeficientul  $\lambda_2$ , care este calculat în SPSS și poate fi raportat în locul lui Cronbach alfa ( $\lambda_2$  este identic cu *alfa*). Astfel, de exemplu, în cazul scalei de evaluare a adaptării la stres SEAS-37 (Popa, 2010), valoarea lui Cronbach

*alfa* este de 0.830, în timp ce coeficientul  $\lambda_2$  este 0.835. Diferența dintre cei doi coeficienți nu este neapărat notabilă, dacă ținem cont de faptul că valoarea coeficientului de fidelitate se raportează, în mod normal, cu două zecimale.

De asemenea, o opțiune alternativă de luat în considerare, este coeficientul de fidelitate *teta* ( $\theta$ ), care se bazează pe varianța explicată de primul factor și este strâns legat de *Cronbach alfa* (Dugard, Todman, & Staines, 2010). Interpretarea coeficientului *teta* se face de regulă după același standard utilizat în cazul lui *Cronbach alfa*. Practic, pentru calcularea coeficientului *teta* este necesară efectuarea preliminară a unei analize factoriale exploratorii, reținându-se valoarea *eigenvalue* a primului factor.\* Formula de calcul este următoarea:

$$\theta = \left( \frac{n}{n-1} \right) \left( 1 - \frac{1}{\lambda} \right) \quad (6)$$

unde  $n$ =numărul itemilor scalei, iar  $\lambda$ =valoarea *eigenvalue* a primului factor, pentru soluția nerotită.

De exemplu, în cazul scalei *SEAS-37*, coeficientul de fidelitate *teta* este:

$$\theta = \left( \frac{37}{37-1} \right) \left( 1 - \frac{1}{6,625} \right) = 0,85$$

valoare ceva mai mare decât *Cronbach alfa* (0,83)

Din perspectiva completitudinii, coeficientul *Cronbach alfa* nu reprezintă „fidelitatea în general“, ci un context particular al acesteia. În conformitate cu *teoria generalizabilității*, fidelitatea se referă la capacitatea măsurării de a rămâne constantă de la un moment la altul, independentă de orice sursă posibilă de eroare (Cronbach, Rajaratnam, & Gleser, 1963). Esența acestei teorii constă în recunoașterea existenței mai multor surse de eroare a măsurării: utilizarea altor subiecți, utilizarea altor itemi, utilizarea altor evaluatori ș.a.. Pentru fiecare din aceste situații există o altă formă de validitate și implicit, un alt mod de estimare a acesteia. Coeficientul *Cronbach alfa* este adecvat doar pentru situația în care sursa de eroare pe care o considerăm relevantă este dată de utilizarea mai mul-

tor itemi, care descriu un anumit construct. Pentru situația în care sursa de eroare relevantă o reprezintă evaluatorii, de exemplu, se va apela la metoda fidelității inter-evaluatori.

Una dintre consecințele importante a teoriei generalizabilității este aceea că fidelitatea cuantificată cu ajutorul indicelui *Cronbach alfa* trebuie înțeleasă, nu doar ca o caracteristică a testului, ci și a populației din care a fost selecționat eșantionul (presupunând că este reprezentativ), precum și a condițiilor de administrare a testului (Pitariu & Albu, 1996). Din acest motiv, este extrem de important ca raportarea valorii coeficientului *Cronbach alfa* să nu se limiteze doar la valoarea lui ca atare, ci și la toate aspectele care descriu situația de măsurare. Vom putea utiliza cu încredere acel instrument pentru alte măsurări, numai dacă noile condiții sunt similare celor în care a fost calculat inițial *Cronbach alfa*. În caz contrar, se impune o nouă evaluare a consistenței interne.

## 6. *Cronbach alfa*, indicator afectat de caracteristicile datelor

Fiind o expresie a analizei corelaționale, nu ar trebui să fie o surpriză sensibilitatea coeficientului *Cronbach alfa* față de natura și caracteristicile datelor analizate. Teoretic, calcularea coeficientului *Cronbach alfa* presupune că variabilele sunt măsurate pe o scală cu valori cantitative, de tip continuu. Totuși, la fel ca și în cazul SEM, sunt acceptate și variabile măsurate la nivel ordinal, cum sunt, de exemplu, scalele de tip *Lickert*. Utilizând o serie de simulări cu metoda Monte Carlo, pe scale *Lickert* cu un număr diferit de variante de răspuns, Lozano et al. (2008) au constatat că valoarea lui *Cronbach alfa* tinde să fie cu atât mai mare cu cât numărul variantelor de răspuns la itemi este mai mare. Proprietățile psihometrice ale scalelor sunt afectate de numărul alternativelor de răspuns. Cu cât subiecții pot alege dintre mai multe variante, cu atât *Cronbach alfa* tinde să fie mai mare. Scalele cu 5 sau 7 variante de răspuns sunt de preferat. De asemenea, este important de remarcat faptul că cel mai scăzut nivel al lui *alfa* a fost obținut pe

\* Valorile *eigenvalue* reprezintă suma pătratelor coeficienților de corelație dintre fiecare item și factorul respectiv.

scafele cu trei niveluri de răspuns, chiar mai scăzut decât pentru scafele cu răspuns dihotomic.

*Cronbach alfa* este recunoscut pentru lipsa de robustețe în raport cu diversele încălcări ale condițiilor impuse de modelul de măsurare. Atât varianța cât și covarianța itemilor pot fi substanțial modificate chiar și de prezența câtorva valori eronate, cu atât mai mult în situațiile în care eșantionul este relativ redus (Christmann & Van Aelst, 2006). Lipsa de omogenitate a eșantionului poate fi, la rândul ei, o sursă de alterare a valorii lui *alfa*, dacă presupunem existența unui grup de subiecți care fac parte dintr-o altă populație decât cea vizată. În acest sens, Liu (2007) și Liu et al. (2010), atrag atenția cu privire la impactul valorilor extreme asupra lui *alfa*. Dacă valorile extreme simetrice nu afectează nivelul lui *alfa*, cele asimetrice, plasate doar pe una dintre laturile distribuției, au un efect inflaționist asupra lui *Cronbach alfa*. Aceste concluzii sugerează importanța unei analize preliminare atente a itemilor angajați în analiza de fidelitate.

### 8. *Cronbach alfa*, indicator afectat de eroarea de eșantionare

Cea mai paradoxală dintre greșelile de interpretare ale lui *Cronbach alfa* este aceea că, deși este un indicator al erorii de măsurare, raportarea lui se face, de regulă, fără a ține cont de faptul că el însuși este supus erorii. La fel ca orice indicator statistic, *alfa* este afectat de eroarea de eșantionare (Cortina, 1993; Duhachek & Iacobucci, 2004; Fan & Thompson, 2001; Koning & Franses, 2003). Cu alte cuvinte, valoarea calculată a lui *alfa* nu este decât o estimare a valorii adevărate a lui *alfa* la nivelul populației din care este selecționat eșantionul. Asumarea acestui adevăr ridică problema preciziei cu care valoarea calculată estimează valoarea adevărată, deci a limitelor de încredere pentru *Cronbach alfa* adevărat.

Principial, limitele de încredere pentru *alfa* pot fi înțelese, la fel ca pentru orice alt indicator statistic, ca intervalul care ar include valoarea adevărată a lui *alfa*, dacă acesta ar fi calculat pe un număr mare de eșantioane aleatoare de același fel. Practic, mărimea intervalului de încredere (precizia) lui *Cronbach alfa*, depinde de

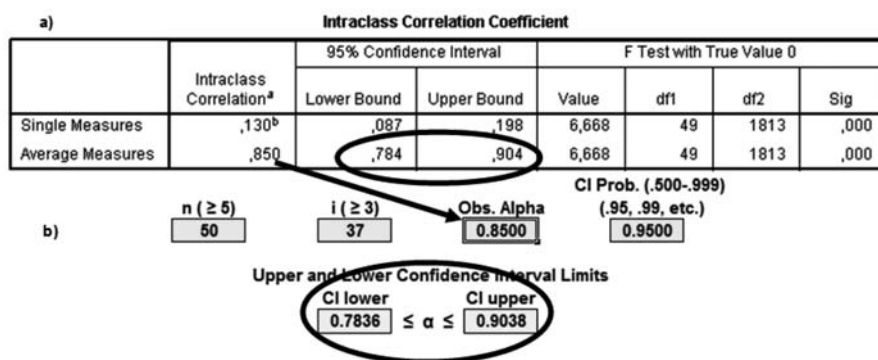
eroarea sa standard, care poate fi descrisă ca fiind expresia variabilității corelațiilor interitemi (Nunnally, apud Cortina, 1993). Precizia cu care *alfa* estimează valoarea sa adevărată depinde de variabilitatea coeficienților de corelație dintre itemi, indiferent care este sursa acestei împrăștieri. O împrăștiere mare a corelațiilor inter-itemi poate rezulta, fie din multidimensionalitatea constructului, fie ca urmare a erorii de eșantionare (Schmitt, 1996). Eroarea standard a lui *alfa* este cu atât mai mică cu cât corelațiile dintre itemi, volumul eșantionului sau numărul itemilor sunt mai mari (Duhachek & Iacobucci, 2004). Doi coeficienți *alfa* de aceeași valoare pot avea precizii diferite, în funcție de una sau alta dintre caracteristicile menționate, sau de combinația specifică dintre aceștia.

Așa cum există diferite metode de calcul al fidelității, eroarea standard este și ea calculabilă în moduri diferite. Din perspectiva practicienilor, abordarea limitelor de încredere pentru *Cronbach alfa* depinde de existența unor programe accesibile care să rezolve această problemă. Din fericire, astfel de programe există. Unul dintre acestea este *ScoreRel CI*, realizat în *Excel* de Barnette (2005), care este disponibil gratuit și poate fi descărcat de pe internet. Un alt program care calculează limitele de încredere pentru *Cronbach alfa* este chiar *SPSS*, sub forma valorii medii a coeficientului de corelație intraclasă (casetă *Reliability Analysis: Statistics* cu setările: *Intraclass correlation coefficient/Model: Two-Way Random/Type: Consistency*). Este posibilă și setarea unei valori de referință (implicit este 0), în raport cu care se va testa semnificația statistică a coeficientului *Cronbach alfa*, dar Fan și Thompson (2001) consideră că această opțiune nu are sens în acest context.

Pentru exemplificare, am calculat *Cronbach alfa* și limitele intervalului de încredere, pentru un eșantion de 50 de cazuri selecționate aleatoriu din cele 4180 aplicări ale scalei *SEAS-37*, cu *SPSS* și cu *ScoreRel CI* (fig. 4a și 4b).

Așa cum se poate observa, ambele programe dau valori practic identice pentru limitele intervalului de încredere. Având valoarea calculată 0,85, valoarea reală, cu un nivel de încredere de 95% se află în intervalul 0,78-0,90. În tabelul 4 am sintetizat o serie de valori ale lui *alfa* și limitele lor de încredere, calculate pe

Figura 4. Limitele de încredere pentru Cronbach alfa=0.899, calculate cu SPSS (a) și *Sorel CI* (b)



Tabelul 4. Limite de încredere ale lui *alfa* pentru eșantioane de mărimi diferite

N	<i>alfa</i>	Lim. inf	Lim. sup	Mărime interval
30	0,761	0,620	0,869	0,249
50	0,850	0,726	0,878	0,152
100	0,835	0,716	0,839	0,123
500	0,823	0,760	0,825	0,065
1000	0,827	0,758	0,797	0,039
4000	0,824	0,763	0,783	0,020

Tabelul 5. Variația limitelor de încredere pentru un chestionar cu 20 itemi, în funcție de N și valoarea lui *alfa*

N	<i>alfa</i> =0.60		<i>alfa</i> =0.70		<i>alfa</i> =0.80		<i>alfa</i> =0.90	
	inf	sup	inf	sup	inf	sup	inf	sup
20	0.29	0.81	0.46	0.86	0.64	0.90	0.82	0.95
30	0.35	0.78	0.51	0.83	0.67	0.89	0.83	0.94
40	0.39	0.75	0.54	0.81	0.69	0.87	0.84	0.93
50	0.41	0.74	0.56	0.80	<b>0.70</b>	<b>0.87</b>	0.85	0.93
60	0.43	0.73	0.57	0.79	0.71	0.86	0.85	0.93
70	0.45	0.72	0.58	0.79	0.72	0.86	0.86	0.93
80	0.46	0.71	0.59	0.78	0.73	0.85	0.86	0.92
90	0.46	0.71	0.60	0.78	0.73	0.85	0.86	0.92
100	0.47	0.70	0.60	0.77	0.73	0.85	0.86	0.92
200	0.51	0.67	0.63	0.75	0.75	0.83	0.87	0.91
300	0.53	0.66	0.64	0.74	0.76	0.83	0.88	0.91

eșantioane aleatorii de diferite mărimi, ale aceleiași scale *SEAS-37*. Așa cum se poate observa, indiferent de variațiile lui *alfa*, care sunt induse

de selecția de eșantionare, mărimea intervalului de încredere (diferența dintre limita superioară și cea inferioară) tinde să scadă sistematic, pe

**Tabelul 6.** Variația limitelor de încredere (95%) pentru un chestionar cu 50 itemi, în funcție de N și valoarea lui *alfa*

N	alfa=0.60		alfa=0.70		alfa=0.80		alfa=0.90	
	Inf	Sup	inf	sup	inf	sup	inf	sup
20	0.30	0.81	0.47	0.86	0.65	0.90	0.82	0.95
30	0.36	0.77	0.52	0.83	0.68	0.88	0.84	0.94
40	0.40	0.75	0.55	0.81	<b>0.70</b>	<b>0.87</b>	0.85	0.93
50	0.42	0.74	0.56	0.80	0.71	0.87	0.85	0.93
60	0.44	0.73	0.58	0.79	0.72	0.86	0.86	0.93
70	0.45	0.72	0.59	0.79	0.72	0.86	0.86	0.93
80	0.46	0.71	0.58	0.78	0.73	0.85	0.86	0.92
90	0.47	0.70	0.60	0.78	0.73	0.85	0.86	0.92
100	0.47	0.70	0.60	0.77	0.73	0.85	0.86	0.92
200	0.51	0.67	0.63	0.75	0.75	0.83	0.87	0.91
300	0.53	0.66	0.64	0.74	0.76	0.83	0.88	0.91

măsură ce volumul eșantionului crește. Acest lucru înseamnă că *alfa* este din ce în ce mai precis, iar valoarea lui prezintă mai multă încredere atunci când este calculată pe eșantioane mai mari.

Pentru a scoate în evidență variația limitelor de încredere pentru diferite niveluri ale lui *alfa* și mărimea eșantionului, am calculat valorile din tabelul 5, pentru un chestionar cu 20 de itemi, și din tabelul 6, pentru un chestionar cu 50 de itemi.

Asa cum se poate observa, dacă luăm în considerare valoarea calculată a lui *alfa* considerată acceptabilă (0,70) și un volum al eșantionului de 30 de subiecți, considerat de mulți (în mod gresit) ca acceptabil, atunci valoarea reală a lui *alfa* este cuprinsă, cu o probabilitate de 95%, în intervalul 0,51-0,83 (tab. 5), respectiv 0,52-0,83 (tab. 6). Pentru un chestionar cu 20 de itemi, limita inferioară a intervalului de încredere 0,70 este atinsă prima dată abia pentru un *alfa*=0,80 și 40 de subiecți (tab. 5), iar în cazul chestionarului cu 50 de itemi, pentru un *alfa*=0,80 și un eșantion de 50 de subiecți (tab. 6). Aceste valori ne ajută să înțelegem faptul că alegerea volumului eșantionului pentru calcularea fidelității trebuie să țină cont de mărimea estimată (dorită) a lui *Cronbach alfa* și de numărul de itemi ai chestionarului.

## 9. Concluzii

Încrederea în rezultatul măsurărilor și, de fapt, în toate concluziile pe care le vom trage utilizând măsurările respective, depinde de gradul lor de fidelitate. Acesta este motivul esențial pentru care problema fidelității este una centrală în evaluarea psihologică și în cercetarea științifică din acest domeniu (Wilkinson L. & Task Force on Statistical Inference, 1999). Coeficientul *Cronbach alfa* continuă să fie cel mai utilizat indicator al consistenței interne, dar interpretarea trebuie să țină cont de limitele acestuia, care ar putea fi astfel sintetizate:

- nu este o expresie a fidelității testului, ci a fidelității măsurării în condițiile respective și pentru populația din care este selecționat eșantionul;

- deși este condiționat de unidimensionalitatea itemilor, nu este un indicator al unidimensionalității;

- deși vizează consistența internă, nu este doar o expresie a acesteia, fiind influențat de numărul itemilor, mărimea corelațiilor inter-itemi și de volumul eșantionului;

- este afectat de anomalii ale datelor;

- este supus erorii de eșantionare, fiind imprecis în estimarea valorii reale la nivelul populației.

O soluție alternativă de optimizare a estimării fidelității, care câștigă popularitate printre psihologi, o reprezintă abordarea prin metoda modelării ecuației de structură (SEM). Avantajul SEM constă în faptul că permite abordarea unor modele variate, inclusiv cele extrem de complexe sau cu diverse constrângeri (Green & Yang, 2009). Totuși, practica arată că *alfa Cronbach* continuă să rămână una dintre cele mai accesibile metode de evaluare a erorii de măsurare în psihologie. În consecință, utilizarea corectă a acestui coeficient presupune respectarea unui set de recomandări de bune practici:

În absența altor abordări ale erorii, *Cronbach alfa* rămâne o opțiune viabilă, cu condiția interpretării lui corecte. Utilizarea unor instrumente psihologice fără nici o referire la eroarea de măsurare pe care o implică, reprezintă o limită serioasă în practica psihologică și în cercetarea științifică.

Reevaluarea fidelității unui instrument psihologic ori de câte ori se modifică condițiile de măsurare și populația pe care este aplicat. Un instrument cu fidelitate dovedită într-un anumit context, nu este în mod necesar fidel în orice context.

Interpretarea lui *Cronbach alfa* și alegerea „pragului acceptabil“ trebuie să țină cont de numărul itemilor, de mărimea și variabilitatea corelațiilor inter-itemi, precum și de volumul eșantionului. Raportarea matricei de corelații inter-itemi poate ajuta la înțelegerea mai bună a fidelității.

Evaluarea complementară a structurii factoriale și raportarea rezultatelor acesteia. Raportarea exclusivă a lui *Cronbach alfa* poate crea o imagine greșită cu privire la unidimensionalitatea itemilor.

Un aspect esențial, considerat obligatoriu, constă în evaluarea și raportarea preciziei lui *Cronbach alfa*, prin intermediul limitelor intervalului de încredere.

## Referințe bibliografice

Barnette, J. J. (2005). Scorerel Ci: An Excel Program For Computing Confidence Intervals For Commonly Used Score Reliability Coefficients. *Educa-*

*tional and Psychological Measurement*, 65(6), 980-983.

Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for Measurement Error Attenuation in Structural Equation Models: Some Important Reminders. *Educational and Psychological Measurement*, 57(5), 785-799.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71(3), 425-440.

Carretta, T. R., & Ree, M. J. (2001). Pitfalls of Ability Research. *International Journal of Selection and Assessment*, 9(4), 325-335.

Christmann, A., & Van Aelst, S. (2006). Robust estimation of Cronbach's alpha. *Journal of Multivariate Analysis*, 97(7), 1660-1674.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98-104.

Crețu, R. Z. (2005). *Evaluarea personalitatii: metode alternative*. Iași: Polirom.

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*(16), 3.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, 16(2), 137-163.

Dugard, P., Todman, J., & Staines, H. (2010). *Approaching Multivariate Statistics: A Practical Introduction* (Second ed.). London and New York: Routledge, Taylor Francis Group.

Duhachek, A., & Iacobucci, D. (2004). Alpha's Standard Error (ASE): An Accurate and Precise Confidence Interval Estimate *Journal of Applied Psychology*, 89(5), 792-808.

Ercan, I., Yazici, B., Sigirli, D., Ediz, B., & Kan, I. (2007). Examining Cronbach Alpha, Theta, Omega Reliability Coefficients According to the Sample Size. *Journal of Modern Applied Statistical Methods*, 6(1), 291-303.

Fan, X., & Thompson, B. (2001). Confidence Intervals About Score Reliability Coefficients, Please: An EPM Guidelines Editorial. *Educational and Psychological Measurement*, 61(4), 517-531.

Field, A. (2009). *Discovering Statistics Using SPSS* (Third ed.): SAGE Publications.

Garson, D. (2010). Statnotes: Topics in Multivariate Analysis – Reliability Analysis. Retrieved 08.12.2010, from <http://faculty.chass.ncsu.edu/garson/PA765/reliab.htm>.

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.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of Coefficient Alpha as an Index of Test Unidimensionality. *Educational and Psychological Measurement*, 37(4), 827-838.

- Green, S. B., & Yang, Y. (2009). Reliability of Summed Item Scores Using Structural Equation Modeling: An Alternative to Coefficient Alpha. *Psychometrika*, 74(1), 155-167.
- Gulliksen, H. (1950). *Theory Of Mental Tests*. New York: John Wiley & Sons, Inc. (versiune online: <http://www.questia.com/PM.qst?a=o&d=8818547#>).
- Huysamen, G. K. (2006). Coefficient Alpha: Unnecessarily Ambiguous; Unduly Ubiquitous. *SA Journal of Industrial Psychology*, 32(4), 34-40 (<http://www.sajip.co.za/index.php/sajip/article/viewFile/242/239>).
- Kanyongo, G. Y., Brook, G. P., Kyei-Blankson, L., & Gocmen, G. (2007). Reliability and Statistical Power: How Measurement Fallibility Affects Power and Required Sample Sizes for Several Parametric and Nonparametric Statistics. *Journal of Modern Applied Statistical Methods*, 6(1), 81-90.
- Kline, T. J. B. (2005). *Psychological testing: a practical approach to design and evaluation*. Thousand Oaks: Sage Publications.
- Koning, A. J., & Franses, P. H. (2003). *Confidence Intervals for Cronbach's Coefficient Alpha Values*: Erasmus Research Institute of Management: Report Series Research in Management (RS-2003-041-MKT) (<http://publishing.eur.nl/ir/repub/asset/431/ERS-2003-041-MKT.pdf>).
- Liu, Y., Wu, A. D., & Zumbo, B. D. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Ordinal/rating scale item responses. *Educational and Psychological Measurement*, 70(1), 5-21.
- Liu, Y., & Zumbo, B. D. (2007). The Impact of Outliers on Cronbach's Coefficient Alpha Estimate of Reliability: Visual Analogue Scales. *Educational and Psychological Measurement*, 67, 620-634.
- Lozano, L. M., Garcia-Cueto, E., & Muniz, J. (2008). Effect of the Number of Responses Categories on the Reliability and Validity of Rating Scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 4(2), 73-79.
- McNemar, Q. (1946). Opinion-Attitude Methodology. *Psychological Bulletin* 43, 289-374.
- Muchinsky, P. M. (1996). The Correction for Attenuation. *Educational and Psychological Measurement*, 56(1), 63-75.
- Pitariu, H., & Albu, M. (1996). *Psihologia personalului: I. Măsurarea și interpretarea diferențelor individuale*. Cluj-Napoca: Editura Presa Universitară Clujeană.
- Popa, M. (2010). *Scala de evaluare a adaptării la stres (SEAS-37)*. Unpublished manuscript.
- Raykov, T. (1997). Estimation of Composite Reliability for Congeneric Measures. *Applied Psychological Measurement*, 21(2), 173-184.
- Ree, M. J., & Carretta, T. R. (2006). The Role of Measurement Error in Familiar Statistics. *Organizational Research Methods*, 9(1), 99-112.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8, 350-353.
- Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*, 74(1), 107-120.
- Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15(1), 72-101.
- Tan, Ș. (2009). Misuses of KR-20 and Cronbach's Alpha Reliability Coefficients. *Education and Science*, 34(152), 101-112 (<http://egitimvebilim.ted.org.tr/index.php/EB/article/viewFile/595/175>).
- ten Holt, J. C., van Duijn, M. A. J., & Boomsma, A. (2010). Scale construction and evaluation in practice: A review of factor analysis versus item response theory applications *Psychological Test and Assessment Modeling*, 52(3), 272-297
- Widhiarso, W. (2007). Estimate Reliability Measurement for Multidimensional Scales *Unpublished Research Summary. Faculty of Psychology. Gadjah Mada University* ([http://academia.edu.documents.s3.amazonaws.com/1803835/Estimate\\_Reliability\\_Measurement\\_for\\_Multidimensional\\_Scales.pdf](http://academia.edu.documents.s3.amazonaws.com/1803835/Estimate_Reliability_Measurement_for_Multidimensional_Scales.pdf)).
- Wilkinson L. & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist* (10.04.2009: <http://www.loyola.edu/library/ref/articles/Wilkinson.pdf>), 54, 594-604.
- Zeller, R. A. (Ed.) (2001) *Encyclopedia of Sociology* (Second ed., Vols. 4). Macmillan Reference USA.
- Zimmerman, D. W., & Williams, R. H. (1997). Properties of the Spearman Correction for Attenuation for Normal and Realistic Non-Normal Distributions. *Applied Psychological Measurement*, 21(3), 253-270.