

雙因素模型下測驗分數信度係數之比較

鄧祈允¹ 黃柏儔²

摘要

在估計測驗分數的信度時，應先考慮測驗的向度（dimensionality），若測驗結構為多向度時，則應使用多向度信度估計法，常見的方法包括分層 α 、Revelle 的 β 、多向度 ω ，以及階層 ω 。然而，過往研究並未仔細檢驗前述多向度估計法的差異，故本研究在真實模型為雙因素模型（bi-factor model）的假設下，系統性地操弄一般因素負荷量、子因素負荷量、子因素數目，以及樣本數，以探討前述四種信度估計法之表現。當以各信度係數之母體參數作為真實數值時，模擬研究結果顯示僅在子因素負荷量較高時，階層 ω 才會與另外三個信度係數有明顯不一致之真實數值，此意味著在子因素負荷量的真實數值較小時，分層 α 、 β 、多向度 ω 所得到的測驗分數信度估計值仍具有相當高之參考性，其中的 β 與階層 ω 展現最為一致之真實數值。當比較各方法的估計表現時，結果顯示相較於多向度 ω 、分層 α ，以及 β ，階層 ω 橫跨所有情境都展現較大的相對偏誤（relative bias）與變異性（variability），此顯示階層 ω 雖在估計概念較為合理，但於有限樣本之實徵表現不佳。本研究結果建議，在估計多向度測驗分數的信度時，不應僅參考階層 ω ，仍需使用其他多向度信度估計指標做為輔助，以對測驗分數的信度有較佳的估計與詮釋。

關鍵詞：Revelle 的 β 、分層 α 、多向度 ω 、信度、階層 ω 、雙因素模型

1. 鄧祈允，國立政治大學心理學系博士生

2. 黃柏儔，國立政治大學心理學系副教授

收件日期：2020.05.18；完成修改：2021.05.20；正式接受：2021.06.21

通訊作者：黃柏儔；Email：psyphh@nccu.edu.tw

地址：116011 臺北市文山區指南路二段 64 號 國立政治大學心理學系

A Comparison of Reliability Coefficients for Test Scores Under Bi-factor Models

Chi-Yun Deng¹ Po-Hsien Huang²

Abstract

When estimating a test score's reliability, the dimensionality of the test needs to be considered. If the test is not unidimensional, multidimensional reliability coefficients, such as stratified alpha, Revelle's beta, multidimensional omega, and hierarchical omega, should be used. However, these multidimensional coefficients might not be carefully distinguished in practical works. This study aimed to investigate these coefficients' meanings and properties. We used bi-factor models as true models for data generation and manipulated loadings of general and group factors, numbers of group factors, and sample sizes. The results showed that: (1) the true value of hierarchical omega was significantly different from other coefficients only when group factor loadings were high; (2) the true value of Revelle's beta was most similar to the hierarchical omega; (3) hierarchical omega yielded larger relative bias and variability than other coefficients under all conditions. These results implied other multidimensional reliability coefficients were still valuable when group factor loadings were low and hierarchical omega didn't perform well under finite sample settings, although hierarchical omega's meaning was most interpretable. These results suggested hierarchical omega should not be considered as "golden standard", other coefficients were still useful to interpret measurement results.

Keywords: bi-factor model, hierarchical omega, multidimensional omega, reliability, Revelle's beta, stratified alpha

-
1. Chi-Yun Deng, Ph.D. student, Department of Psychology, National Chengchi University
2. Po-Hsien Huang, Associate Professor, Department of Psychology, National Chengchi University

Received: 2020.05.18; Revised: 2021.05.20; Accepted: 2021.06.21

Corresponding Author: Po-Hsien Huang; Email: psyphh@nccu.edu.tw

Address: No. 64, Sec. 2, Zhinan Rd., Wenshan Dist., Taipei City 116011, Taiwan
Department of Psychology, National Chengchi University

壹、緒論

一、研究問題與背景

在社會科學領域中，測驗（test）常用於評估個人的潛在能力水準以協助決策，而信度指的是測驗分數之精確性（Lord & Novick, 1968）。在真實分數模型（true score model）的架構下，測驗分數被假設是由真分數（true score）和誤差（error）所組成。常用的信度估計法，包括：再測信度（test-retest）、複本信度（alternative form），以及內部一致性信度係數（如 Cronbach, 1951; Guttman, 1945; Murphy & Davidshofer, 1988）。其中， α 係數是最常被使用的內部一致性指標（Green & Yang, 2009a, 2009b; Hogan, Benjamin, & Brezinski, 2000; Sijtsma, 2009）。儘管如此， α 係數只有在試題滿足本質 τ 相等（essentially tau-equivalency）之條件時，才能正確地反映測驗分數的信度（Green & Yang, 2009a, 2009b, 2015; Sijtsma, 2009）。相較之下，在因素分析架構下所發展之 ω 係數，僅假設各試題需滿足同屬結構（congeneric structure），故 ω 係數適用於較廣泛之應用情境，不少學者亦認為 ω 係數乃為較適切之內部一致性指標（如 Green & Yang, 2009a; McDonald, 1970; Sijtsma, 2009）。

前述的 α 與 ω 係數皆僅在測驗符合單向度結構時，才能正確地反映信度。然而，測驗工具常展現多向度之結構（Drolet & Morrison, 2001; Green & Yang, 2009a），以「流行病學研究中心之憂鬱量表」（Center for Epidemiological Studies Depression, CES-D）為例，其包含了生理症狀、人際互動、正向情緒、負向情緒等四個因素（Williams et al., 2007）。此時，測驗編製者應採用考慮到多向度結構之信度指標，才可正確地反映測驗分數之信度。本研究的第一個研究目的為探討數種多向度信度指標之估計表現，包括：分層 α （stratified-alpha）（Cronbach, Schönemann, & McKie, 1965）、Revelle 的 β （Revelle, 1979）、多向度 ω （multidimensional-omega）、階層 ω （hierarchical-omega）（McDonald, 1970, 1999）。

然而，過往有些研究對於多向度組合測驗的背後或估計法的模型結構沒有做詳細之區分（如 Cronbach et al., 1965; Kamata, Turhan, & Darandari, 2003; Margono, 2015），使得估計法背後所假定模型中的真分數為何，以及得出的

信度係數意義為何，成了值得探討的議題。從因素分析的觀點來看，多向度組合測驗的結構可能為二階（second-order）／高階因素模型（high order factor model）或雙因素模型（bi-factor model），其中雙因素模型較能清晰的表徵與解釋測驗分數信度，也較能涵蓋其他模型的表徵方式（Reise, 2012; Revelle & Condon, 2019），故本研究的第二個目的為在雙因素模型架構下，探討測驗組合分數與信度之意涵。

二、單向度測驗組合分數的信度估計

在古典測驗理論（classical test theory）的架構下，信度之定義仰賴真分數模型，其假設個體於第 i 題之觀測分數（observed score, X_i ），可寫為真分數（true score, T_i ）與誤差分數（error, E_i ）之加總，即 $X_i = T_i + E_i$ ，而觀測分數之信度則被定義為真分數變異數所占觀察分數變異數之比例，即 $\rho = \frac{\text{var}(T)}{\text{var}(X)}$ ，從迴歸的角度來看，其反映了觀測分數能被真分數解釋之百分比（Green & Yang, 2015; Kelley & Pornprasertmanit, 2016; Lord & Novick, 1968; McDonald, 1999; Murphy & Davidshofer, 1988）。

古典測驗理論的信度定義於測驗領域中被廣為使用（Murphy & Davidshofer, 1988），而在諸多的信度估計程序與指標中， α 係數乃時常被使用之指標，其公式為：

$$\alpha = \frac{p}{p-1} \left(1 - \frac{\sum_{i=1}^p \sigma_{X_i}^2}{\sigma_X^2} \right) \quad \text{公式(1)}$$

在上述公式中， p 為總題數， σ_X^2 為各試題加總分數之變異數，而 $\sigma_{X_i}^2$ 為各題分數之變異數。 α 係數的優點在於容易計算，但其只有在各試題分數符合本質 τ 相等（essentially tau-equivalency），即各試題所對應之真分數僅相差一常數時，才會是信度的一致估計式（Cronbach, 1951; Green & Yang, 2009a; Lord & Novick, 1968; McDonald, 1999）。

相較於古典測驗理論僅區分真分數與誤差分數，模型取向（model based）則進一步引入潛在因素，以描述個體特質水準與觀測分數間之關係。以單向度測驗為例，其考慮單因素模型 $X_i = \mu_i + \lambda_i F + E_i$ ，在這裡， F 為個體之潛在因素水準， E_i 為第 i 題之誤差， λ_i 為第 i 題之因素負荷量（factor loading）， μ_i 為第 i 題之截距。McDonald（1970）將前述單因素模型中的 $\mu_i + \lambda_i F$ 視為真

分數，發展了 ω 係數，其公式為：

$$\omega = \frac{\text{var}(\sum_{i=1}^p \lambda_i F)}{\sigma_X^2} = \frac{(\sum_{i=1}^p \lambda_i)^2}{\sigma_X^2} \quad \text{公式(2)}$$

相較於 α 係數， ω 係數的優勢在於並未假設各試題須符合本質 τ 相等，僅需滿足同屬量測（congeneric measurement）之假設，即各試題皆用於測量潛在因素 F 。過去的研究顯示， ω 係數在更複雜的測驗結構，或違反古典測驗理論所需的假設時（如不符合本質 τ 相等假設），會有較小的估計偏誤（如 McDonald, 1970, 1981, 1999; Yang & Green, 2010）。

三、多向度測驗組合分數的信度估計

由於實際場域中多是使用多向度組合測驗，因此不少學者提出其他的估計法，以解決多向度測驗組合分數信度的估計議題（如 Cronbach et al., 1965; Green & Yang, 2009b; Kelley, & Pornprasertmanit, 2016; McDonald, 1970, 1981, 1999; Revelle, 1979）。在本研究中，研究者考慮分層 α 、Revelle 的 β 、多向度 ω ，以及階層 ω 等方法。接下來，將對各方法進行簡短之介紹。

（一）分層 α

分層 α ，其簡寫為 α_s ，是 Cronbach 等人（1965）將 α 係數推廣至多向度情況下之估計法，其適用由 k 個分量表組成，或有子因素（group factor）存在之測驗結構。分層 α 假設各題目可被分類（classified）到各個分層（strata），且同一分層的試題須滿足本質 τ 相等的關係（Cronbach et al., 1965）。分層 α 的信度計算方法為：

$$\alpha_s = 1 - \frac{\sum_{j=1}^k \sigma_j^2 (1 - \alpha_j)}{\sigma_X^2} \quad \text{公式(3)}$$

在上述計算公式中， σ_j^2 為第 j 個分量表分數之變異數， α_j 為第 j 個分量表分數之信度， σ_X^2 為總分之變異數。

（二）Revelle 的 β

β 乃藉由 α 係數是所有折半信度的平均值此一概念作為出發點所發展之

估計法，其被定義為最差（the worst）的折半信度數值（Revelle, 1979; Zinbarg, Revelle, Yovel, & Li, 2005）。透過 β 之定義，可用來檢視一份測驗是否為單向度。若一測驗分數的 β 值高，表示兩折半測驗分數間的相關很高，暗示測驗內容一致性高，很可能為單向度測驗；反之，若是得到低的 β 值，則表示兩折半測驗間未必只有一般因素（general factor），可能還有其他子因素，因此整份測驗未必是單向度測驗。從 β 係數的定義與特性來看，它不只可估計非單向度測驗分數的信度，也可做為檢視測驗向度的指標，且適用於各種測驗結構的信度估計（Revelle, 1979）。當兩折半測驗（ X 和 X' ）的長度相等時，其計算公式為：

$$\beta = \frac{4\sigma_{XX'}}{\sigma_{(X+X')}^2} \quad \text{公式(4)}$$

其中， $\sigma_{XX'}$ 乃所有折半方式中最小之共變數，而 $\sigma_{(X+X')}^2$ 為總分之變異數。若兩折半測驗的長度不相等時，則計算方式需稍作調整（Revelle, 1979）。

（三）多向度 ω

原用於單向度情境的 ω 係數（McDonald, 1970, 1999; Equation 6.20a），後來被拓展成多向度情境之信度估計法——多向度 ω （如 Kamata et al., 2003; Margono, 2015; Revelle & Condon 2019; Widhiarso & Ravand, 2014），簡寫為 ω_t 。 ω_t 除了因素和試題間須為線性關係與誤差間彼此無相關外，幾乎不需其他假設。 ω_t 假設觀測分數對應的因素模型為 $X_i = \mu_i + \lambda_{i1}F_1 + \lambda_{i2}F_2 + \lambda_{i3}F_3 + \cdots + \lambda_{ij}F_k + E_i$ ，其中 i 為題號， j 為因素之下標， X_i 、 E_i 分別表示第 i 題的觀測分數與測量誤差， μ_i 為第 i 題之截距， F_1 、 F_2 、 $F_3 \dots F_k$ 為該題對應的 k 個因素， λ_{i1} 、 λ_{i2} 、 $\lambda_{i3} \dots \lambda_{ik}$ 則為該題對應各因素之因素負荷量。在此架構下， ω_t 的計算方式為：

$$\omega_t = \frac{\text{var}(\sum_{j=1}^k \sum_{i=1}^p \lambda_{ij}F_j)}{\sigma_x^2} = 1 - \frac{\sum_{i=1}^p \psi_i^2}{\sigma_x^2} \quad \text{公式(5)}$$

在此， ψ_i^2 表示 E_i 之變異數。從公式(5)可見 ω_t 承襲公式(2)的觀測分數拆解方式，且將所有和因素有關的分數皆視為真分數。

(四) 階層 ω

McDonald (1999) 提出 ω_t 外，也考慮有階層結構的測驗之信度估計問題。階層結構意指所有題目除有一般因素外，某些題目可能還會受到其他子因素影響。在此模型中，觀測分數對應的因素模型為 $X_i = \mu_i + \lambda_i G + \lambda_{i1} F_1 + \lambda_{i2} F_2 + \dots + \lambda_{ik} F_k + E_i$ ，其中 G 為一般因素，是所有題目都反映的潛在構念， λ_i 為一般因素負荷量， F_1 、 F_2 、 $F_3 \dots F_k$ 為該題對應的子因素，是部分題目才反映的潛在構念， λ_{i1} 、 λ_{i2} 、 $\lambda_{i3} \dots \lambda_{ik}$ 則為該題對應各子因素之因素負荷量。McDonald 僅將一般因素效果視為真分數，故階層 ω (hierarchical-omega)，其簡寫為 ω_h ，可用以下的計算方式：

$$\omega_h = \frac{\text{var}(\sum_{i=1}^p \lambda_i G)}{\sigma_x^2} = \frac{(\sum_{i=1}^p \lambda_i)^2}{\sigma_x^2} \quad \text{公式(6)}$$

從上述可知，多因素時的 ω_h 估計值通常會小於 ω_t 。

過去關於多向度信度估計的研究中，多是使用實徵資料或模擬資料展示各方法的實徵估計表現，結果顯示 α_s 、 ω_t 等方法比 α 係數更適用於多向度測驗組合分數的信度估計（如 Kamata et al., 2003; Margono, 2015; Widhiarso & Ravand, 2014）；有些研究則從數學公式的角度說明各估計法的推導歷程、估計結果的大小關係，例如：證明 β 恆為信度最小值或等於 α （如 Zinbarg et al., 2005）、 ω_t 和 ω_h 於因素模型、信度估計公式之差異展示（如 Green & Yang, 2015; Margono, 2015）。然而，從實務面來看，這些研究的因素數目多不超過三，但許多領域（如人格、臨床）常是三因素以上的測驗，顯然過去研究探討的因素數並不夠；此外，不少研究的模型設定沒有階層結構的情境，也相當缺乏 ω_h 計量特性研究，因此這項估計法以及有階層結構的多因素模型尚有許多可探討的研究設計與議題。

四、多向度測驗信度係數估計與模型轉換

前文指出，不同的真分數計算方式不僅影響估計法的選擇，也代表研究者如何詮釋構念的心理實質性。在實務上，有時研究者會將多向度測驗中的試題分數或分量表分數加總，而得到整體測驗的組合分數與解釋其意義，例如：加總「魏氏成人智力測驗」（WAIS）中各分量表分數，其組合分數代表智力（Wechsler, 2014）。然而，加總所有題目或分量表分數時，須注意加

法的意涵。根據測量理論（measurement theory），加法不只是數字運算，也代表「量」（quantity）之間的關係，例如： $a + b = c$ ，意指 c 這個量是由 a 、 b 兩個量所組成（Michell, 1997）。以前述的智力測驗為例，各分量表分數是指數值大小（magnitude），量表反映的潛在構念是量，加總各分量表分數表示有一個量是由各分量表組成，這個量指的是智力。將此觀點引進測驗的脈絡，加總試題分數代表研究者認為在多向度測驗中，有更高層次或廣泛（general）的因素，並用加總分數表徵此因素；換言之，加總所有試題分數代表研究目的在探討多向度測驗中的一般因素。於此，相關的議題便是如何表徵多向度測驗的因素結構，才能精確連結研究者的理論觀點與信度估計？

通常，多向度測驗之結構為二階／高階因素模型或雙因素模型（McDonald, 1999; Reise, 2012; Revelle & Condon, 2019），以 CES-D 為例，圖 1 為二階因素模型觀點，圖 2 為雙因素模型觀點。

圖 1 的二階因素模型奠基於共同因素模型（common factor model）。McDonald（1999）指出，當測驗結構符合共同因素模型時，可使用 ω_i 估計信度。然而， ω_i 的估計方式為將所有因素效果皆當作真分數，在上述 CES-D 的脈絡中會產生幾個問題。首先，研究者對憂鬱的觀點並非無階層之結構，而是有層次的二階因素模型假設，因此用 ω_i 當作信度估計值即忽略了此階層性。再者， ω_i 並非只納入了憂鬱的效果，而是將所有的共同因素納入計算，此時的信度意義可能是模糊不清的（Green & Yang, 2015）。即便考慮模型的階層問題，其信度估計依然有其困難點，以圖 1 中的任一試題為例，假

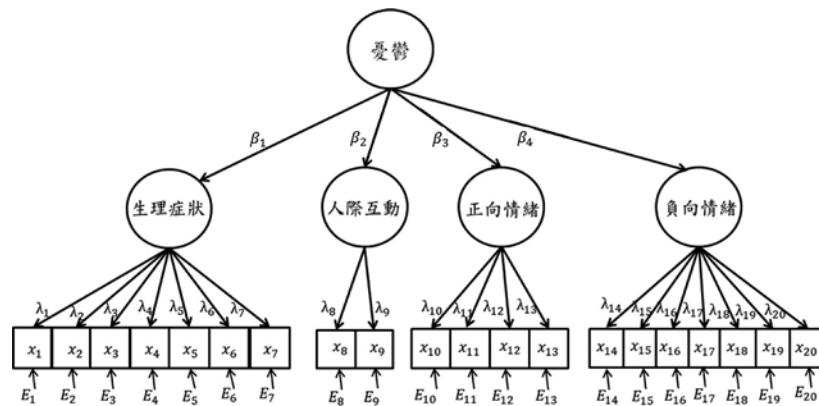
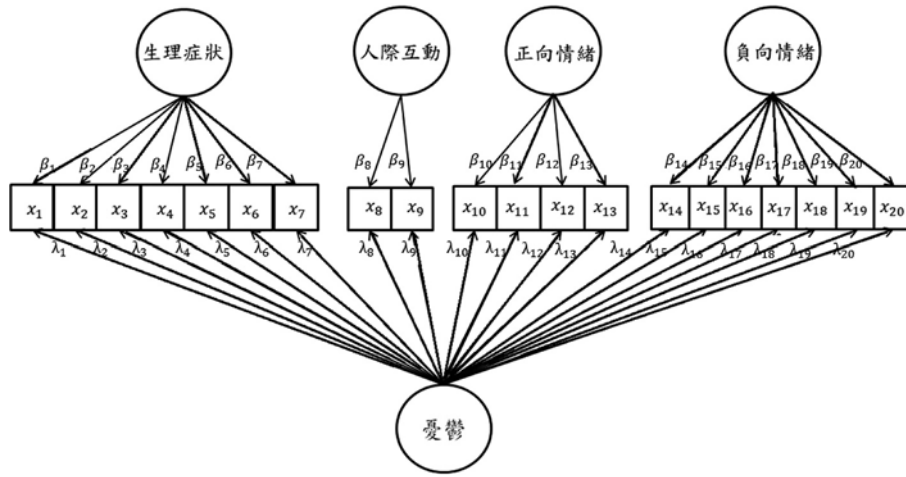


圖 1 CES-D 的二階因素模型示意圖



註：理論上需附上公式中的誤差項，但受限於篇幅影響，故圖中不特意畫出。

圖 2 CES-D 的雙因素模型示意圖

設試題分數為 X ，二階因素為 ζ ，一階因素為 F ，則因素與試題分數間的關係及兩因素間的關係，可分別用公式(7)與公式(8)表達：

$$X = \mu + \lambda F + \varepsilon \quad \text{公式(7)}$$

$$F = \tau + \beta \zeta + \delta \quad \text{公式(8)}$$

在公式(7)中， μ 為截距， λ 為 F 對 X 的因素負荷量， ε 為 X 的測量誤差；在公式(8)中， τ 為截距， β 為 ζ 對 F 的因素負荷量， δ 為 F 的測量誤差。此時，將公式(7)的 F 以公式(8)代入，可得公式(9)，表示 X 與二階因素間的關係：

$$X = \mu + \lambda(\tau + \beta \zeta + \delta) + \varepsilon \quad \text{公式(9)}$$

此時，真分數部分是 $\mu + \lambda(\tau + \beta \zeta)$ ，估計出 $\lambda\beta$ 便能得出信度。然而，實際上難以得到 λ 和 β 各自的估計值，表示無法得知一階與二階因素各自對試題分數的影響力，造成不易解釋此時的信度意義。相較之下，雙因素模型假設一般因素影響所有試題，而有些試題除了受一般因素影響外，也受其他子因素的影響（Green & Yang, 2015; McDonald, 1999; Reise, 2012; Revelle & Condon, 2019），此想法使雙因素模型有幾項二階因素模型不具備的優勢。

首先，雙因素模型的表徵方式較貼近實務中對多向度構念的理解與試題編寫的歷程：題目分數「同時」受到一般因素與多個子因素影響。一般而言，研究者加總所有試題分數之目的在測量一般因素之效果，只加總單項分測驗分數旨在估計子因素的效果，這些分數信度在雙因素模型下都能得到解決。針對一般因素的效果可使用 ω_h 進行估計（McDonald, 1999; Reise, 2012; Revelle & Condon, 2019）；某分測驗之整體分數信度可用 ω_t 估計，而將分測驗之 ω_t 扣除分測驗之 ω_h ，則反映該分測驗的「單純」子因素效果，稱為 ω_s （omega subscale），此時亦能藉由 ω_s 評估子因素於該分測驗的效果，推論試題分數主要反映何種因素（Reise, 2012）。以 CES-D 為例， ω_h 估計的信度值是反映一般因素的分數：憂鬱的效果；生理症狀之 ω_t 是反映生理症狀加憂鬱的效果； ω_s 則反映扣除憂鬱後的「單純」生理症狀效果。上述說明展示了雙因素模型描繪多向度構念的生動性（vividness），以及清楚區分各樣加總分數意涵的優勢，這也是 Reise（2012）建議使用雙因素模型與 ω 系列之係數的原因。

其次，所有二階因素模型皆可轉換成雙因素模型（Yung, Thissen, & McLeod, 1999），其能和許多不同類型的測驗有好的配適度（如 Bados, Gómez-Benito, & Balaguer, 2010; Gignac, Palmer, & Stough, 2007; Patrick, Hicks, Nichol, & Krueger, 2007），此顯示雙因素模型在模型適合度不遜於二階因素模型（Reise, 2012）。再者，從因素分析架構來看，Reise（2012）認為二階因素模型沒有在分數變異來源中區分一般因素與子因素的效果，Revelle 與 Condon（2019）認為多向度測驗之結構可用雙因素模型表徵，表示雙因素模型涵蓋了過去常用的共同因素模型，說明雙因素模型除了概念表徵外，在數理模型上亦有優勢。

綜合多向度測驗構念表徵、信度意義與因素模型等議題的論述，本研究認為雙因素模型在心理學研究中於探討多向度測驗更具有廣泛性與優勢，故本研究以雙因素模型作為真實模型產生模擬資料，評估不同信度估計法之相對表現。此相對表現包含兩層面：首先，研究者想了解在不同的雙因素模型下，各信度估計法之真值（true value）與 ω_h 真值之差異；其次，研究者想了解在有限樣本下，這些估計法之估計品質為何，包括偏誤（bias）與穩定性（stability）之表現。除了模擬研究外，本研究於下節使用一筆 CES-D 之實徵資料進行分析，展示本研究提及的各個多向度信度係數之實徵表現。

五、實徵範例

本研究使用 R 語言 (R Core Team, 2019) 中的 ds4psy 套件 (Neth, 2020) 中名為 posPsy_participants 的資料 (Woodworth, O'Brien-Malone, Diamond, & Schüz, 2018) 當作實徵範例。此資料來自 Woodworth、O'Brien-Malone、Diamond 與 Schüz (2017) 進行的正向心理學治療 (positive psychology interventions) 效果之調查研究，有 992 位研究參與者，資料中的 29 至 48 列 (column) 為研究參與者的 CES-D 作答反應，共 20 題。依據 Longmire 與 Knight (2010) 的研究，CES-D 可分為負向情緒 (depressed affect)、生理症狀 (somatic and retarded activity)、正向情緒 (positive affect)、人際互動 (interpersonal feelings) 等四個因素，故本研究使用這四個因素做為 CES-D 資料的因素結構。而在雙因素模型下，CES-D 中的四個因素被視為子因素，一般因素為憂鬱，因此本研究將 CES-D 對應的雙因素模型設定為一個一般因素加上四個子因素，接著使用 psych 套件 (Revelle, 2019) 的 omega 函數、sirt 套件 (Robitzsch, 2020) 的 stratified.cronbach.alpha、splitHalf 函數，依序估計 ω_t 、 ω_h 、 α_s 、 α 、 β 五個信度係數，結果如表 1 所示。

表 1 CES-D 實徵資料分析結果

量表加總分數之信度		
$\alpha = .94$, $\alpha_s = .86$, $\omega_t = .95$, $\omega_h = .83$, $\beta = .89$		
子因素	試題編號	分量表分數之 α
負向情緒	3, 6, 9, 10, 14, 17, 18	.90
生理症狀	1, 2, 5, 7, 11, 13, 20	.84
正向情緒	4, 8, 12, 16	.86
人際互動	15, 19	.69

從分析結果可知，由於最差折半信度 β 為 .89 已屬於高信度值，代表 CES-D 的內部一致性高，這也反應在以加總分數為觀點的 α 、 ω_t 指標，兩者的估計值分別為 .94、.95，顯示 CES-D 確實有高內部一致性。在分數解釋上，研究者加總 CES-D 分數之目的在捕捉一般因素效果的實際效果，若使用 α 、 ω_t 或 α_s 解釋，研究者看到的數值是混合憂鬱、負向情緒、生理症狀、正向情緒、人際互動等所有因素之效果，而非評估憂鬱的「單純」效果，在憂鬱的概念解釋上並不合理。若使用雙因素模型的觀點搭配 ω_h ，例如：在本例

中， ω_h 估計值為 .83，表示測驗分數的總變異有 83% 可被「單純」的憂鬱因素解釋；若是將 ω_t 扣除 ω_h 則會得到 .12，代表在排除憂鬱的效果下，負向情緒、生理症狀、正向情緒、人際互動的總和效果僅能解釋 12% 的分數變異量，此顯示 ω_h 能更準確的評估憂鬱效果，對憂鬱的概念解釋較合理，說明使用雙因素模型搭配 ω_h 在理解多向度構念上更為合理，且讓多向度測驗分數的解釋更為完整、全面。

貳、研究方法

一、操弄變項

從文獻回顧可知，子因數數目、樣本數、不同的一般因素負荷量等變項會影響信度估計之結果，故本研究操弄的四個獨變項與水準數如下：

1. 子因素數目：數目涵蓋三至六個子因素，此設計不僅包含 Green 與 Yang (2009a) 的研究水準數，還拓展至五或六個子因素，原因在於實務測驗（如魏氏智力測驗）為四因素以上的測驗（Wechsler, 2014），因此研究四個子因素以上的情境有實務上之必要。另外，考慮因素分析中的模型辨證（identification）與收斂問題，本研究採用 Bollen (1989) 的建議，設定每一子因素有四題。

2. 不同的樣本數：此變項參考 Kamata 等人 (2003) 以及 Kelley 與 Pornprasertmanit (2016) 使用的數值，有 50、100、200、400、600、800，這些樣本數除了為模擬研究常用的數值外，小樣本數（如 50、100）是考慮有些臨床測驗的樣本人數少，因此模擬小樣本的情境有其實務之重要性；而較大樣本數則是考慮有些實務測驗或調查，例如：世界價值觀調查（world value survey）的樣本數通常為大樣本。同時，希望透過較為連續的數值設計看出各方法的估計表現趨勢。

3. 不同的子因素負荷量：此變項參考 Green 與 Yang (2009a) 的研究，原為 .2 與 .6 兩種水準，但考慮子因素負荷量於實徵資料中多未達 .5 (Reise, Morizot, & Hays, 2007)，因此本研究將 .6 調為 .4。簡言之，子因素負荷量有 .2、.4 兩種水準。

4. 不同的一般因素負荷量：此變項在 Green 與 Yang (2009a) 的研究原為 .3、.8 兩種數值，但考慮一般因素負荷量於實徵資料中多在 .5 以上 (Reise et

al., 2007)，因此本研究將 .3 調為 .5，故一般因素負荷量有 .5、.8 兩種水準。

在評估各信度估計法的表現時，本研究主要參考：(1)各信度指標之真值和 ω_h 真值的差異；(2)各信度指標的相對偏誤（relative bias）百分比，其計算公式為 $\frac{\bar{\hat{\theta}} - \theta}{\theta} \times 100\%$ ，其中的 $\bar{\hat{\theta}}$ 為參數估計值之平均數， θ 則為各指標之母群體數值；(3)以各信度指標的標準差來評估信度估計法的穩定性，其計算公式為 $\sqrt{\frac{\sum_{i=1}^N (\hat{\theta}_i - \bar{\hat{\theta}})^2}{N-1}}$ ，其中的 N 為模擬次數， $\hat{\theta}_i$ 為每次模擬的參數估計值， $\bar{\hat{\theta}}$ 為各指標估計值之整體平均數。

二、模擬流程

本研究根據各情境的雙因素模型，先使用統計軟體 R 語言中 lavaan 套件（package）（Rosseel, 2012）的 simulateData 函數產生連續觀察資料，例如：在三個子因素的情境下，本研究設定所有因素之間皆為正交（orthogonal），即所有因素之間沒有相關；所有子因素負荷量相等，如皆為 .2 或皆為 .5，所有試題的誤差分數皆無相關。然後，據此設定搭配 .5 或 .8 的一般因素負荷量，完成該情境的雙模型設定，接著由設定好的模型使用 simulateData 函數產生資料¹，再計算各情境中各信度指標的真值。

在信度估計方面，本研究使用 α_s 、 β 、 ω_t 、 ω_h 、 α 五個信度指標，每一情境重複模擬（replicate）1,000 次，再計算各信度估計值之平均數與標準差。在各信度指標之計算方面， α_s 是使用 sirt 套件的 stratified.cronbach.alpha 函數計算， β 是使用 psych 套件的 splitHalf 函數計算，其得出的最小折半信度估計值即為 β ， ω_t 使用 psych 套件的 omega 函數計算，因素萃取法（factor extraction method）設為主軸因子法（principal axes），轉軸法（rotation）設為斜交轉軸（oblimin），搭配使用 Schmid-Leiman 方法（Reise, Moore, & Haviland, 2010; Schmid & Leiman, 1957），即得出有階層結構的因素模型解。理論上，Schmid-Leiman 方法是將原先的斜交因素模型轉成有二階（second-order）因素的模型結構，並且假設一階因素和二階因素都為簡單結構（simple structure）。接著利用前述的公式(6)至公式(8)的轉換過程，將觀察變項以二階因素的型式表徵，是為公式(8)，再將公式(8)轉換成以下形式：

1 模擬研究產生資料之程式碼，可向第一作者寄信索取。

$$X = Bh + \Theta u_0 \quad \text{公式(9)}$$

其中，公式(9)中的 h 為二階因素之轉置矩陣和一階誤差項轉置矩陣的聯合矩陣， B 為經過 SL 法轉換後的因素負荷量矩陣， Θ 為誤差之因素負荷量矩陣， u_0 為誤差項之向量。此時，可將觀察變項對應的相關矩陣（correlation matrix） Σ 改寫為：

$$\Sigma = BB^T + \Theta^2 \quad \text{公式(10)}$$

其中， B 矩陣同時擁有一般因素負荷量與子因素負荷量，進而得出雙因素模型的參數解。

參、結果

一、各方法之真值比較

表 2 呈現不同情境下各估計法的真值，其中的 g 為一般因素負荷量、 s 為子因素負荷量、 m 為子因素數目。從表 2 可看出，當一般因素負荷量增加或子因素數目增加時，所有估計法的真值皆會提高；而當子因素負荷量增加時，僅 ω_h 和 β 之外的方法才會上升。在子因素負荷量為 .2 時， β 真值和 ω_h 真值會相當一致或差距不超過 .01，顯示在低子因素負荷量的情況下， β 和 ω_h 的估計結果會十分接近；其他方法之真值則在子因素負荷量為 .2 和一般因素負荷量為 .8 的情況下，才會和 ω_h 真值較接近。此結果顯示， ω_h 較其他方法更能捕捉到一般因素的效果。

二、信度估計之相對偏誤百分比與標準差

圖 3 呈現各方法信度估計之相對偏誤百分比。首先，研究者觀察到 α 與 α_s 幾乎沒有展現偏誤，多在 2% 以內；其次， ω_r 僅在樣本數為 200 以下時，才會展現偏誤，但仍屬於多數學者可接受的偏誤不到 10% 之範圍（Green & Yang, 2009a）；最後， ω_h 和 β 展現的相對偏誤最大， ω_h 於小樣本情境（如 50、100）等狀況時會有大於 10% 的相對偏誤，甚至可超過 30%， β 的狀況好一些，最大相對偏誤為 14.63%，其餘皆在 10% 以內。隨著樣本數增加，一般因素負荷量提升，或子因素負荷量提升， ω_h 和 β 的相對偏誤百分比皆呈現縮小的趨勢。

表 2 不同情境下各估計法的真值

		$g = .5$ $m = 3$	$g = .5$ $m = 4$	$g = .5$ $m = 5$	$g = .5$ $m = 6$	$g = .8$ $m = 3$	$g = .8$ $m = 4$	$g = .8$ $m = 5$	$g = .8$ $m = 6$
$s = .2$	α	.81	.85	.87	.89	.96	.97	.97	.98
	β	.79	.82	.86	.88	.95	.96	.97	.97
	ω_h	.78	.82	.85	.87	.94	.96	.96	.97
	ω_t	.82	.85	.88	.90	.96	.97	.98	.98
	α_s	.82	.85	.88	.90	.96	.97	.98	.98
$s = .4$	α	.83	.86	.88	.90	.97	.97	.98	.98
	β	.76	.76	.82	.85	.93	.92	.95	.96
	ω_h	.71	.72	.80	.83	.90	.90	.94	.95
	ω_t	.86	.89	.91	.92	.98	.98	.99	.99
	α_s	.86	.89	.91	.92	.98	.98	.99	.99

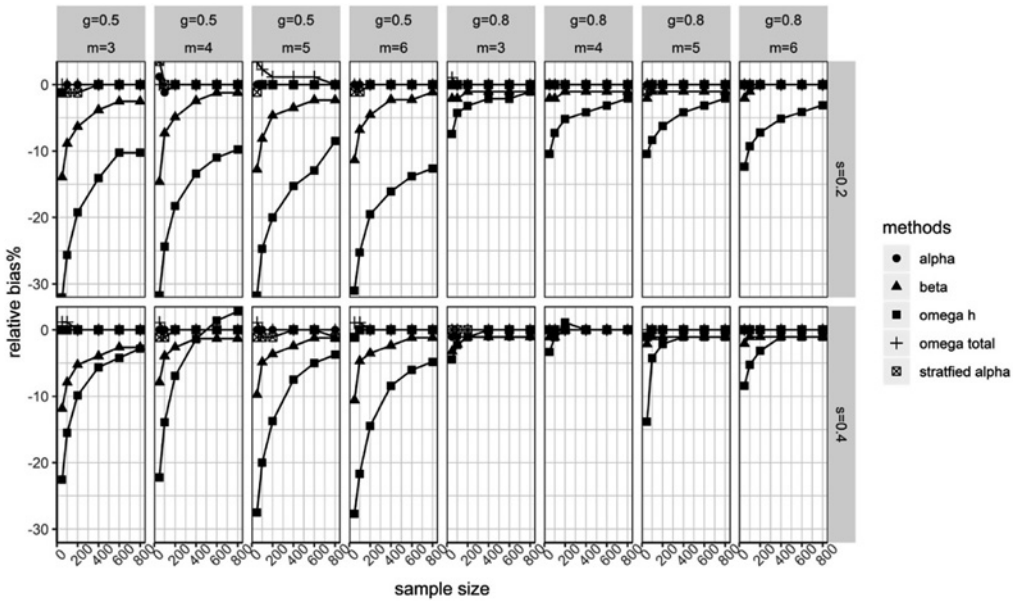


圖 3 各樣本數下各方法的信度估計相對偏誤百分比

圖 4 呈現各方法信度估計之標準差，結果顯示隨著樣本數增加，一般因素負荷量以及子因素負荷量的提升，各方法的標準差有縮小的趨勢。而子因素數目的增加，僅在一般因素負荷量為 .5 時會使各方法估計標準差較明顯縮小，其他情境下的子因素數目並無造成各方法的估計標準差有明顯變化。最後，研究者再次觀察到 ω_h 和 β 的估計標準差較其他指標來得大。

從上述結果可知，儘管在概念上， ω_h 乃最為合適之信度估計指標，但其在估計偏誤與穩定性之表現皆不理想，而真值與 ω_h 較為接近的 β ，亦同樣展現相對差的表現，兩指標皆僅在較大樣本數、高一般因素負荷量，以及高子因素負荷量時，才有較好的估計品質。

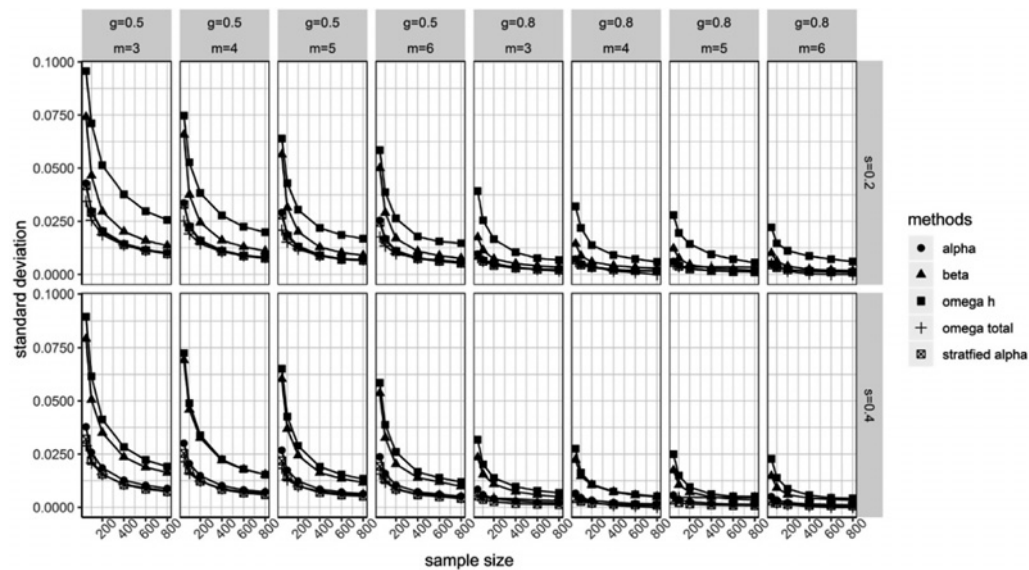


圖 4 各樣本數下各方法的信度估計標準差

肆、結論

一、真值、相對偏誤與標準差

整體而言，各指標之間真值的比較，以 α_s 、 ω_t 和 ω_h 彼此間的真值差異最大， α 係數和 ω_h 間的真值差異次之， β 和 ω_h 間的真值差異則最小；但在低子因素負荷量、高一般因素負荷量的情境下，各方法的真值較為接近。在多數

情境中， ω_t 和 α_s 的真值相等，顯示兩者皆適用於多向度結構測驗的信度估計（Cronbach et al., 1965; Green & Yang, 2009a; Margono, 2015; Sijtsma, 2009），也表示在滿足 α_s 的假設時，古典測驗理論取向與模型取向的估計結果會一致（Kamata et al., 2003; Widhiarso & Ravand, 2014）。

在一般因素負荷量較高的情境下，各方法皆有較小的相對偏誤與較穩定的估計表現，顯示清晰的構念或高效度的測驗較能被精確的測量，這和過去測驗理論中的信度觀點一致（如 Murphy & Davidshofer, 1988）。當子因素負荷量低時，所有方法的估計相對偏誤較大；而當子因素數目多時， ω_h 和 β 的估計相對偏誤較大，但對其他方法的估計相對偏誤則無明顯影響。

在小樣本的情境下， α 、 α_s 、 ω_t 的估計穩定性相對佳，而 β 與 ω_h 的估計穩定性相對差。由於 w_h 屬於模型取向的方法，因此需要大樣本才有較好的估計表現，這和過去理論與研究一致（如 Kelley & Pornprasertmanit, 2016）。另外，本研究意外發現 ω_t 於小樣本有不錯的估計表現。而 β 屬於古典測驗理論取向的方法，從過去研究（如 Cronbach, 1951; Cronbach et al., 1965）會推測其估計表現應較不受樣本數影響，然本研究結果顯示小樣本情境下 β 的估計表現稍顯不穩定，可能原因在於本研究使用的估計函數設定是隨機挑選測驗的折半組合，從中選取最小的折半信度值作為 β ，因而造成此指標表現稍不穩定，但概括而言， β 仍和古典測驗理論取向方法的估計標準差之變化趨勢一致。而樣本數較大（如 600 以上）時，其模型取向的方法（如 ω_h 和 ω_t ）皆呈現較好的估計表現，這和過去的研究結果一致（如 Kelley & Pornprasertmanit, 2016）。

從上述討論各方法和 ω_h 的真值差異與估計表現情況後，研究者建議在面臨一般因素負荷量低、子因素數目少且因素負荷量高的情況時，只用 ω_h 作為單一信度指標並不合適，還需要參考其他方法並檢視量表結構是否合理。當一般因素負荷量小、子因素數目多且因素負荷量低時，雖然其他方法的真值和 ω_h 真值之差異不超過 0.1，但只單看 ω_h 真值亦不合適，因為此時研究者所關心的主要構念與子構念都未被充分測量，可能需要審思構念的理論詮釋或搭配其他信度指標，再以不同觀點探討測量結果。當一般因素負荷量高、子因素數目少且因素負荷量低時，此時只用 ω_h 做為單一指標於理論與邏輯上並無重大瑕疵，然考慮到其他方法的真值和 ω_h 真值也縮小，仍建議參考 Reise (2012) 增加其他指標輔以不同觀點的信度評估，使測驗的分數解釋較為全面。若是一般因素負荷量高、子因素數目多且因素負荷量高的情

境，雖然各方法的真值和 ω_h 真值之差異不大，且大樣本下估計表現相似，但此情境中的子構念影響力不容忽視，會建議除了 ω_h 外還需參考其他多向度信度指標，並評估是否測量過多的構念或需調整測驗結構。

整體而言，所有估計法的真值於本研究中呈現的結果與過去的理論或研究結果一致，除驗證本研究結果的正確性與可靠度，也驗證過去的理論和研究結果應是正確與穩定。

二、綜合討論與研究限制

本研究探討了 α 、 α_s 、 β 、 ω_t 、 ω_h 五種信度估計法，除了 α 、 β 係數外，其餘為專用於多向度測驗結果的信度估計法。本研究結果除可作為多向度測驗結果之信度估計法選擇的參考外，也部分彌補先前較少將 β 和 ω_h 同時放入研究探討的缺陷。

本研究旨在探討各方法在雙因素模型下於不同樣本數、不同子因素數目、不同一般因素負荷量與子因素負荷量下的估計偏誤與穩定性。結果發現，古典測驗理論取向的估計法（如 α 、 α_s 、 β ）都有不錯的估計表現，和 ω_h 的差異多數情況下不會超過 0.15；而模型取向的估計法（如 ω_t ）於各情境下有不錯的估計表現，其表現趨勢和 α_s 相似，皆可用於估計多向度測驗中所有因素效果混合後的「整體信度」。而 ω_h 在實徵意義上是最為清晰與合理的估計法，但在本研究設計的某些情境中卻穩定性最低，這是否為該方法的特性之一，還需後續研究。然就本研究的觀點會建議，在估計多向度測驗結果的信度時，仍應以 ω_h 為主要參照的指標，只是需考慮前文中提及的幾種常見之一般因素負荷量、子因素數目與因素負荷量的組合情況，評估是否需加入其他估計法一同作為信度指標。然而，本研究有幾項限制，在推論研究結果時宜多加留意，以下分段說明。

1. 本研究之主要目的是探討探索性（exploratory）研究取向下的信度估計表現。一般而言，在發展測驗工具前期的研究取向多為探索性研究取向，研究者在這個階段可能還沒有清晰、完整的理論架構，需要透過蒐集資料輔助理論的建構。因此，本研究希望聚焦於探索性研究取向下的多向度信度係數之數理意涵與估計表現，提供研究者概念上較為正確的信度使用方式和各情境下的估計表現。但不可否認的，當社會科學某些理論或測驗工具發展至較成熟、完整的階段時，即需要使用驗證性（confirmatory）研究取向檢視理論或測驗結構的正確性，這是本研究未特別聚焦的部分，未來可進一步探討

在驗證性研究取向下的不同因素結構之的信度係數估計表現，讓多向度信度的研究更為完整。

2.本研究發現因素分析的轉軸方法會影響 ω_h 的估計表現。本研究使用 R 語言中 psych 套件裡 omega 函數中的斜交轉軸法，能得出與理論相符的真值，但 ω_h 估計偏誤較大；若使用 omega 函數中的 promax 轉軸法則能得出較小之 ω_h 估計偏誤，但所得出的真值與理論並不相符。因此，會建議研究者或實務工作者使用 omega 函數估計 ω_h 時需留意選取的轉軸法。從轉軸議題可知，儘管本研究中 ω_h 在理論探討方面認為此指標意義清晰，但探索性研究時估計表現不佳則是研究限制之一。

3.本研究的雙因素模型設定是因素間無相關且無誤差間相關之條件，因此本研究結果僅限於雙因素模型且所有因素皆為正交、誤差間無相關的情境。或許有研究者建議可探討子因素間有相關或因素間為斜交的情境議題，然而有學者（如 Reise, 2012）並不認同子因素間有相關或因素間斜交的模型合理性，原因是當子因素間有相關時，會讓一般因素在概念上難以解釋，本研究抑是相同觀點。

參考文獻

- óBados, A., Gómez-Benito, J., & Balaguer, G. (2010). The state-trait anxiety inventory, trait version: Does it really measure anxiety? *Journal of Personality Assessment*, 92(6), 560-567. doi:10.1080/00223891.2010.513295
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16(3), 297-334. doi:10.1007/BF02310555
- Cronbach, L. J., Schönemann, P., & McKie, D. (1965). Alpha coefficients for stratified-parallel tests. *Educational and Psychological Measurement*, 25(2), 291-312. doi:10.1177/001316446502500201
- Drolet, A. L., & Morrison, D. G. (2001). Do we really need multiple-item measures in service research? *Journal of Service Research*, 3(3), 196-204. doi:10.1177/109467050133001
- Gignac, G. E., Palmer, B. R., & Stough, C. (2007). A confirmatory factor analytic investigation of the tas-20: Corroboration of a five-factor model and suggestions for improvement. *Journal of Personality Assessment*, 89(3), 247-257. doi:10.1080/00223890701629730
- Green, S. B., & Yang, Y. (2009a). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74(1), 121-135. doi:10.1007/s11336-008-9098-4
- Green, S. B., & Yang, Y. (2009b). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, 74(1), 155-167. doi:10.1007/s11336-008-9099-3
- Green, S. B., & Yang, Y. (2015). Evaluation of dimensionality in the assessment of internal consistency reliability: Coefficient alpha and omega coefficients. *Educational Measurement: Issues and Practice*, 34(4), 14-20. doi:10.1111/emip.12100
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10(4), 255-282. doi:10.1007/BF02288892
- Hogan, T. P., Benjamin, A., & Brezinski, K. L. (2000). Reliability methods: A note on the frequency of use of various types. *Educational and Psychological Measurement*, 60(4), 523-531. doi:10.1177/00131640021970691
- Kamata, A., Turhan, A., & Darandari, E. (2003, April). *Estimating reliability for multidimensional composite scale scores*. Paper presented at the annual meeting of American

- Educational Research Association, Chicago, IL.
- Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods*, 21(1), 69-92. doi:10.1037/a0040086
- Longmire, C. V. F., & Knight, B. G. (2010). Confirmatory factor analysis of the Center for Epidemiologic Studies-Depression Scale in black and white dementia caregivers. *Ageing and Mental Health*, 14(8), 962-970. doi:10.1080/13607863.2010.501060
- Lord, F. M., & Novick, M. (1968). *Statistical theories of mental test scores*. New York, NY: Addison-Westley.
- Margono, G. (2015). Multidimensional reliability of instrument for measuring students' attitudes toward statistics by using semantic differential scale. *American Journal of Educational Research*, 3(1), 49-53. doi:10.12691/education-3-1-10
- McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, 23(1), 1-21. doi:10.1111/j.2044-8317.1970.tb00432.x
- 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: Lawrence Erlbaum Associates.
- Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88(3), 355-383. doi:10.1111/j.2044-8295.1997.tb02641.x
- Murphy, K. R., & Davidshofer, C. O. (1988). *Psychological testing: Principles, and applications*. Englewood Cliffs, NJ: Prentice-Hall.
- Neth, H. (2020). *Data science for psychologists*. Retrieved from <https://bookdown.org/hneth/ds4psy/>
- Patrick, C. J., Hicks, B. M., Nichol, P. E., & Krueger, R. F. (2007). A bifactor to modeling the structure of the psychopathy checklist-revised. *Journal of Personality Disorders*, 21(2), 118-141. doi:10.1521/pedi.2007.21.2.118
- R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing (<https://www.R-project.org/>).
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667-696. doi:10.1080/00273171.2012.715555
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544-559. doi:10.1080/00223891.2010.496477

- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, 16(1), 19-31. doi:10.1007/s11136-007-9183-7
- Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate Behavioral Research*, 14(1), 57-74. doi:10.1207/s15327906mbr1401_4
- Revelle, W. (2019). *Psych: Procedures for personality and psychological research*. Evanston, IL: Northwestern University. Retrieved from <https://CRAN.R-project.org/package=psych>
- Revelle, W., & Condon, D. M. (2019). Reliability from α to ω : A tutorial. *Psychological Assessment*, 31(12), 1395-1411. doi:10.1037/pas0000754
- Robitzsch, A. (2020). *Sirt: Supplementary item response theory models*. R package version 3.9-4. Retrieved from <https://CRAN.R-project.org/package=sirt>
- Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. doi:10.18637/jss.v048.i02
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 53-61. doi:10.1007/BF02289209
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120. doi:10.1007/s11336-008-9101-0
- Wechsler, D. (2014). *Wechsler Adult Intelligence Scale, fourth edition (WAIS-IV)*. TX: Pearson.
- Widhiarso, W., & Ravand, H. (2014). Estimating reliability coefficient for multidimensional measures: A pedagogical illustration. *Review of Psychology*, 21(2), 111-121.
- Williams, C. D., Taylor, T. R., Makambi, K., Harrell, J., Palmer, J. R., Rosenberg, L., & Adams-Campbell, L. L. (2007). CES-D four-factor structure is confirmed, but not invariant, in a large cohort of African American women. *Psychiatry Research*, 150(2), 173-180. doi:10.1016/j.psychres.2006.02.007
- Woodworth, R. J., O'Brien-Malone, A., Diamond, M. R., & Schüz, B. (2017). Web-based positive psychology interventions: A reexamination of effectiveness. *Journal of Clinical Psychology*, 73(3), 218-232. doi:10.1002/jclp.22328
- Woodworth, R. J., O'Brien-Malone, A., Diamond, M. R., & Schüz, B. (2018). Data from, 'Web-based Positive Psychology Interventions: A Reexamination of Effectiveness'. *Journal of Open Psychology Data*, 6(1). doi:10.5334/jopd.35
- Yang, Y., & Green, S. B. (2010). A note on structural equation modeling estimates of reliability. *Structural Equation Modeling*, 17(1), 66-81. doi:10.1080/10705510903438963
- Yung, Y. F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-

order factor model and the hierarchical factor model. *Psychometrika*, 64(2), 113-128. doi:10.1007/BF02294531

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α , Revelle's β , and McDonald's ω h: Their relations with each other and two alternative of conceptualizations reliability. *Psychometrika*, 70(1), 123-133. doi:10.1007/s11336-003-0974-7

