

---

# Tra Alfa e Omega c'è di mezzo la CFA?\*

Massimiliano Pastore

Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova.

**Sommario** L'Alpha di Cronbach è uno degli indicatori di affidabilità più utilizzati nella ricerca psicologica nonostante un'ampia letteratura recente ne abbia messo in discussione l'utilità. In questo lavoro riassumiamo brevemente il contesto di applicazione e le criticità di Alpha. Quindi, con un semplice esperimento Monte Carlo confrontiamo le performance di tale indice con tre alternative: l'indice Omega e due indici di adattamento comunemente utilizzati nell'analisi fattoriale confermativa (CFA). Dall'esperimento è emerso che Alpha performa bene solo in particolari condizioni ottimali, difficilmente riscontrabili nella pratica. In altri contesti, in particolare quando i fattori sottostanti sono due piuttosto che uno solo, tende a produrre valori eccessivamente alti. La conclusione è che questo indice non può essere considerato la migliore opzione per la valutazione dell'affidabilità. Piuttosto, si suggerisce di utilizzare un approccio basato sulla CFA, che fornisce sia informazioni più complete sulle strutture fattoriali sottostanti, sia indici che permettono una migliore valutazione dell'affidabilità.

**Parole chiave:** Affidabilità, Alpha di Cronbach, Omega, Analisi fattoriale confermativa.

## Is there CFA between Alpha and Omega?

**Summary** Cronbach's Alpha is one of the most popular measure of reliability in psychological research, despite recent literature has extensively questioned its utility. In this work we summarize briefly contexts of application and critical aspects. By presenting a simple Monte Carlo experiment, we compare the performance of Alpha index with three alternatives: the Omega index and two fit indices used in the confirmatory factor analysis (CFA). The experiment suggests that Alpha shows adequate performance only in specific conditions, which are hard to find in practice. In other contexts, and especially when the underlying factors are two rather than one, Alpha tends to lead too high values. We conclude that Alpha cannot be considered the best option for reliability assessment. Rather, it is suggested the adoption a CFA-based approach that provides both more complete information on underlying factor structures, and indices that allow a more accurate reliability assessment.

*La corrispondenza va inviata a*

*Massimiliano Pastore*

*c/o Dipartimento di Psicologia dello Sviluppo e della Socializzazione*

*Via Venezia, 8 I-35131 Padova (PD), Italy.*

*Email: massimiliano.pastore@unipd.it*

---

\* Please cite as: Pastore, M. (2017). Tra Alfa e Omega c'è di mezzo la CFA? *Giornale Italiano di Psicologia*, 44, 761–780.

## 1 INTRODUZIONE

L'Alpha di Cronbach è certamente una tra le misure di affidabilità più diffuse nella Psicologia, e forse la statistica più diffusa in assoluto come indicatore della qualità di un punteggio di un test (Sijtsma, 2009). Paradossalmente, il successo e la diffusione di questo indice non sembrano diminuire nonostante una letteratura sempre più ampia ne abbia messo più volte in evidenza le criticità, in passato (es. Green, Lissitz, & Mulaik, 1977; Schmitt, 1996) ed anche recentemente (Dunn, Baguley, & Brunsten, 2014; Peters, 2014). In particolare, Cho and Kim (2015) hanno evidenziato sei errate interpretazioni di Alpha molto diffuse, tra cui il fatto che sia ritenuto un indicatore di consistenza interna e la migliore scelta tra i vari coefficienti di affidabilità. In realtà le alternative migliori sono già da tempo disponibili e sono state ampiamente discusse (si veda ad es.: Bentler, 2009; Green & Yang, 2009; Revelle & Zinbarg, 2009; Zinbarg, Revelle, Yovel, & Li, 2005); tra le varie proposte, quella che sta raccogliendo un ampio consenso è il coefficiente Omega (McDonald, 1999). Si tratta di un indice che viene calcolato a partire dagli stessi algoritmi di stima utilizzati nei modelli di analisi fattoriale confermativa (*Confirmatory Factor Analysis*, CFA; Bollen, 1989) ma, per quanto ne sappiamo, ad oggi non è mai stato messo direttamente a confronto con i tradizionali indici che si usano con questi modelli per valutare la bontà di adattamento.

L'obiettivo di questo lavoro è di confrontare le performance di Alpha, Omega e due indici di adattamento comunemente utilizzati nella CFA in un semplice esperimento Monte Carlo al fine di valutare se questi ultimi non siano già sufficienti a fornire informazioni utili in relazione all'affidabilità. Nella seconda sezione del lavoro discuteremo il significato dell'affidabilità nel contesto della Teoria Classica dei Test (TCT), dei problemi connessi all'uso di Alpha, e di Omega come proposta alternativa, condivisa in letteratura, per il calcolo dell'affidabilità. Nella terza sezione descriveremo l'esperimento Monte Carlo effettuato e, di seguito, nella quarta i risultati più interessanti. Infine, discuteremo le conclusioni e le implicazioni più rilevanti.

## 2 ALPHA, OMEGA E LA TEORIA CLASSICA DEI TEST

Secondo la TCT il punteggio osservato,  $x_{ip}$ , di un soggetto  $i$  su un item  $p$  è la somma di due componenti: il punteggio vero  $t_{ip}$  e l'errore  $e_{ip}$

$$x_{ip} = t_{ip} + e_{ip} \quad (1)$$

sotto le assunzioni che: 1) il valore atteso degli errori è zero; 2) gli errori non sono correlati con i punteggi veri. Il punteggio di un soggetto su una scala si ottiene sommando tutti i punteggi nei  $P$  item che la compongono:  $X_i = \sum_{p=1}^P x_{ip}$ . Allo stesso modo possiamo definire le quantità  $T_i = \sum_{p=1}^P t_{ip}$  e  $E_i = \sum_{p=1}^P e_{ip}$  rispettivamente come il punteggio vero e l'errore di una scala, da cui si ha  $X_i = T_i + E_i$ . La varianza dei punteggi di una scala può essere quindi scomposta nella somma

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \quad (2)$$

L'affidabilità di una scala è definita come il rapporto tra la varianza vera e la varianza totale della scala:

$$\rho_{XX'} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2} = \frac{\sigma_T^2}{\sigma_X^2} \quad (3)$$

Dalla lettura di quest'ultima formula è chiaro che l'affidabilità risulta influenzata da tutto ciò che agisce sull'errore; di conseguenza, più che essere una proprietà di un test, l'affidabilità diventa la proprietà di una scala applicata in uno specifico contesto ad una particolare popolazione (Dunn et al., 2014; Miller, 1995; Thompson & Vacha-Haase, 2000). Un'altra importante limitazione, sempre derivante dalla definizione, è l'impossibilità di conoscere l'esatto valore di affidabilità, valido per ogni possibile situazione/contesto a causa della variabilità associata a  $T$  ed  $E$ , che non può essere perfettamente stimata in quanto soggetta all'errore di campionamento. Inoltre, per come è definito, il modello non è identificato in quanto esistono infinite combinazioni di modi in cui la varianza di  $X$  può essere ripartita tra  $T$  ed  $E$  (Graham, 2006).

Di conseguenza, per poter stimare l'affidabilità è necessario imporre dei vincoli o definire delle assunzioni che, tanto più sono restrittive, tanto meno facilmente possono essere presenti nei dati osservati (Dunn et al., 2014). In particolare, il modello che definisce Alpha si caratterizza per le seguenti proprietà: 1) gli item misurano lo stesso tratto latente (unidimensionalità), 2) le varianze dei punteggi veri degli item devono essere costanti. Questa seconda proprietà, chiamata *essentially tau-equivalence* (Raykov, 1997a), viene raramente rispettata in pratica nelle scale psicologiche (Green & Yang, 2009) e pertanto le stime di Alpha risultano spesso distorte (Lord & Novick, 1968; Graham, 2006).

Nel lungo dibattito che prosegue da anni, vari autori hanno messo in evidenza che la stima di Alpha può essere distorta se aumenta la correlazione tra gli errori di ciascun item ed il numero di item (Cortina, 1993; Yuan & Bentler, 2002), che dipende dalla qualità del campionamento, in particolare se i punteggi derivano da popolazioni diverse (Waller, 2008), che può essere influenzata dalla dimensione campionaria, dal numero di categorie di risposta utilizzate e dal numero di item (Cortina, 1993; Duhachek, Coughlan, & Iacobucci, 2005; Lissitz & Green, 1975; Weng, 2004). Ancora, Alpha non è una misura della consistenza interna di un test (Sijtsma, 2009) e neppure di unidimensionalità (Schmitt, 1996; Ten Berge & Sočan, 2004). Per questo, Alpha è considerato da molti, piuttosto che una vera misura di affidabilità, una specie di limite inferiore di affidabilità (Sijtsma, 2009). Alcuni autori ritengono che, per una interpretazione più consona e più "affidabile", sia necessario almeno tenere in considerazione e valutare l'intervallo di confidenza ad esso associato (Maydeu-Olivares, Coffman, & Hartmann, 2007). Altri autori invece, sono concordi sull'abolizione dell'utilizzo di questo indice per sostituirlo con indicatori di migliore qualità informativa (Cho & Kim, 2015; Dunn et al., 2014; Peters, 2014; Sijtsma, 2009).

Una misura che si sta affermando come alternativa è il coefficiente Omega (McDonald, 1999) che si è rivelato essere un indice più sensibile della consistenza interna sia rispetto all'Alpha che ad altre alternative proposte (es.: Graham, 2006; Raykov, 1997a; Revelle & Zinbarg, 2009; Zinbarg et al., 2005; Zinbarg, Yovel, Revelle, & McDonald, 2006; Zinbarg, Revelle, & Yovel, 2007). Omega infatti performa meglio nelle situazioni, frequenti in psicologia, di violazione della *tau-equivalenza*, ed in generale riduce il rischio, rispetto ad Alpha, di sotto o sovra stimare l'affidabilità. In altri termini, gli assunti richiesti da Omega sono meno restrittivi e più realistici ed il calcolo dell'intervallo di confidenza associato a tale indice riflette molto meglio la variabilità del processo di stima offrendo un maggiore grado di accuratezza delle stime stesse (Dunn et al., 2014).

## 2.1 Calcolo di Alpha e Omega

Il calcolo di Alpha si basa sulla matrice di covarianza  $\mathbf{S}$  ottenuta sui  $p$  item di una scala; sia  $s_{ij}$ , con  $i, j = 1, \dots, p$ , l'elemento  $ij$ -esimo della matrice  $\mathbf{S}$ , la formula per il calcolo di Alpha è:

$$\alpha = \frac{p}{p-1} \left( \frac{1 - \sum_i s_{ii}}{\sum_{ij} s_{ij}} \right) \quad (4)$$

Il calcolo di Omega è un po' più complesso e si basa sulla struttura di un modello fattoriale caratterizzato da una sola componente latente. In altri termini, coerentemente con l'assunto di unidimensionalità, si considera un modello di misura in cui i  $p$  item siano espressione di un unico tratto latente o fattore. Su tale modello vengono stimate le saturazioni (o *factor loadings*;  $\lambda_i$ ,  $i = 1, \dots, p$ ) e le varianze d'errore ( $\psi_{ii}$ ) associate agli item. Si rammenta che le saturazioni misurano la relazione tra il fattore latente ed i singoli item, e sulla base di tale relazione possiamo valutare quanto bene gli item siano espressione di quest'ultimo. Quindi, la formula per il calcolo è (McDonald, 1999):

$$\omega = \frac{(\sum_i \lambda_i)^2}{(\sum_i \lambda_i)^2 + \sum_i \psi_{ii}} \quad (5)$$

La stima delle saturazioni avviene utilizzando i classici metodi dei modelli CFA. In particolare, nel presente lavoro faremo riferimento al metodo di Massima Verosimiglianza (*Maximum Likelihood*, ML) ed al *Diagonally Weighted Least Squares* (DWLS). Il primo è il metodo di stima più noto e richiede che le variabili utilizzate siano quantitative ed abbiano distribuzioni approssimativamente simmetriche. Il secondo è considerato migliore quando si hanno variabili discrete con poche categorie di risposta, come ad esempio le scale Likert, ed è quindi più indicato quando si trattano variabili tipiche dei questionari psicologici (Flora & Curran, 2004; Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Lionetti, Keijsers, Dellagiulia, & Pastore, 2016; Rhemtulla, Brosseau-Liard, & Savalei, 2012; Yang-Wallentin, Jöreskog, & Luo, 2010).

Sulla base della formula (5) siamo pertanto legittimati ad accostare l'indice Omega ai modelli di CFA ed a come questi ultimi possano essere utilizzati per la valutazione dell'affidabilità. L'idea non è nuova, Bollen (1989, pag. 221) aveva già descritto come il valore di  $R^2$  associato ad una variabile  $x_i$  in un modello confermativo possa essere considerato una misura di affidabilità alternativa ad Alpha. Più recentemente, Green and Yang (2009) hanno illustrato come, concretamente, l'affidabilità possa essere stimata direttamente utilizzando i valori dei parametri ottenuti con CFA sia con uno stimatore ML, che con approccio robusto basato su DWLS. In sostanza, dal momento che il calcolo di Omega deriva dallo stesso algoritmo che stima i parametri del modello fattoriale, ci possiamo aspettare che vi sia una associazione tra Omega e gli indici che misurano la bontà di adattamento della CFA.

Operativamente, per il calcolo di Omega si possono utilizzare vari pacchetti di R (R Core Team, 2016), disponibili gratuitamente, tra cui: `coefficientalpha` (Zhang & Yuan, 2015, 2016), `MBESS` (Kelley, 2016), `psych` (Revelle, 2016), `semTools` (semTools Contributors, 2016). Quest'ultimo pacchetto permette di calcolare Omega, a partire dall'output di un modello di analisi fattoriale confermativa, utilizzando l'approccio descritto da Green and Yang (2009). Ancora, Deng and Chan (2016) hanno messo a disposizione on-line una routine apposita, sempre scritta in R, scaricabile alla pagina web: [https://www.psy.cuhk.edu.hk/psy\\_media/WChan\\_Page/alpha-omega.txt](https://www.psy.cuhk.edu.hk/psy_media/WChan_Page/alpha-omega.txt). In alternativa si può procedere con un qualunque software che stimi i parametri del modello fattoriale e successivamente utilizzare la formula (5).

### 3 ESPERIMENTO

L'obiettivo del nostro esperimento è confrontare le performance di Alpha e Omega con alcuni indici di adattamento dei modelli fattoriali confermativi. In particolare, abbiamo preso in considerazione 2 indici comunemente utilizzati nella letteratura: il *Tucker-Lewis Index* o *Nonnormed Fit index* (TLI; Tucker & Lewis, 1973; Bentler & Bonett, 1980) ed il *Root-Mean-Square Error of Approximation* (RMS; Steiger & Lind,

1980). Questi due indici misurano l'adattamento del modello testato e si considerano buoni quando sono rispettivamente maggiore di 0.95 e minore di 0.06 (Hu & Bentler, 1999).

Per semplicità ci siamo concentrati solo sulla condizione ideale in cui tutti gli item hanno le stesse saturazioni, varianze residue identiche e presentano distribuzioni approssimativamente simmetriche. Si tratta in pratica di una condizione definita *parallel model*, più restrittiva della *tau-equivalenza* in quanto prevede che i punteggi veri  $T$  e gli errori  $E$  siano uguali in tutti gli item (Raykov, 1997a, 1997b; Graham, 2006). Questa scelta, pur essendo limitata, ha comunque il vantaggio di mettere in evidenza le criticità associate alle performance di Alpha, più volte ribadite in letteratura, nelle sue condizioni ideali di applicazione, che sappiamo essere raramente rispettate nell'ambito delle scale psicologiche (Green & Yang, 2009).

### 3.1 Simulazione

Il programma per la simulazione Monte Carlo è stato scritto in ambiente R (R Core Team, 2016) utilizzando i pacchetti `lavaan` (Rosseel, 2012), e `semTools` (semTools Contributors, 2016). Sono stati considerati i seguenti 4 fattori sperimentali:

- Tipologia di modello vero ( $M$ ) a 4 livelli: un solo fattore ( $m_1$ ), due fattori con lo stesso numero di item e correlazione  $\phi = 0, 0.3, 0.6$  (rispettivamente  $m_2, m_3$  e  $m_4$ ); le saturazioni (i *factor loadings*) sono state fissate a 0.6 per tutti gli item, indipendentemente dal modello.
- Numero di item ( $P$ ) della scala, con 2 livelli: 8 e 12.
- Tipo di scala Likert ( $Q$ ), con 3 livelli: 3, 5, 7 modalità di scelta.
- Numerosità campionaria ( $N$ ), con 3 livelli: 30, 100, 300.

Dati  $m, p, q$  e  $n$  i distinti livelli dei fattori  $M, P, Q$  e  $N$ , rispettivamente, sono stati ripetuti 1000 volte per ciascuna delle  $4 \times 2 \times 3 \times 3 = 72$  combinazioni i seguenti passaggi:

1. Generato un set di  $p$  item di dimensione  $n$  sulla base del modello  $m$  (con struttura latente a uno o due fattori con correlazione  $\phi$ ) e  $q$  valori discreti. Per definire questi ultimi sono state utilizzate soglie basate sulla distribuzione binomiale in modo da garantire una sottostante distribuzione approssimativamente normale (per maggiori dettagli si veda ad es.: Jöreskog & Sörbom, 1996; Pastore, 2013).
2. Stimato gli indici di adattamento del modello con il metodo della Massima Verosimiglianza (ML) e calcolato sugli stessi dati Alpha e Omega.
3. Stimato gli indici di adattamento del modello in forma robusta con il metodo *Diagonally Weighted Least Squares* (DWLS) e calcolato Alpha e Omega trattando i dati come variabili ordinali (per maggiori dettagli si veda Green & Yang, 2009).

Indipendentemente dal tipo di modello generativo dei dati ( $M$ ), per le stime dei parametri (ed il calcolo degli indici di adattamento) è stato utilizzato un modello a fattore singolo. In questo modo, nel caso  $m_1$  il modello fattoriale era specificato correttamente mentre negli altri tre casi no, in quanto la vera struttura generativa degli item era a due fattori. Pertanto, nella condizione ideale di corretta specificazione del modello, ci aspettiamo che tutti gli indici funzionino bene e che siano correlati tra di loro. Negli altri tre casi ci attendiamo dei diversi comportamenti degli indici in funzione della vera correlazione tra i fattori. Sappiamo che TLI e RMS sono sensibili all'errata specificazione del modello (si veda ad es. Fan & Sivo, 2005, 2007) e pertanto ci aspettiamo che tali indici si modifichino indicando la non appropriatezza del modello ad un fattore in questi casi, mentre Alpha ed Omega, nelle stesse condizioni, siano meno influenzati.

Al termine della procedura sono stati generati 72000 set di dati e calcolati altrettanti indici Alpha, Omega, TLI e RMS, rispettivamente con i due metodi di stima LM e DWLS.

	ML								DWLS							
	$n$	$\bar{x}$	ds	med	min	max	skew	kurt	$n$	$\bar{x}$	ds	med	min	max	skew	kurt
Alpha	70673	0.72	0.11	0.74	0.00	0.93	-1.00	1.73	71438	0.75	0.10	0.77	0.01	0.95	-1.15	2.39
Omega	70688	0.65	0.19	0.71	0.00	0.94	-1.02	0.22	71402	0.69	0.24	0.77	0.00	1.00	-1.71	1.88
TLI	62522	0.64	0.22	0.62	0.00	1.00	0.01	-1.03	63798	0.67	0.26	0.71	0.00	1.00	-0.45	-0.99
RMS	70688	0.10	0.05	0.11	0.00	0.30	-0.30	-0.55	71456	0.11	0.07	0.12	0.00	0.50	0.03	-0.45

**Tabella 1:** Statistiche descrittive degli indici calcolati nella simulazione. ML = stime di massima verosimiglianza, DWLS = stime Diagonally Weighted Least Squares;  $n$  = numerosità effettiva,  $\bar{x}$  = media, ds = deviazione standard, med = mediana, min = minimo, max = massimo, skew = skewness, kurt = kurtosis.

## 4 RISULTATI

### 4.1 Soluzioni accettabili

Dal totale di 72000 campioni di dati generati, abbiamo escluso i casi per i quali sono stati riscontrati problemi di convergenza o di stima dei parametri. In relazione ai modelli fattoriali abbiamo una percentuale che potremmo definire *fisiologica* di “non convergenze/soluzioni non accettabili” pari a circa l’1% dei casi, più o meno equivalente nei due metodi di stima. Gli indici Alpha ed Omega sono stati stimati il 100% delle volte con ML, mentre con DWLS in 6 casi (circa lo 0.01%) non è stato possibile. Per uniformità di trattamento dei dati abbiamo comunque deciso di tenere solo i casi per cui avessimo ottenuto le stime di tutti gli indicatori considerati ovvero 70688 (circa il 98%) con gli item trattati come variabili quantitative e 71456 (circa il 99%) con gli item trattati come variabili ordinali. Va ribadito, ad ulteriore supporto di questa scelta, che Omega viene calcolato utilizzando i parametri stimati nel modello CFA (vedi equazione 5) pertanto se questi ultimi non sono accettabili di conseguenza non può esserlo neppure Omega.

Successivamente abbiamo analizzato le distribuzioni dei valori di Alpha, Omega e dei 2 indici di fit selezionati (TLI e RMS) e stimati sulla base dei due metodi (ML e DWLS). Sono stati quindi esclusi i valori di Alpha minori di zero (rispettivamente 15 casi su 70688, 0.02%, con ML e 18 casi su 71456, 0.02%, con DWLS), i valori di Omega maggiori di uno (54 casi su 71456, 0.08%, con DWLS) ed i valori di TLI fuori dal range 0-1 (rispettivamente 8166 casi su 70688, 11.55%, con ML e 7658 casi su 71456, 10.72%, con DWLS). Negli altri casi non sono stati trovati dei valori anomali o comunque al di fuori del range atteso. In tabella 1 sono riportate le statistiche descrittive relative a ciascuno degli indici considerati sui casi rimasti.

### 4.2 Effetti dei fattori sperimentali

Come prima operazione abbiamo stimato quanto i vari fattori sperimentali avessero effetto sul comportamento degli indici. Dato che le distribuzioni campionarie ottenute presentavano livelli più o meno accentuati di asimmetria abbiamo scelto di utilizzare dei modelli lineari generalizzati (*Generalized Linear Models*, GLM) con funzione legame la distribuzione Gamma (McCullagh & Nelder, 1989; Dobson, 2002; Wood, 2006). Le analisi con i GLM (una per ciascun indice trattato come variabile dipendente) hanno permesso di ripartire la quota di devianza spiegata dal modello nelle differenti componenti del disegno sperimentale. Per ciascun indice è stato possibile pertanto stimare la devianza attribuita ad un fattore ( $D_{source}$ ) e con questa calcolare un indice di dimensione dell’effetto con la seguente statistica:

$$\delta = 100 \times \frac{D_{source}}{D_{null}} \quad (6)$$

in cui  $D_{null}$  è la devianza del modello nullo ovvero il modello contenente la sola intercetta; in tabella 2 sono riportati tutti i valori ottenuti in dettaglio.

	ML				DWLS			
	Alpha	Omega	TLI	RMS	Alpha	Omega	TLI	RMS
<i>M</i>	43	46	50	19	43	16	49	28
<i>P</i>	13	1	0	0	12	1	0	0
<i>Q</i>	2	0	0	0	0	0	0	1
<i>N</i>	0	0	3	1	0	0	1	0
<i>M:P</i>	3	0	0	0	3	0	0	0
<i>M:Q</i>	0	0	0	0	0	0	0	1
<i>P:Q</i>	0	0	0	0	0	0	0	0
<i>M:N</i>	0	1	2	12	0	0	11	10
<i>P:N</i>	0	0	1	0	0	0	0	0
<i>Q:N</i>	0	0	0	0	0	0	0	0
<i>M:P:Q</i>	0	0	0	0	0	0	0	0
<i>M:P:N</i>	0	0	1	0	0	0	1	0
<i>M:Q:N</i>	0	0	0	0	0	0	0	0
<i>P:Q:N</i>	0	0	0	0	0	0	0	0
<i>M:P:Q:N</i>	0	0	0	0	0	0	0	0
<i>genR</i> <sup>2</sup>	61	49	56	32	58	17	62	41

**Tabella 2:** Percentuale di devianza spiegata ( $\delta$ ) negli indici considerati a carico dei fattori del disegno sperimentale:  $M$  = tipo di modello,  $P$  = numero di item,  $Q$  = tipo di scala likert,  $N$  = numerosità campionaria. Nell'ultima riga è riportata la stima percentuale di devianza spiegata complessivamente dal modello ( $genR^2$ ; Wood, 2006), calcolata con la formula:  $(D_{null} - D_{source})/D_{null}$ .

Dalla lettura di questa tabella risulta evidente che la tipologia di modello (fattore  $M$ ) influenza molto il comportamento dei vari indici: quando i valori sono stimati con ML si va dal 19% per RMS al 50% per TLI, quando sono trattati come variabili ordinali il range varia tra il 16% (Omega) ed il 49% (TLI). Alpha risulta essere influenzato in parte anche dal numero di item (fattore  $P$ ), rispettivamente per il 13 e 12% circa; inoltre vi sono un effetto di interazione  $M \times N$  sull'indice RMS (rispettivamente 12 e 10%) e sull'indice TLI, quando viene stimato con DWLS (circa 11%). Tutti gli altri effetti sono al massimo del 3% per cui li possiamo considerare ininfluenti. Pertanto, nella sezione successiva, considereremo le performance degli indici solo in relazione al fattore sperimentale  $M$ .

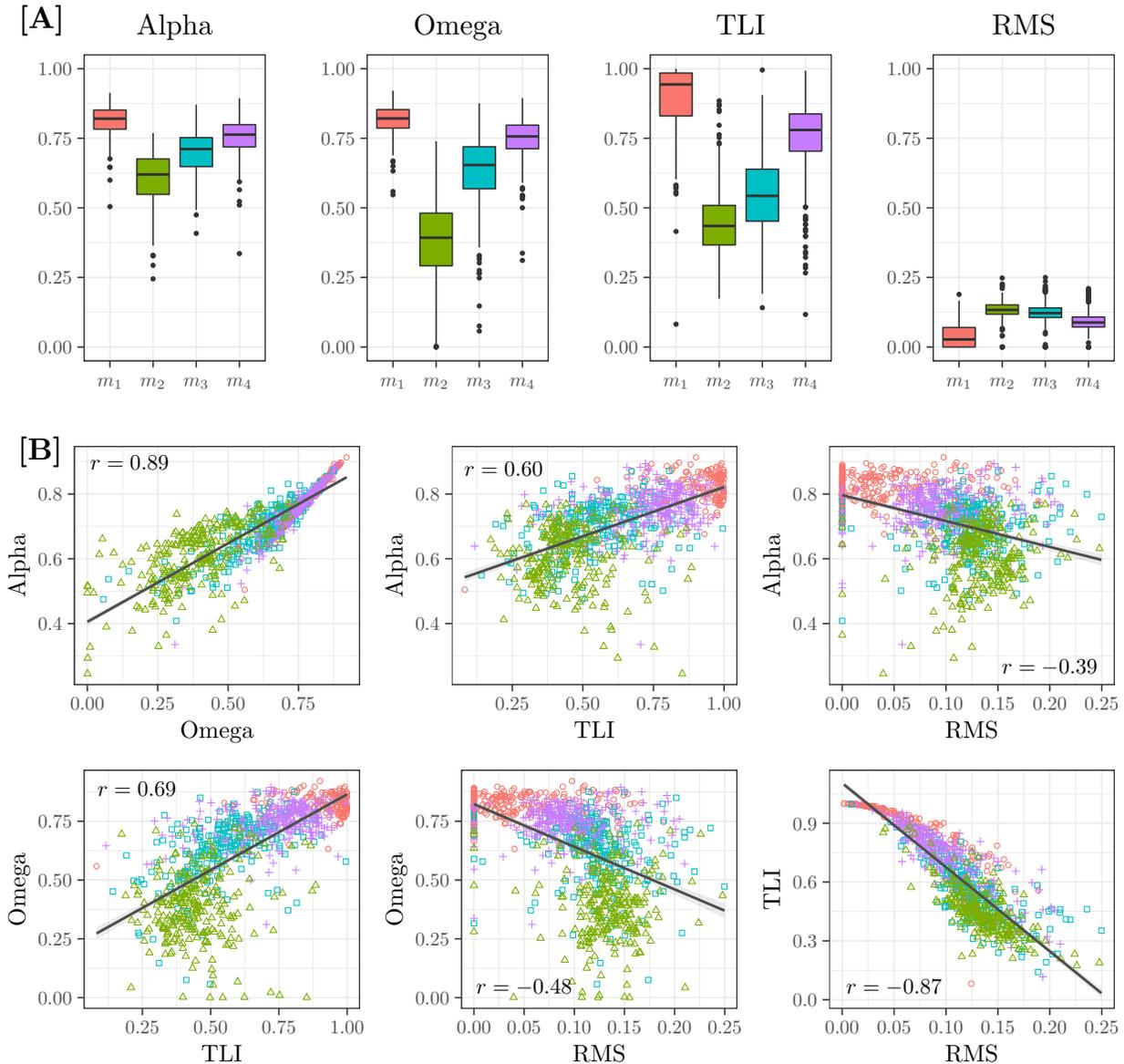
### 4.3 Distribuzioni univariate e correlazioni tra gli indici

Per valutare le distribuzioni e la relazione tra i quattro indici considerati prendiamo in esame prima le stime ottenute con ML e successivamente le stime ottenute con DWLS.

#### Stime ML

La figura 1 è composta di 2 pannelli che presentano due tipi di informazioni relative agli indici considerati. Nel pannello [A] sono rappresentate le distribuzioni univariate degli indici. Ognuno dei boxplot riportati nei quadranti rappresenta in particolare la distribuzione di un indice rispetto ad un livello del fattore  $M$ , da sinistra a destra rispettivamente:  $m_1$ ,  $m_2$ ,  $m_3$ ,  $m_4$ . Il livello  $m_1$  indica il caso ideale in cui gli item fanno parte di un solo fattore, gli altri livelli sono relativi a condizioni in cui gli item sono in realtà parte di due fattori latenti con correlazioni pari a 0, 0.3, 0.6 rispettivamente. Nel pannello [B] sono rappresentate le distribuzioni bivariate delle coppie di indici, sempre in funzione dei livelli di  $M$ , riconoscibili per i diversi tipi di simboli:  $m_1$  cerchi ( $\circ$ ),  $m_2$  triangoli ( $\Delta$ ),  $m_3$  quadrati ( $\square$ ) e  $m_4$  il simbolo +.

A livello univariato, si osserva che gli indici performano al meglio nella condizione ideale (livello  $m_1$ ), come ci si poteva attendere, ed in generale migliorano all'aumentare della correlazione tra i fattori. Nel caso peggiore, ovvero quando ci sono due fattori non correlati (livello  $m_2$ ), Alpha presenta comunque valori generalmente troppo alti: il 75% dei valori risulta superiore a 0.55, il 25% superiore a 0.68. Nella stessa condizione gli altri indici sembrano essere decisamente più sensibili: il 75% dei valori di Omega è inferiore a 0.49, il 98% dei valori di TLI è inferiore alla soglia di accettazione (0.95) e così pure il 98% dei valori di RMS al di sopra della soglia di accettazione (0.06). Questo risultato conferma quindi che Alpha non è molto sensibile alla presenza di più fattori, anche se non correlati. L'aumento delle correlazioni tra i fattori



**Figura 1:** Stime ML: [A] distribuzioni univariate dei quattro indici Alpha, Omega, TLI e RMS rispetto ai quattro livelli del fattore  $M$ ; [B] distribuzioni bivariate delle coppie di indici rispetto ai livelli del fattore  $M$ : un solo fattore (cerchi,  $\circ$ ), due fattori con correlazione 0 (triangoli,  $\Delta$ ), 0.3 (quadrati,  $\square$ ) e 0.6 (+). I punti rappresentati (1058) sono un campione pari all'1.5% del totale di valori generati con MC.

(condizioni  $m_3$  e  $m_4$ ) comporta anche una minore sensibilità di tutti gli indici alla errata specificazione. Si osserva però che il 92% dei valori di TLI e l'86% dei valori di RMS continuano ad indicare che il modello ad un fattore non è accettabile anche nella condizione  $m_4$ , ovvero quando i fattori sono molto correlati ( $\phi = .6$ ). Nella stessa condizione, i valori di Alpha e Omega sono superiori a .70 nel 75% dei casi.

A livello bivariato osserviamo che Alpha correla, in generale, con gli altri indici, precisamente  $r = 0.89$  con Omega, 0.6 con TLI e  $-0.39$  con RMS, rispettivamente. Omega ha correlazioni più alte sia con TLI (0.69) che con RMS ( $-0.48$ ). Se teniamo conto dei diversi livelli del fattore  $M$ , rappresentati in figura 1, pannello [B], da simboli diversi, è interessante osservare soprattutto le relazioni tra gli indici nella condizione peggiore ( $m_2$ ). In questo caso, Alpha correla poco sia con TLI ( $r = -0.03$ ) che con RMS ( $r = 0.17$ ); lo stesso vale per Omega (0.13, con TLI e 0.03 con RMS). In pratica, nessuno di questi due indici sembra essere sensibile ad una specificazione del modello così nettamente errata.

### Stime DWLS

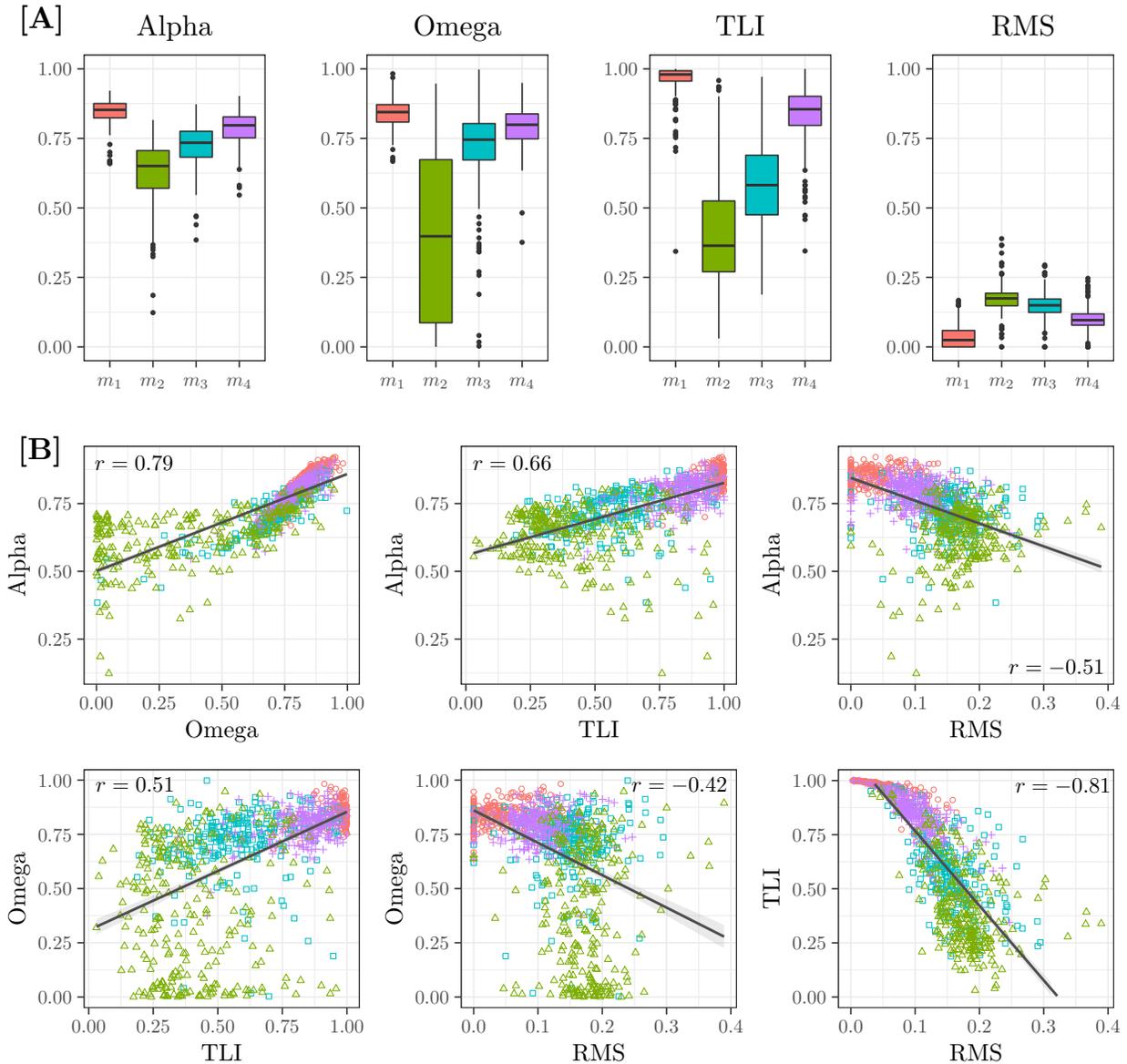
In figura 2 sono rappresentate le stesse informazioni della precedente figura 1 sui valori ottenuti con le stime DWLS. A livello univariato (pannello [A]) osserviamo ancora che tutti gli indici risultano sensibili alla presenza dei due fattori ma in maniera piuttosto diversa tra loro. Come nel caso delle stime ML, e come atteso, le prestazioni migliori si hanno nel caso ideale ( $m_1$ , primo box a sinistra), le peggiori nel caso in cui ci siano due fattori indipendenti ( $m_2$ ), con l'aumento della correlazione tra i fattori, tutti i valori degli indici tendono a migliorare. In particolare, nella condizione  $m_2$  (secondo box da sinistra), il 50% dei valori di Alpha risulta superiore a 0.65. Nella stessa condizione, i valori di Omega superiori a 0.65 sono circa il 29%, i valori di TLI superiori alla soglia di accettabilità (0.95) solamente lo 0.54% e quelli di RMS inferiori a 0.06 appena l'1.34%. Nella condizione  $m_4$  (ultimo box a destra), il 75% dei valori di Alpha e Omega risulta superiore a 0.75, mentre TLI e RMS presentano valori che portano a non rigettare il modello rispettivamente nel 9.48% e 11.73% dei casi.

Le correlazioni tra gli indici si mantengono medio/alte anche con le stime DWLS (figura 2, pannello [B]): Alpha presenta in generale i valori 0.79 con Omega, 0.66 con TLI e  $-0.51$  con RMS, rispettivamente; Omega ha correlazioni più basse sia con TLI (0.59) che con RMS ( $-0.42$ ). Ancora, nella condizione peggiore ( $m_2$ ), le correlazioni di Alpha con gli altri tre indici sono rispettivamente 0.58,  $-0.06$  e 0.16, le correlazioni di Omega rispettivamente 0.07 con TLI e 0.05 con RMS. Si conferma anche con questo metodo di stima che Alpha e Omega sono poco sensibili alla presenza di due fattori piuttosto che uno solo.

### 4.4 Analisi delle soglie ottimali

Per una valutazione più specifica, abbiamo considerato le performance dei quattro indici in relazione a delle soglie comunemente indicate come ottimali. In tabella 3 sono riassunte le percentuali di casi in cui i quattro indici hanno superato tali soglie in funzione del tipo di modello vero  $M$ . Per Alpha e Omega abbiamo utilizzato come soglia il valore 0.7 (Cho & Kim, 2015), per TLI e RMS quelli suggeriti da Hu and Bentler (1999); dato che questi ultimi suggeriscono che i due indici debbano essere letti in coppia, abbiamo anche valutato la percentuale di casi in cui contemporaneamente si avesse  $TLI \geq 0.95$  e  $RMS \leq 0.06$  e riportato queste percentuali nella colonna indicata con HB.

La prima riga della tabella ci illustra che, in condizioni ideali, Alpha ed Omega performano molto bene, con una percentuale di modelli accettati intorno al 97/98%, superiore a quella degli altri due indici. Il caso che comunque ci sembra più interessante è quello relativo a  $m_2$  in cui vediamo che Alpha, stimato con ML,



**Figura 2:** Stime DWLS: [A] distribuzioni univariate dei quattro indici Alpha, Omega, TLI e RMS rispetto ai quattro livelli del fattore  $M$ ; [B] distribuzioni bivariate delle coppie di indici rispetto ai livelli del fattore  $M$ : un solo fattore (cerchi,  $\circ$ ), due fattori con correlazione 0 (triangoli,  $\Delta$ ), 0.3 (quadrati,  $\square$ ) e 0.6 (+). I punti rappresentati (1058) sono un campione pari all'1.5% del totale di valori generati con MC.

supera la soglia il 15% delle volte. Sarebbe come a dire che c'è una probabilità del 15% avere un valore alto di Alpha nonostante la scala non sia composta da un unico fattore. Se teniamo conto della natura ordinale dei dati, tale probabilità praticamente raddoppia, fino al 29%. Nella stessa condizione osserviamo invece che TLI e RMS, presi in coppia (come suggerito da Hu & Bentler, 1999), portano correttamente al rifiuto del modello circa il 97% delle volte, indipendentemente dal metodo di stima utilizzato.

$M$	ML				HB	DWLS				HB
	Alpha	Omega	TLI	RMS		Alpha	Omega	TLI	RMS	
$m_1$	97.4	97.3	49.2	73.1	48.8	98.9	98.7	95.5	3.8	60.5
$m_2$	15.1	2.7	0.3	2.4	0.3	29.0	22.5	39.3	17.8	0.4
$m_3$	51.8	30.9	0.6	3.9	0.6	66.4	67.3	59.0	14.9	0.8
$m_4$	82.8	79.5	3.3	14.5	3.3	92.7	90.6	83.2	10.0	5.8

**Tabella 3:** Percentuali di casi in cui gli indici hanno superato i valori soglia ottimali in funzione del fattore  $M$ : Alpha e Omega  $\geq 0.7$ , TLI  $\geq 0.95$ , RMS  $\leq 0.06$ . HB indica la percentuale di casi in cui contemporaneamente TLI  $\geq 0.95$  e RMS  $\leq 0.06$  (Hu & Bentler, 1999).

## 5 CONCLUSIONI

L'esperimento presentato in questo lavoro è volutamente limitato ad una situazione semplice e finalizzato a valutare la performance di Alpha e Omega in una situazione particolarmente critica (presenza di due fattori non correlati). In questa condizione, nessuno dei due indici si comporta in maniera soddisfacente, invece gli indici di adattamento della CFA funzionano bene, presentando valori al di sotto della soglia di accettazione del modello. Rimandiamo gli interessati ad altri lavori, disponibili in letteratura, in cui sono stati presi in considerazione casi più complessi, ad esempio distribuzioni asimmetriche sugli item (Green & Yang, 2009) oppure modelli con fattori di secondo ordine (Cho & Kim, 2015).

Come ci attendevamo, nessuno dei due indici ha un comportamento ottimale al di fuori della situazione ideale, a differenza di TLI e RMS che invece si comportano secondo le attese. È plausibile che, in condizioni peggiori e più realistiche, Alpha e Omega vadano ancora peggio, mentre la vasta letteratura sugli indici della CFA dimostra che sono più informativi in molte situazioni non ottimali ad esempio quando il modello non è correttamente specificato oppure è congenerico, ovvero non rispetta la proprietà di tau-equivalenza (es.: Enders & Finney, 2003; Fan & Sivo, 2005, 2007; Fan, Thompson, & Wang, 1999; Fan & Wang, 1998; Gerbing & Anderson, 1993; Hu & Bentler, 1998, 1999; Marsh, Hau, & Wen, 2004; Wu & West, 2010; Yu & Muthén, 2002).

Da molti autori Alpha viene visto come una misura di consistenza interna (Christensen, Johnson, & Turner, 2011; Nunnally & Bernstein, 1994; Rubin & Babbie, 2008; Thompson, 2003) che secondo Cho and Kim (2015, pag. 215) può essere interpretata in tre modi: come omogeneità, interrelazione tra un insieme di item, o come saturazione relativa ad un fattore generale. L'omogeneità può essere riferita all'unidimensionalità di un insieme di item (Cortina, 1993; Green et al., 1977; McDonald, 1981; Schmitt, 1996) ovvero alla presenza di un fattore latente da essi misurato (Hattie, 1985). L'interrelazione viene definita come media aritmetica dei coefficienti di correlazione interitem (Cronbach, 1951). Infine, il fattore di saturazione generale indica la proporzione di varianza del test che si attribuisce a questo fattore generale (Revelle & Zinbarg, 2009). Appare evidente, anche dalla nostra semplice simulazione, che gli indici di adattamento TLI e RMS premettono di valutare tutti questi aspetti mentre Omega solo in parte ed Alpha no.

In pratica, l'approccio CFA permette di avere informazioni sulla consistenza interna e sull'affidabilità sia dei singoli item che dell'insieme completo (si veda ad es. Bollen, 1989, pagg. 218-222) anche in situazioni in cui gli item non rispettino le assunzioni ottimali richieste da Alpha. In più, possiamo avere informazioni generali sulla qualità della scala, ad esempio evidenziando la presenza di situazioni problematiche se l'algoritmo di stima dei parametri non converge, o dei singoli item, dall'analisi delle saturazioni e delle varianze residue. Ancora, la CFA dispone di molti altri indici di adattamento, non presi in considerazione nel presente lavoro, e permette di valutare anche strutture più articolate e complesse. E sempre nel contesto della CFA, sono disponibili varie alternative ad Alpha proposte nel corso degli anni come ad esempio Fornell and Larcker

(1981), Werts, Jöreskog, and Linn (1972) o, più recentemente, Green and Yang (2009), Raykov and Shrout (2002), lo stesso indice Omega, come abbiamo visto, si basa sulle stime della CFA.

In conclusione, è ancora utile calcolare l'Alpha? Dipende dall'uso che ne vogliamo fare; di certo dobbiamo essere consapevoli che si tratta di una misura che dà solo una informazione piuttosto limitata, ossia, un valore alto di Alpha è solo preliminare ad un'analisi più approfondita e necessaria. Come minimo, se proprio volessimo continuare ad utilizzarla, dovremmo sempre associarla ad un intervallo di confidenza (come suggerito da Maydeu-Olivares et al., 2007), per avere almeno un'idea della variabilità associata alla sua stima. Citando lo stesso Cronbach (2004): *The Alpha coefficient was developed out of the history that emphasized a crossed design used for measuring differences among persons. This is now seen to cover only a small perspective of the range of measurement uses for which reliability information is needed*; quindi Alpha non è la migliore opzione per il calcolo dell'affidabilità di una scala. In alternativa, ci sembra che l'approccio CFA, ormai ampiamente diffuso e facilmente accessibile, permetta di avere molte più informazioni e sia decisamente consigliabile.

## Riferimenti bibliografici

- Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74*(1), 137–143.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588–606.
- Bollen, K. (1989). *Structural equations with latent variables*. New York: John Wiley.
- Cho, E., & Kim, S. (2015). Cronbach's coefficient alpha well known but poorly understood. *Organizational Research Methods*, *18*(2), 207–230.
- Christensen, L. B., Johnson, B., & Turner, L. A. (2011). *Research methods, design, and analysis*. Boston MA: Pearson.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98–104.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.
- Cronbach, L. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and psychological measurement*, *64*(3), 391–418.
- Deng, L., & Chan, W. (2016). Testing the Difference Between Reliability Coefficients Alpha and Omega. *Educational and Psychological Measurement*, *77*(2), 185–203.
- Dobson, A. (2002). *An Introduction to Generalized Linear Models* (2nd ed.). Chapman and Hall/CRC.
- Duhachek, A., Coughlan, A. T., & Iacobucci, D. (2005). Results on the standard error of the coefficient alpha index of reliability. *Marketing Science*, *24*(2), 294–301.
- 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.
- Enders, C., & Finney, S. (2003). SEM fit index criteria re-examined: An investigation of ML and robust fit indices in complex models. In *annual meeting of the American Educational Research Association, Chicago*.
- Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, *12*(3), 343–367.
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529.

- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56–83.
- Fan, X., & Wang, L. (1998). Effects of potential confounding factors on fit indices and parameter estimates for true and misspecified SEM models. *Educational and Psychological Measurement*, 58(5), 701–735.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods*, 9(4), 466–491.
- Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, 16(4), 625–641.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, 39–50.
- Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 40–65). Sage Publications.
- 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.
- Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied psychological measurement*, 9(2), 139–164.
- Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, 3(4), 424.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Jöreskog, K. G., & Sörbom, D. (1996). *PRELIS 2 User's Reference Guide: A Program for Multivariate Data Screening and Data Summarization: a Preprocessor for LISREL*. Scientific Software International.
- Kelley, K. (2016). MBESS: The MBESS R Package [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=MBESS> (R package version 4.0.0)
- Lionetti, F., Keijsers, L., Dellagiulia, A., & Pastore, M. (2016). Evidence of Factorial Validity of Parental Knowledge, Control and Solicitation, and Adolescent Disclosure Scales: When the Ordered Nature of Likert Scales Matters. *Frontiers in Psychology*, 7, 941. doi: 10.3389/fpsyg.2016.00941
- Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60(1), 10–13.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Cambridge, MA: Addison-Wesley.
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural equation modeling*, 11(3), 320–341.
- Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological methods*, 12(2), 157–176.

- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. CRC press.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*(1), 100–117.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Laurence Erlbaum Associates.
- Miller, M. B. (1995). Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. *Structural Equation Modeling*, *2*(1), 255–273.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
- Pastore, M. (2013). Le tecniche di simulazione Monte Carlo. In C. Barbaranelli & S. Ingoglia (Eds.), *I modelli di equazioni strutturali: temi e prospettive* (pp. 295–324). LED, Milano.
- 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*, *16*(2), 56–69.
- R Core Team. (2016). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Raykov, T. (1997a). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, *21*(2), 173–184.
- Raykov, T. (1997b). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research*, *32*(4), 329–353.
- Raykov, T., & Shrout, P. E. (2002). Reliability of scales with general structure: Point and interval estimation using a structural equation modeling approach. *Structural equation modeling*, *9*(2), 195–212.
- Revelle, W. (2016). psych: Procedures for Psychological, Psychometric, and Personality Research [Computer software manual]. Evanston, Illinois. Retrieved from <http://CRAN.R-project.org/package=psych> (R package version 1.6.6)
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145–154.
- Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological methods*, *17*(3), 354–373.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from <http://www.jstatsoft.org/v48/i02/>
- Rubin, A., & Babbie, E. (2008). *Research methods for social work*. Belmont, CA: Thompson Brooks/Cole.
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment*, *8*(4), 350–353.
- semTools Contributors. (2016). semTools: Useful tools for structural equation modeling [Computer software manual]. Retrieved from <http://cran.r-project.org/package=semTools> (R package version 0.4-12)
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120.
- Steiger, J. H., & Lind, J. C. (1980, May). Statistically based tests for the number of common factors. In *Annual meeting of the Psychometric Society, Iowa City, IA* (Vol. 758, pp. 424–453).
- Ten Berge, J. M., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*(4), 613–625.
- Thompson, B. (2003). Understanding reliability and coefficient alpha, really. In B. Thompson (Ed.), *Score reliability: Contemporary thinking on reliability issues* (pp. 3–23). Thousand Oaks, CA: SAGE.
- Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational*

- and *Psychological Measurement*, 60(2), 174–195.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10.
- Waller, N. G. (2008). Commingled samples: A neglected source of bias in reliability analysis. *Applied Psychological Measurement*, 32(3), 211–223.
- Weng, L.-J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological Measurement*, 64(6), 956–972.
- Werts, C. E., Jöreskog, K. G., & Linn, R. L. (1972). A multitrait-multimethod model for studying growth. *Educational and Psychological Measurement*, 32(3), 655–678.
- Wood, S. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.
- Wu, W., & West, S. G. (2010). Sensitivity of fit indices to misspecification in growth curve models. *Multivariate Behavioral Research*, 45(3), 420–452.
- Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory Factor Analysis of Ordinal Variables With Misspecified Models. *Structural Equation Modeling-A Multidisciplinary Journal*, 17(3), 392–423.
- Yu, C.-Y., & Muthén, B. (2002). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. In *annual meeting of the American Educational Research Association, New Orleans, LA*.
- Yuan, K.-H., & Bentler, P. M. (2002). On robustness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika*, 67(2), 251–259.
- Zhang, Z., & Yuan, K.-H. (2015). coefficientalpha: Robust Coefficient Alpha and Omega with Missing and Non-Normal Data [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=coefficientalpha> (R package version 0.5)
- Zhang, Z., & Yuan, K.-H. (2016). Robust Coefficients Alpha and Omega and Confidence Intervals With Outlying Observations and Missing Data Methods and Software. *Educational and Psychological Measurement*, 76(3), 387–411.
- Zinbarg, R. E., Revelle, W., & Yovel, I. (2007). Estimating  $\omega_h$  for structures containing two group factors: Perils and prospects. *Applied Psychological Measurement*, 31(2), 135–157.
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's  $\alpha$ , Revelle's  $\beta$ , and McDonald's  $\omega_H$ : Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133.
- Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for  $\omega_h$ . *Applied Psychological Measurement*, 30(2), 121–144.