

「臺灣簡明版世界衛生組織生活品質問卷」 之信度概化研究

林立中¹ 蘇意晴² 姚開屏³

摘要

與「簡明版世界衛生組織生活品質問卷」(WHOQOL-BREF)相同,「臺灣簡明版世界衛生組織生活品質問卷」已廣泛應用至國內各研究領域。雖分數信度為效度係數、效果量等統計量的基礎,然文獻上缺乏該量表之分數信度的研究。信度概化法(reliability generalization, RG)採用後設分析方法,以探討某一測量工具之跨研究分數信度的心理計量特徵,及其分數信度與研究特徵的關係,為檢驗分數信度的重要工具。因此,本研究採用 RG 以探討「臺灣版 WHOQOL-BREF」之四範疇分數信度的心理計量表現。本研究從「臺灣博碩士論文知識加值系統」與「華藝線上圖書館」兩個資料庫,以「WHOQOL」或「世界衛生組織生活品質」作為關鍵字蒐集 2000 年至 2015 年 8 月使用「臺灣版 WHOQOL-BREF」的博碩士論文,刪除不符合研究目的之論文後,共納入 248 篇。除分析其分數信度的描述統計,並採用多重方法如迴歸分析、相關分析、*t* 考驗、變異數分析,以檢驗分數信度與研究特徵的關係。研究結果顯示,四範疇之跨研究的平均分數信度高於 .70,且其標準差低於 .09,顯示此四範疇的分數信度具跨研究穩定性。四範疇之量表分數標準差可顯著預測該範疇之分數信度,且部分範疇之分數信度受教育程度、宗教信仰、樣本年齡的影響。建議未來在評估是否使用「臺灣版 WHOQOL-BREF」時,需考量這些研究變項,尤其是在生理健康與環境範疇上。

關鍵詞：信度概化、臺灣版 WHOQOL-BREF

-
1. 林立中,國立臺灣大學心理學系博士生;康橋國際學校林口校區行政專員
 2. 蘇意晴,國立臺灣大學心理學系碩士;歐立威科技股份有限公司資料分析師
 3. 姚開屏,國立臺灣大學心理學系教授
- 收件日期:2019.12.08;完成修改:2020.06.18;正式接受:2020.06.24
通訊作者:姚開屏;Email:kaiping@ntu.edu.tw
地址:臺北市大安區羅斯福路四段1號 國立臺灣大學心理學系

Reliability Generalization Study on the Taiwan Version of the WHOQOL-BREF

Li-Chung Lin¹ Yi-Ching Su² Grace Yao³

Abstract

As the short version of the World Health Organization Quality of Life questionnaire (WHOQOL-BREF), Taiwan version of WHOQOL-BREF is widely used in various research fields. Although the score reliability is the foundation of statistics, such as validity coefficient and effect size, the study about score reliability of WHOQOL-BREF Taiwan version is rare. By using meta-analysis method, reliability generalization (RG) is a powerful approach for examining the psychometric properties of score reliability and assessing the study characteristics associated with score reliability. The present study therefore utilized RG to examine psychometric performance of the score reliability from four domains of the Taiwan version of the WHOQOL-BREF. This study used the keywords, WHOQOL and World Health Organization Quality of Life, to collect doctoral and master dissertations (from 1998 to the May 2019) which used the WHOQOL-BREF Taiwan version from two databases: National Digital Library of Theses and Dissertations in Taiwan and Airiti Library. After deleted dissertations which were incompatible with the purpose of this study, 248 dissertations were included. In addition to analyzing descriptive statistics of score reliability from these dissertations, variety of statistic methods (such as regression analysis, correlation analysis, *t*-test, and analysis of variance) were conducted to examine the relationship between score reliability and study characteristics. The findings showed that the mean score reliability of four domains of WHOQOL-BREF Taiwan version were higher than .70 and their standard deviations were lower than .09, indicating that WHOQOL-BREF Taiwan revealed consistency across studies. Moreover, the standard deviations of scores from these four domains could significantly predict their score reliability, as well as parts of their score reliability were influenced by study characteristics such as education level, religious beliefs, and age of study participants. We suggested that when assessing whether to use WHOQOL-BREF Taiwan version or not, researchers need to consider the influence of these study characteristics, especially for physical health and environmental domains.

Keywords: reliability generalization, Taiwan version of WHOQOL-BREF

1. Li-Chung Lin, Doctoral student, Department of Psychology, National Taiwan University; Administration staff, Kang Chiao International School Linkou Campus

2. Yi-Ching Su, Master's Degree, Department of Psychology, National Taiwan University; Data Analyst, Omniwaresoft Technology Inc.

3. Grace Yao, Professor, Department of Psychology, National Taiwan University

Received: 2019.12.08; Revised: 2020.06.18; Accepted: 2020.06.24

Corresponding Author: Grace Yao; Email: kaiping@ntu.edu.tw

Address: No. 1, Sec. 4, Roosevelt Rd., Da'an Dist., Taipei City 106, Taiwan

Department of Psychology, National Taiwan University

壹、緒論

近 30 年來，生活品質（quality of life, QOL）已成為心理、醫藥和臨床研究的重要變項（Chen, Wu, & Yao, 2006; Trompenaars, Masthoff, Van Heck, Hodi-amont, & Vries, 2005）。為能夠更適切且精準地測量 QOL，世界衛生組織（World Health Organization, WHO）同時考量病患的生活品質、藥物與醫療政策，以及生活品質之跨文化性等重要議題，而發展出國際生活品質測量工具，定名為「世界衛生組織生活品質問卷」（WHOQOL-100）。隨後，WHO 精簡其題目而發展出「簡明版世界衛生組織生活品質問卷」（WHO-QOL-BREF），以適用於實徵臨床研究與流行病學調查。WHOQOL-BREF 為一跨文化評估個人生活品質的普同測量工具（The WHOQOL Group, 1995, 1998），目前已發展 50 餘國的版本，為研究者最熟悉的測量生活品質工具之一（Nedjat, Monazeri, Holakouie, Mohammad, & Majdzadeh, 2008）。「臺灣簡明版世界衛生組織生活品質問卷」（簡稱「臺灣版 WHOQOL-BREF」）於 2002 年由 Yao、Chung、Yu 與 Wang 翻譯而成，其使用率逐年增加，至 2020 年 4 月底申請使用之研究件數已達 1,700 件，使用單位遍布各級學校、醫院診所與公司行號等，涵蓋護理相關、醫院各科、精神科與公共衛生等研究部門，並用以評估一般成人、學生、病患等不同族群的生活品質，其應用範圍實已超出原先發展「臺灣版 WHOQOL-BREF」時所依據的樣本組成，造成實徵應用者無法確認奠基於該樣本的心理計量特徵是否能夠類推至其他樣本類型。由於信度係數為效度係數與其他統計量的基礎，如低信度會降低統計考驗力與效果量的估計，因此本研究主要著重在探討 WHOQOL-BREF 的信度計量特徵。

信度意指某一測量工具之測量結果的一致性，是一組測量分數的特徵，而非測量工具本身。因此，測驗分數的信度（本文統稱為分數信度）較測驗信度對於信度概念之描述更為恰當（Gronlund & Linn, 1990）。分數信度對心理計量研究與實徵研究至關重要（Graham & Christiansen, 2009），低信度不僅會降低統計量（如相關係數、迴歸係數），亦會低估變項間的實際關係（如因高測量誤差而低估變項間的關係）。再者，由於分數信度為樣本和測量工具的函數，不同的樣本規模或特徵亦會影響其估計值。據此，Vacha-Hase 與 Thompson（2011）指出，在進行量化研究的資料分析之前，應先報告

該研究所使用之測量工具的分數信度。相似的觀點亦出現在《教育與心理測量》期刊 (*Educational and Psychological Measurement*)、美國教育研究協會 (American Educational Research Association)、美國心理學協會 (American Psychological Association, APA)，以及 APA 統計推論小組 (APA Task Force on Statistical Inference) 的編輯政策上。因此，選擇一個具有良好分數信度的測量工具為量化研究的首要任務。藉由探討測量工具之分數信度的心理計量基礎，例如：某一量表之分數信度在不同研究中的表現（如平均表現和分散情形），以及檢驗研究特徵（如樣本大小、性別、地區等）與分數信度的關係，有助於實徵研究者選擇一個具有良好信度的測量工具，以評估某一測量工具是否適用於該研究，故跨研究之分數信度分析顯得非常重要。

在評估測量工具之分數信度的概化程度上，Cronbach、Gleser、Nanda 與 Rajaratnam (1972) 首先提出概化理論 (generalizability theory, GT)，以檢驗某一測量工具的可靠性 (dependability)。GT 採用變異數分析估計測量工具之多重測量誤差來源的大小，例如：測量情境 (occasion)、評分者 (rater)、題目 (item) 等，並藉由確認這些可能的潛在測量誤差來源，用以評估測量工具的概化普同性 (universe of generalizability) (Shavelson, Webb, & Rowley, 1989)。隨後，Vacha-Haase (1998) 提出信度概化 (reliability generalization, RG)，藉由對某一測量工具之分數信度的後設分析，以統整、歸納並檢驗該測量工具之跨研究分數信度係數的心理計量特徵。透過 RG 可獲得分數信度之跨研究的平均數與變異數，並藉由檢驗研究特徵與分數信度的關係，以探討哪些研究特徵（如性別、年齡）會影響分數信度的變異性，據此可提供選擇一份良好測量工具的訊息基礎 (Botella, Suero, & Gambara, 2010)，是檢驗測量工具之分數信度的重要方法。自 Vacha-Haase 提出 RG 後，其方法學與實徵研究逐漸增加，並已廣泛地應用至多種測量工具 (Vacha-Haase & Thompson, 2011)，例如：「狀態—特質焦慮量表」 (State-Trait Anxiety)、「內外控量表」 (Locus of Control) 等。然根據研究者的文獻回顧結果，目前在臺灣並未有任何量表的 RG 研究出版，更遑論「臺灣版 WHOQOL-BREF」的 RG 研究，使得研究者對於該量表之分數信度的心理計量特徵，以及哪些研究特徵會影響其分數信度的表現並不清楚，可能導致在實徵應用上，研究者未能知曉該研究情境是否適合使用「臺灣版 WHOQOL-BREF」，或該測量工具在何種研究情境可獲得較高的分數信度。RG 藉由後設分析法以探索測量工具的分數信度的心理計量表現，其分析結果可被用以

做為選擇測量工具的訊息基礎 (Botella et al., 2010)。因此，本研究採用 RG 以檢驗「臺灣版 WHOQOL-BREF」之分數信度的心理計量特徵，有助於心理計量研究者深入了解該測量工具之分數信度的心理計量表現，其研究分析結果亦可提供重要訊息給實徵研究者做為是否採用「臺灣版 WHOQOL-BREF」的基礎。

貳、文獻探討

一、信度概化

Vacha-Haase (1998) 指出，因測驗受到樣本特徵與樣本規模等的影響，而有不穩定性，所以探索其分數信度對實徵研究是很重要的，而提出信度概化法，用以研究分數信度的心理計量特徵。RG 為效度概化法 (validity generalization) 的延伸擴展，是一種探討測量工具之分數信度的後設分析方法，可用以描繪分數信度之跨研究的集中情形與變異程度，以及檢驗研究特徵對於跨研究之分數信度係數的變異量之貢獻程度 (Vacha-Haase, 1998; Vacha-Haase, Henson, & Caruso, 2002)。在實際分析上，RG 除了計算某一測量工具之分數信度的跨研究描述統計量 (如平均數、變異數等) 外，並以各篇研究的分數信度為依變項，以研究特徵 (如樣本數、樣本類型等) 為獨變項，採用迴歸分析、相關分析、變異數分析等統計法，以檢驗分數信度的變異性受到哪些研究特徵的影響，或哪些研究特徵可用以解釋或預測分數信度的變異性。研究者可依據能夠顯著影響或預測分數信度的研究特徵來修改其研究設計，以最大化分數信度、效果量、統計考驗力，以及統計顯著性 (Henson & Thompson, 2002)。

Vacha-Haase 等人 (2002) 指出，隨著 RG 研究的累積，可以增進對於信度理論的理解，在測驗使用上更為精準地應用分數信度，更能清楚地了解哪些研究特徵在統計上可顯著影響分數信度，並進一步鼓勵研究者採用 RG，以檢驗各種不同測量工具的分數信度。隨後，Vacha-Haase 與 Thompson (2011) 回顧 PsycInfo 和 ERIC 近 12 年來 RG 的實徵研究，共發現 47 篇研究，涵蓋心理病理 (psychopathology)、生活滿意 (life satisfaction) 等量表。經過研究者們的推廣和鼓勵，本研究仿照 Vacha-Haase 與 Thompson 的搜尋條件，發現從 1998 至 2019 年 10 月以來，有關 RG 的實徵研究已累積至 123

篇，其應用範圍擴展至「兒童與青少年完美主義量表」(Child and Adolescent Perfectionism Scale)、「父母威權問卷」(Parental Authority Questionnaire)、「兒童焦慮與憂鬱量表」(The Revised Child Anxiety and Depression Scale)、「兒童自閉症評估量表」(Childhood Autism Rating Scale)等量表。然根據我們的回顧結果，即使「臺灣版 WHOQOL-BREF」已廣泛應用在測量 QOL 上，但目前尚未有此問卷的 RG 相關研究。根據 Vacha-Haase 等人(2002)的建議與啟發，本研究採用 RG 以檢驗「臺灣版 WHOQOL-BREF」的分數信度，藉以增廣分數信度的文獻累積，同時亦能夠促進研究者對於 WHOQOL-BREF 之分數信度的心理計量知識。

二、「臺灣版 WHOQOL-BREF」

WHO 鑒於先前生活品質之測量工具大多著重在測量疾病或失能對個人所造成的影響，而非建立自完整的生活品質概念，且這些測量工具大多侷限在歐美文化體系，未考量文化適用性的議題。WHO 因而開展 QOL 的研究計畫，招集 15 個國家或地區的專家學者共同研擬、討論、釐清生活品質的概念，並將之定義為：「個人在所處之文化與價值體系中的感受程度，涉及個人對目標、期望、標準與關懷的生活感受。涵蓋個人生活品質之身體健康、心理狀態、獨立程度、社會關係、個人信念、環境六大方面」。WHO 透過訪談、焦點團體等質化方法發展出 1,800 道題目，隨後採用試題分析、因素分析等量化方法挑選試題，進而發展出「世界衛生組織生活品質問卷」(WHOQOL-100)。WHOQOL-100 包含 100 道具文化共通性之健康相關生活品質的題目，用以測量生活品質的 24 個層面，其分屬生理、心理、獨立程度、社會關係、環境，以及心靈／宗教／個人信念六大範疇(The WHOQOL Group, 1995, 1998)。

由於 WHOQOL-100 的題目太多，不適用於需考慮時間及實用性的臨床研究及流行病學調查，WHO 精簡其題目並發展出「簡明版世界衛生組織生活品質問卷」(WHOQOL-BREF)。WHOQOL-BREF 共有 26 題，其中 2 題分別測量整體生活品質及一般健康狀態，其餘 24 題來自於生活品質的 24 個層面，此 24 個層面分屬於四個主要範疇：生理健康範疇(涵蓋原先的生理和獨立程度範疇)、心理範疇(涵蓋原先的心理及心靈／宗教／個人信念範疇)、社會關係範疇、環境範疇。在應用上，研究者依據此四範疇分數評估個體的生活品質。WHOQOL-BREF 目前已翻譯超過 50 餘國的語言(Redko,

Rogers, Bule, Siad, & Choh, 2015)，為實徵研究中，經常被用以測量生活品質的工具之一。「臺灣版 WHOQOL-BREF」於 1997 年由 Yao 等人（2002）發展而成，該量表除保留原始 WHOQOL-BREF 的測量結構：1 題測量整體生活品質、1 題測量一般健康、7 題測量生理健康範疇（physical domain）、6 題測量心理範疇（psychological domain）、3 題測量社會關係範疇（social relationships domain），以及 8 題測量環境範疇（environment domain）外，並於社會關係和環境範疇各加入 1 道本土題，共計 28 題。該量表施測於 18 歲以上的樣本，採自陳報告以評估自己最近兩個星期的主觀生活品質。題目為 5 點量尺，部分負向題目經過轉換後，分數愈高表示有較高的生活品質。研究小組自全臺 17 家醫院蒐集 1,068 份有效問卷，樣本涵蓋不同疾病的病患與一般健康人，以檢驗「臺灣版 WHOQOL-BREF」的計量特徵，四範疇的分數信度 Cronbach's α 分別為生理健康範疇 .76、心理範疇 .70、社會關係範疇（含本土題）.72、環境範疇（含本土題）.77。之後，該問卷也曾被使用於由國家衛生研究院及國民健康局所聯合主導的 2001 年國民健康訪問調查（2001 National Health Interview Survey, NHIS）中，蒐集到一萬三千多筆資料（石曜堂、張新儀、樂冠華、林明珠、蘇文川，2002），經心理計量分析，該筆資料所表現出的信度與效度，與從全臺 17 家醫院蒐集到的資料類似。

如同 WHOQOL-BREF 於國際間受到研究者的廣泛使用，在臺灣亦同樣受到重視，申請使用「臺灣版 WHOQOL-BREF」的研究逐年增加，其使用範圍涵蓋醫學、公共衛生、教育學、心理學等領域，其樣本類型包含大專生、一般成人、老人、病患等。檢視其實徵研究的應用範圍，實已超出原先發展該量表所採用的樣本族群，導致研究者無法確認建立自該樣本的「臺灣版 WHOQOL-BREF」之心理計量表現是否能夠類推至其他樣本族群。再者，即便該量表於國內應用廣泛，目前仍未有研究針對「臺灣版 WHOQOL-BREF」之分數信度的心理計量特徵進行系統性研究。實徵研究者對於「臺灣版 WHOQOL-BREF」之分數信度的心理計量表現，以及哪些研究特徵會影響其分數信度的表現並不清楚，因而無法提供實徵應用者做為選擇「臺灣版 WHOQOL-BREF」的參考訊息。因此，本研究採用信度概化法，以探討「臺灣版 WHOQOL-BREF」之生理健康、心理、社會關係和環境等四範疇的分數信度之心理計量特徵，並檢驗其與研究特徵間的關係。本研究依據所回顧的文獻，整理出 11 個研究特徵（如樣本的平均年齡、樣本男性比例、大專以上畢業或肄業比例、地區等），以作為後續分析的預測變項。透過本研

究，除可促進對於「臺灣版 WHOQOL-BREF」之分數信度的心理計量知識，並可提供實徵研究者是否選擇該測量工具的訊息基礎。

參、方法

考慮到研究所需變項資料之完整度，本研究採用臺灣博碩士論文為樣本資料，檢視「臺灣版 WHOQOL-BREF」之分數信度的心理計量表現，以研究特徵為預測變項，分別以四範疇之分數信度為依變項，來執行迴歸分析、 t 檢驗、變異數分析，以檢驗有哪些研究特徵可顯著地預測、解釋或影響跨研究之分數信度的變異性。同時計算四範疇之分數信度與研究特徵的相關係數，以檢驗兩者是否有顯著關係。

一、資料來源

本研究從「臺灣博碩士論文知識加值系統」與「華藝線上圖書館」兩個資料庫，以「WHOQOL」或「世界衛生組織生活品質」作為關鍵字蒐集 2000 年至 2015 年 8 月的臺灣博碩士論文。能納入本研究的論文必須符合：(1) 有使用 WHOQOL-BREF，並對問卷內容無任何修改（如刪題、更改量尺語詞或題目敘述）與(2)有報告 Cronbach's α 兩個條件。在刪除重複出現於此兩個資料庫的文獻後，共獲得 1,332 篇論文，進一步刪除 170 篇無電子全文亦搜尋不到紙本、612 篇未使用「臺灣版 WHOQOL-BREF」作為測量工具，以及 36 篇更改量表或未使用 Cronbach's α 的研究論文，共納入 514 篇論文，以進行信度概化研究。

在此 514 篇論文中，有 56 篇（10.9%）使用「臺灣版 WHOQOL-BREF」卻未提供分數信度的訊息、有 206 篇（40.1%）直接引用「臺灣版 WHOQOL-BREF」使用手冊或先前其他研究的分數信度係數、有 4 篇（0.8%）僅報告四範疇之分數信度的區間範圍，剩餘 248 篇（48.2%）有採用研究蒐集的樣本資料計算 Cronbach's α ，並呈現其數值。因此，本研究僅以此 248 篇論文納入後續的 RG 分析。由於部分研究論文報告兩個以上的分數信度，如預試階段與正式施測，或分別呈現實驗組和對照組，造成生理健康、心理、社會關係和環境四範疇之分數信度係數的分析資料數有所差異，此四範疇之 Cronbach's α 的分析資料數依次分別有 259、268、254、258 個。

二、預測變項

本研究檢視了近年來有關分數信度的實徵研究（30 篇），發現研究參與者的相關變項，如性別、樣本數、年齡、樣本類型、種族變項（如 Aguayo, Pecino, de la Fuente Solana, & Fernández, 2011; Caruso, 2000; Henson, Kogan, & Vacha-Haase, 2001; Kieffer & Reese, 2002; Reese, Kieffer, & Briggs, 2002; Vacha-Haase, 1998; Vacha-Haase, Tani, Kogan, Woodall, & Thompson, 2001）、測驗工具的相關變項，如問卷版本（如 Aguayo et al., 2011; Caruso, Witkiewitz, Belcourt-Dittloff, & Gottlieb, 2001; Ross, Blackburn, & Forbes, 2005）、問卷題數（如 Beretvas & Pastor, 2003; Caruso et al., 2001; Kieffer & Reese, 2002），以及與分數信度的心理計量特徵相關變項，如測驗分數標準差（如 Aguayo et al., 2011; Caruso et al., 2001; Reese et al., 2002），常被採用作為預測變項。其中，以年齡、樣本類型以及測驗分數標準差經常與量表的分數信度有顯著關係（如 Caruso et al., 2001; Kieffer & Reese, 2002; Vacha-Haase et al., 2001）。由於本研究的主旨是檢驗「臺灣版 WHOQOL-BREF」的信度概化，數個在先前研究常見的預測變項（如種族、問卷版本、問卷題數）不適合應用於本研究中。再者，除了納入先前信度概化之實徵研究經常採用的預測變項，研究者亦根據所回顧論文的研究特徵，歸納、整理出 11 個預測變項，其中 9 個變項為連續量尺，分別為：(1)「臺灣版 WHOQOL-BREF」得分的平均分數，及其(2)標準差、(3)樣本數、(4)樣本的平均年齡，及其(5)年齡標準差、(6)樣本男性比例，因所回顧之文獻多為男女混合樣本，本研究依據先前文獻採用男性比例做為樣本男性比例變項（Botella et al., 2010; Caruso et al., 2001; Shields & Caruso, 2003, 2004），即男性人數除以總人數，介於 0~1。(7)大專以上畢業或肄業比例，為了最大化分析的論文數目，將教育程度定義為大專校院以上畢業或肄業的比例，數值介於 0~1。(8)有宗教信仰比例：計算參與者有宗教信仰的比例，數值介於 0~1。(9)已婚比例：為了盡可能保留納入 RG 分析的論文數目，將婚姻狀態定義為已婚的比例；其餘 2 個預測變項為類別量尺。(10)樣本類型：將完全病人組編碼為 0，非完全病人組編碼為 1，非完全病人組涵蓋學生、一般民眾，混合了病人與非病人。(11)地區：依樣本的居住地區，分北部、中部、南部、東部與其他五組，採虛擬編碼（dummy coding）方式，將地區變項再劃分為 4 個變項，分別為地區 1（1 = 北部，0 = 其他）、地區 2（1 = 中部，0 = 其他）；地區 3（1 = 南部，0 = 其他）與地區 4

(1 = 東部，0 = 其他)。

三、資料分析

本研究依據先前研究的建議 (Henson & Thompson, 2002; Mason, Allam, & Brannick, 2007; Thompson & Vacha-Haase, 2000) 與多數 RG 實徵研究的作法 (Sánchez-Meca, López-López, & López-Pina, 2012)，採用量表的原始分數信度係數 Cronbach's alpha 作為依變項。採用統計軟體 SAS 9.4 進行資料分析，首先藉由描述統計以探討「臺灣版 WHOQOL-BREF」的四範疇之跨研究 Cronbach's alpha 的變異情形。Thompson 與 Vacha-Haase (2000) 指出，RG 並非單指一種分析方法，不會侷限於使用任一種統計分析法，研究者可自行選擇適當的方法。本研究參考此一觀點，並依據預測變項的特徵（連續或間斷變項），以及在分析中盡可能地使用所有的樣本資訊，採用多重方法，如多元迴歸分析、*t* 檢驗、變異數分析、相關分析，以檢驗「臺灣版 WHOQOL-BREF」之四範疇的 Cronbach's alpha 與預測變項的關係，這些統計方法亦經常被先前研究所採用（如 Beretvas & Pastor, 2003; Caruso et al., 2001; Kieffer & Reese, 2002; Reese et al., 2002; Vacha-Haase, 1998）。

在進行多元迴歸分析時，因先前信度概化研究發現樣本同質性會影響分數信度係數 (Botella et al., 2010; Henson et al., 2001; Shields & Caruso, 2003)，本研究先以四範疇分數的標準差為預測變項以檢驗樣本同質性，再放入其他預測變項，以檢驗哪些預測變項可統計上顯著預測四範疇之 Cronbach's alpha 的跨研究變異性。迴歸分析的優點在於可同時檢驗所有預測變項對於四範疇之 Cronbach's alpha 的預測力，然其缺點在於因許多論文未能同時擁有所有考量的預測變項，容易產生大量的遺漏值，導致納入迴歸分析模型中的論文數目很少。為了在 RG 的分析中，盡可能地納入較多的論文數目，本研究進一步地針對連續型與類別型的預測變項分別進行皮爾森相關分析（平均分數、分數的標準差、樣本數、平均年齡、年齡的標準差、樣本男性比例、大專以上畢業或肄業比例、有宗教信仰比例、已婚比例變項）、*t* 檢驗（樣本類型變項），以及變異數分析（地區變項）。再者，透過交叉比對多重統計方法的分析結果，以期望增進了解預測變項與四範疇之分數信度係數的關係。

肆、研究結果

RG 的研究結果經常面臨檔案抽屜問題（file-drawer problem），意指當未發表或未出版之研究所發現的心理計量特徵與已發表或出版的研究發現不一致時，RG 的研究結果可能因為沒有納入這些未發表或未出版的研究，而導致偏誤的研究結果（Howell & Shields, 2008）。為了解決此一問題，Howell 與 Shields（2008）發展出評估檔案抽屜問題的方法，並建議後續研究者於執行 RG 研究前，需先評估其研究結果是否容易受到檔案抽屜問題的影響。本研究根據其建議計算四範疇的失敗—安全研究數目（Fail-Safe N），其生理健康、心理、社會關係與環境範疇之失敗—安全研究數目分別為 276.43、332.5、78.75、480，表示若此四範疇之未出版研究的分數信度低於 .64~.65，分別達到 277、333、79 或 480 篇，將會改變 RG 的研究結果，亦即四範疇之平均母群分數信度低於 .70，據此研究者相信本研究結果是穩定的。

表 1 指出四範疇之跨研究的平均數介於 .79~.71，具有合理的信度，且其標準差介於 .09~.07，表示四範疇之分數信度的跨研究變異情形小，具有一定程度之跨研究的穩定性。由於部分研究之四範疇（生理範疇 1 筆、其餘三範疇各有 2 筆）的分數信度係數過低，可能會影響其後續的統計分析結果，故予以刪除。結果共刪除 9 筆資料，其中 3 筆資料來自單篇研究，其餘 6 筆資料分別來自不同研究。刪除後之數值列於表 1 括弧內。

表 1 「臺灣版 WHOQOL-BREF」之四範疇跨研究的描述統計

範疇	樣本數	平均數	標準差	最小值	最大值
生理健康	259 (258)	.76 (.76)	.08 (.07)	.44 (.51)	.93 (.93)
心理	268 (266)	.77 (.77)	.07 (.07)	.38 (.51)	.94 (.94)
社會關係	254 (252)	.71 (.72)	.09 (.08)	.27 (.44)	.93 (.93)
環境	258 (256)	.79 (.79)	.07 (.06)	.33 (.57)	.95 (.95)

註：括弧內為刪除分數信度過低之論文數的分析結果。

表 2 呈現採用 11 個研究特徵為預測變項，分別對四範疇之分數信度係數的迴歸分析結果。由於同時考量所有預測變項的論文數目很少，因此納入分析之樣本數大幅降低：生理健康範疇僅 53 筆資料、心理範疇僅 54 筆資料、社會關係範疇僅 45 筆資料、環境範疇僅 50 筆資料。

表 2 「臺灣版 WHOQOL-BREF」之四範疇的迴歸分析摘要表

變項	樣本數	自由度	迴歸係數	標準誤	t 值	標準化迴歸係數	R ² (調整後 R ²)	F 值
生理健康範疇	53							
截距		1	0.51	0.04	11.52***	0.00	.45	20.52***
分數標準差		1	0.07	0.01	5.43***	0.59	(.43)	
有宗教信仰比例		1	0.10	0.05	2.08*	0.22		
心理範疇	54						.18	11.00**
截距		1	0.63	0.04	14.45***	0.00		
分數標準差		1	0.06	0.02	3.32**	0.42		
社會關係範疇	45						.16	7.95**
截距		1	0.60	0.05	12.35***	0.00		
分數標準差		1	0.06	0.02	2.82**	0.40		
環境範疇	50							
截距		1	0.69	0.04	19.59***	0.00	.36	8.48**
分數標準差		1	0.04	0.02	2.42*	0.29	(.31)	
樣本數		1	-0.00	0.00	-2.46*	-0.29		
大專以上畢業或肄業比例		1	0.10	0.03	3.62**	0.43		

* $p < .05$, ** $p < .01$, *** $p < .001$

生理健康範疇之分數標準差與有宗教信仰比例可顯著地預測生理健康範疇之分數信度係數，調整後之模型解釋變異量為 $R^2 = .43$ ， $F(2, 47) = 20.52$ ， $p < .0001$ ，其中以生理健康範疇分數標準差的貢獻最大 ($\beta = .59$)，其次為有宗教信仰比例變項 ($\beta = .22$)，顯示當分數的標準差或參與者有宗教信仰的比例愈大時，生理範疇之 Cronbach's α 會愈大。

僅心理或社會關係範疇之分數標準差能夠顯著預測心理或社會關係範疇的分數信度，其模型解釋變異量分別為 $R^2 = .175$ ， $F(1, 52) = 11.00$ ， $p < .01$ 與 $R^2 = .156$ ， $F(1, 43) = 7.95$ ， $p < .01$ ，顯示心理或社會關係範疇之分數標準差愈大，其分數信度愈高。

環境範疇分數標準差、樣本數、大專以上畢業或肄業比例可顯著預測環境範疇之分數信度，調整後之模型解釋變異量為 $R^2 = .314$ ， $F(3, 46) = 8.48$ ， $p = .0001$ ，其中以大專以上畢業或肄業比例變項 ($\beta = .43$) 的貢獻最大，其次為分數標準差 ($\beta = .29$) 和樣本數 ($\beta = -.29$)，顯示當教育程度愈高、環境範疇的分數標準差愈大、樣本數愈小，則環境範疇的分數信度愈高。

表 3 為四範疇之分數信度係數與連續量尺之預測變項的皮爾森相關係數，相較於迴歸分析納入所有預測變項（並非所有研究同時考量全部的預測變項），因皮爾森相關係數僅分析分數信度係數與單一個預測變項間的相關，使納入分析的樣本資料從迴歸分析中的 40 至 60 的樣本資料，大幅提高到 150 至 180。由於樣本數差距甚大，而導致迴歸分析與相關分析結果有所不同。與迴歸分析相似，相關分析亦發現四範疇的分數標準差對於其 Cronbach's α 有顯著正相關，除社會關係範疇之分數標準差與其分數信度之相關係數未符合預期達顯著水準。此結果與先前研究發現一致，分數標準差為預測分數信度係數的重要變項（Caruso et al., 2001; Shields & Caruso, 2004）。

平均年齡與生理健康範疇有顯著正相關，然與環境範疇呈現顯著負相關，顯示隨平均年齡愈大，生理健康範疇之 Cronbach's α 愈高，但環境範疇之 Cronbach's α 則愈低。樣本男性比例與生理範疇有顯著正相關，意指男性樣本數愈多，生理範疇的 Cronbach's α 愈高。然在迴歸分析中，平均年齡與樣本男性比例均無法顯著地預測四範疇的分數信度係數。大專以上畢業或肄業比例與生理健康範疇有顯著負相關，與環境範疇則有顯著正相關，表示教育程度愈高，生理健康範疇的 Cronbach's α 愈低，而環境範疇的 Cronbach's α 則愈高。樣本數、有宗教信仰與四範疇 Cronbach's α 的相關係數未如預期達顯著水準，與迴歸分析結果不一致。

由於納入迴歸分析與相關分析之樣本資料量的差異甚巨，造成此兩種分析在發現與分數信度係數有顯著關係的預測變項上有所不同，於實徵應用上，應需綜合考量這些預測變項對於分數信度係數的影響或關係。

進一步計算排除掉各範疇分數標準差的影響後，研究變項與四範疇之分數信度係數的淨相關，結果顯示平均年齡愈大，生理健康範疇的 Cronbach's α 愈高，其淨相關係數為 .21, $p < .05$ ，但環境範疇的 Cronbach's α 會愈低，淨相關係數為 -.21, $p < .05$ ，與迴歸分析的結果一致；參與者教育程度愈高，環境範疇的 Cronbach's α 會愈高，淨相關係數為 .28, $p < .01$ 。

四範疇之分數信度係數與樣本類型（全病人組與非全病人組）之 t 檢驗以及地區變項（北部、中部、南部、東部、混合五組）之變異數分析結果顯示，四範疇之分數信度係數的平均數在不同樣本類型或不同地區皆未達顯著差異，此結果與迴歸分析之結果一致，為節省空間，結果未列於內文。

表 3 「臺灣版 WHOQOL-BREF」四範疇的分數信度與連續型研究變項的相關係數摘要表

		生理健康	心理	社會關係	環境
平均分數	<i>r</i>	0.00	-0.07	-0.05	-0.11
	<i>N</i>	157	164	148	155
平均分數的標準差	<i>r</i>	0.52***	0.32***	0.04	0.37***
	<i>N</i>	145	150	132	140
樣本數	<i>r</i>	-0.07	-0.04	0.11	0.01
	<i>N</i>	231	239	225	232
平均年齡	<i>r</i>	0.40***	0.03	0.05	-0.21*
	<i>N</i>	138	141	127	132
年齡標準差	<i>r</i>	-0.01	-0.03	-0.05	0.06
	<i>N</i>	127	129	115	120
樣本男性比例	<i>r</i>	0.22**	0.04	0.15	-0.12
	<i>N</i>	179	184	169	176
大專以上畢業或肄業比例	<i>r</i>	-0.21*	0.13	0.14	0.20*
	<i>N</i>	142	147	134	139
有宗教信仰比例	<i>r</i>	0.09	0.04	-0.09	0.09
	<i>N</i>	95	97	85	92
已婚比例	<i>r</i>	-0.08	-0.06	0.04	-0.04
	<i>N</i>	153	159	146	152

* $p < .05$, ** $p < .01$, *** $p < .001$

伍、結論與建議

本研究採用 RG 以探討「臺灣版 WHOQOL-BREF」之四範疇分數信度的心理計量表現，並檢驗有哪些研究特徵可預測或解釋其分數信度的變異性。在回顧的 514 篇臺灣博碩士論文中，有 48% 的研究依據其樣本資料計算並報告四範疇的分數信度。此比例較高於先前文獻的發現，例如：Vacha-Haase、Ness、Nilsson 與 Reetz (1999) 回顧《諮商心理學》(*Journal of Counseling Psychology*)、《心理學與老年》(*Psychology & Aging*)，以及《專業心理學：研究與實作》(*Professional Psychology: Research and Practice*) 等三個心理學期刊，發現僅 36% 的量化研究根據其樣本資料計算與報告分數信度。其部分原因可能是因為博碩士論文的篇幅較一般期刊長，允許研究者詳盡地報

告研究分析結果，然期刊可能受限於篇幅限制，而未能報告分數信度的訊息。另一方面，亦可能是國內研究者已意識到信度是測驗分數結果的特質，而非測驗工具本身（Crocker & Algina, 1986; Gronlund & Linn, 1990; Rowley, 1976），因此多數研究者使用研究所蒐集到的樣本資料，以計算該研究所採用的測量工具之分數信度係數，而非直接引用自該測量工具的使用手冊或先前研究的分數信度結果。

本研究發現「臺灣版 WHOQOL-BREF」之四範疇的跨研究平均分數信度皆在 .70 以上，此一結果相似於林立中、姚開屏（2019）回顧 WHOQOL-BREF 之心理計量特徵的研究發現。就社會科學研究領域，.70 以上為可接受之分數信度係數範圍（Nunnally, 1978），且其跨研究的變異性小，亦即在多元研究環境背景中，如不同的樣本數和樣本類型、不同的研究領域等，多數研究皆能獲得近似的分數信度係數值，顯示 WHOQOL-BREF 具有良好且跨研究穩定的分數信度。

藉由分析「臺灣版 WHOQOL-BREF」之四範疇分數信度與預測變項（研究特徵）的關係發現，四範疇的分數標準差可預測四範疇的分數信度之變異性，此結果與理論預期和先前的 RG 研究發現一致：分數標準差愈大，Cronbach's α 就會愈大。相關分析亦獲得相似的研究結果，生理、心理、環境範疇的分數信度與分數信度之標準差有顯著正相關，僅社會關係範疇之分數信度未達顯著相關。在其餘預測變項中，僅少部分預測變項與分數信度有顯著性的關係，大專以上畢業或肄業比例變項同時在迴歸和相關分析中，一致地呈現出與環境範疇的分數信度具有顯著正向關係，意即參與者的教育程度為大專校院以上的比例愈高，環境範疇之 Cronbach's α 會愈大。在平均年齡變項的分析結果上，與先前其他 RG 的研究發現相似，如 Shields 與 Caruso（2003）針對酒精使用疾患確認檢測（Alcohol Use Disorders Identification Test）進行信度概化研究，平均年齡愈小、分數信度愈高。Ryngala、Shields 與 Caruso（2005）使用「兒童顯性焦慮量表修訂版」（Revised Children's Manifest Anxiety Scale）進行信度概化研究，結果發現都是平均年齡愈大、分數信度愈高。平均年齡對 Cronbach's α 的影響似乎會隨著問卷內容不一樣，而有不同的方向。本研究也有類似的發現，在生理和環境範疇平均年齡的預測方向也是相反的，意即將生理與環境範疇施測在年齡愈小的樣本上，可獲得的分數信度愈高。

內部一致性信度 Cronbach's α 在「臺灣版 WHOQOL-BREF」之四範疇的

信度概化是本文之研究焦點，亦是最常被研究論文所報告的信度係數。然除了著重在內部一致性的信度係數，尚有著重於時間穩定性的再測信度（test-retest reliability），以及聚焦在不同評分者間的一致性評分者間信度（inter-rater reliability）。未來可檢驗此兩種信度係數在「臺灣版 WHOQOL-BREF」之四範疇的信度概化程度，以期能夠更全面性地了解「臺灣版 WHOQOL-BREF」的信度計量特徵。

本研究為了納入較多使用「臺灣版 WHOQOL-BREF」的論文篇數，採用博碩士論文為樣本資料，用以解決進行 RG 時，可能面臨到可分析論文篇數容易偏少的問題。然由於每篇論文的研究設計不盡相同，不同研究所共享的研究特徵更是少數，部分研究特徵由於未被多數研究所考量，而未能將這些研究特徵納入分析，有可能會遺漏了某些影響「臺灣版 WHOQOL-BREF」之四範疇的分數信度變項，如 Youngstrom 與 Green（2003）研究發現，社經地位可顯著地預測「差異性情緒量表」（Differential Emotions Scale）的分數信度，然在本篇研究未能將此預測變項納入分析。因此，本研究的限制在於因為某些研究特徵僅被少數研究所考量，而無法納入後續的分數信度概化分析之中，無法全面性的考量影響分數信度的預測變項。

本研究透過 RG 方法並採用多重統計分析方法以檢驗 11 個預測變項（研究特徵）與「臺灣版 WHOQOL-BREF」之四範疇分數信度係數的關係，可提供實徵研究者選擇使用 WHOQOL-BREF 的訊息基礎，或實徵應用者可根據本研究發現與 WHOQOL-BREF 分數信度有顯著關係的預測變項，以修改自身的研究設計，以期獲得最大化的 WHOQOL-BREF 之分數信度。綜合迴歸、相關等分析法的結果，本研究建議未來在評估是否使用「臺灣版 WHOQOL-BREF」時，因分數標準差為預測四範疇的分數信度之重要變項，實徵應用者須將樣本資料的變異性納入考量。再者，亦需評估大專以上畢業或肄業比例、有宗教信仰比例與年齡變項，尤其是在生理健康、環境範疇上。生理健康範疇在研究參與者具有宗教信仰或年齡較大者，可得到較高的分數信度。環境範疇在研究參與者具有較高的教育程度、較低的年齡或較低的樣本數時，可獲得較高的分數信度。而心理與社會關係範疇僅受到該範疇之分數標準差的影響，不受其他樣本男性比例、樣本數、平均年齡、大專以上畢業或肄業比例、有宗教信仰比例、地區等預測變項的影響，此顯示心理與社會關係範疇之分數信度具有強韌性，適用於多樣的研究背景環境。

林立中 蘇意晴 姚開屏 「臺灣簡明版世界衛生組織生活品質問卷」之信度概化研究

謝誌

本研究得以順利完成，感謝科技部計畫（105-2918-I-002-006、106-2410-H-002-081-SSS、108-2410-H-002-100、109-2410-H-002-091-SS2）之經費補助。

參考文獻

中文部分

- 石曜堂、張新儀、樂冠華、林明珠、蘇文川（2002）。國民健康訪問調查簡介。
國民健康訪問調查研究簡訊，1，1-8。
- 林立中、姚開屏（2019）。WHOQOL-BREF 之心理計量特徵的回顧研究。**臺灣
衛誌**，38，590-603。

英文部分

- Aguayo, R., Pecino, C. V., de la Fuente Solana, E. I., & Fernández, L. M. L. (2011). A meta-analytic reliability generalization study of the Maslach Burnout Inventory. *International Journal of Clinical and Health Psychology*, 11(2), 343-361.
- Beretvas, S. N., & Pastor, D. A. (2003). Using mixed-effects models in reliability generalization studies. *Educational and Psychological Measurement*, 63(1), 75-95. <https://doi.org/10.1177/0013164402239318>
- Botella, J., Suero, M., & Gambara, H. (2010). Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychological Methods*, 15(4), 386. doi:10.1037/a0019626
- Caruso, J. C. (2000). Reliability generalization of the NEO Personality Scales. *Educational and Psychological Measurement*, 60(2), 236-254. doi:10.1177/00131640021970484
- Caruso, J. C., Witkiewitz, K., Belcourt-Dittloff, A., & Gottlieb, J. D. (2001). Reliability of scores from the Eysenck Personality Questionnaire: A reliability generalization study. *Educational and Psychological Measurement*, 61(4), 675-689. doi:10.1177/00131640121971437
- Chen, K. H., Wu, C. H., & Yao, G. (2006). Applicability of the WHOQOL-BREF on early adolescence. *Social Indicators Research*, 79(2), 215-234. doi:10.1007/s11205-005-0211-0
- Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt, Rinehart, and Winston.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of*

- behavioral measurements: Theory of generalizability of scores and profiles*. New York, NY: John Wiley & Sons.
- Graham, J. M., & Christiansen, K. (2009). The reliability of romantic love: A reliability generalization meta-analysis. *Personal Relationships*, 16(1), 49-66. doi:10.1111/j.1475-6811.2009.01209.x
- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching*. London, UK: Collier Macmillan.
- Henson, R. K., & Thompson, B. (2002). Characterizing measurement error in scores across studies: Some recommendations for conducting reliability generalization studies. *Measurement and Evaluation in Counseling and Development*, 35(2), 113-126. doi:10.1080/07481756.2002.12069054
- Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the teacher efficacy scale and related instruments. *Educational and Psychological Measurement*, 61(3), 404-420. doi:10.1177/00131640121971284
- Howell, R. T., & Shields, A. L. (2008). The file drawer problem in reliability generalization: A strategy to compute a fail-safe N with reliability coefficients. *Educational and Psychological Measurement*, 62(1), 120-128. doi:10.1177/0013164407301528
- Kieffer, K. M., & Reese, R. J. (2002). A reliability generalization study of the Geriatric Depression Scale. *Educational and Psychological Measurement*, 62(6), 969-994. doi:10.1177/0013164402238085
- Mason, C., Allam, R., & Brannick, M. T. (2007). How to meta-analyze coefficient-of-stability estimates: Some recommendations on Monte Carlo studies. *Educational and Psychological Measurement*, 67(5), 765-783. doi:10.1177/0013164407301532
- Nedjat, S., Monazeri, A., Holakouie, K., Mohammad, K., & Majdzadeh, R. (2008). Psychometric properties of the Iranian interview-administered version of the World Health Organization's quality of life questionnaire (WHOQOL-BREF): A population-based study. *BMC Health Services Research*, 8, 61-68. doi:10.1186/1472-6963-8-61
- Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.
- Redko, C., Rogers, N., Bule, L., Siad, H., & Choh, A. (2015). Development and validation of the Somali WHOQOL-BREF among refugees living in the USA. *Quality of Life Research*, 24(6), 1503-1513. doi:10.1007/s11136-014-0877-3
- Reese, R. J., Kieffer, K. M., & Briggs, B. K. (2002). A reliability generalization study of select measures of adult attachment style. *Educational and Psychological Measurement*, 62(4), 619-646. doi:10.1177/001316402128775058
- Ross, M. E., Blackburn, M., & Forbes, S. (2005). Reliability generalization of the patterns

- of adaptive learning survey goal orientation scales. *Educational and Psychological Measurement*, 65(3), 451-464. doi:10.1177/0013164404272496
- Rowley, G. L. (1976). Notes and comments: The reliability of observational measures. *American Educational Research Journal*, 13(1), 51-59. doi:10.3102/00028312013001051
- Ryngala, D. J., Shields, A. L., & Caruso, J. C. (2005). Reliability generalization of the revised children's manifest anxiety scale. *Educational and Psychological Measurement*, 65(2), 259-271. doi:10.1177/0013164404272495
- Sánchez-Meca, J., López-López, J. A., & López-Pina, J. A. (2012). Some recommended statistical analytic practices when reliability generalization studies are conducted. *British Journal of Mathematical and Statistical Psychology*, 66(3), 402-425. doi:10.1111/j.2044-8317.2012.02057.x
- Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist*, 44(6), 922-932. doi:10.1037/0003-066X.44.6.922
- Shields, A. L., & Caruso, J. C. (2003). Reliability generalization of the alcohol use disorders identification test. *Educational and Psychological Measurement*, 63(3), 404-413. doi:10.1177/0013164403063003004
- Shields, A. L., & Caruso, J. C. (2004). A reliability induction and reliability generalization study of the CAGE questionnaire. *Educational and Psychological Measurement*, 64(2), 254-270. doi:10.1177/0013164403261814
- The WHOQOL Group. (1995). The World Health Organization Quality of Life Assessment (WHOQOL): Position paper from the World Health Organization. *Social Science Medicine*, 41(10), 1403-1409. doi:10.1016/0277-9536(95)00112-K
- The WHOQOL Group. (1998). Development of the world health organization WHOQOL-BREF quality of life assessment. *Psychological Medicine*, 28(3), 551-558. doi:10.1017/s0033291798006667
- Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and Psychological Measurement*, 60(2), 174-195. doi:10.1177/0013164400602002
- Trompenaars, F. J., Masthoff, E. D., Van Heck, G. L., Hodiament, P. P., & Vries, J. D. (2005). Content validity, construct validity, and reliability of the WHOQOL-BREF in a population of Dutch adult psychiatric outpatients. *Quality of Life Research*, 14(1), 151-160. doi:10.1007/s11136-004-0787-x
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measure-*

- ment, 58(1), 6-20. doi:10.1177/0013164498058001002
- Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development, 44*(3), 159-168. doi:10.1177/0748175611409845
- Vacha-Haase, T., Henson, R. K., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement, 62*(4), 562-569. doi:10.1177/0013164402062004002
- Vacha-Haase, T., Ness, C. M., Nilsson, J., & Reetz, D. (1999). Practices regarding reporting of reliability coefficients: A review of three journals. *Journal of Experimental Education, 67*(4), 335-341. doi:10.1080/00220979909598487
- Vacha-Haase, T., Tani, C. R., Kogan, L. R., Woodall, R. A., & Thompson, B. (2001). Reliability generalization: Exploring reliability variations on MMPI/MMPI-2 validity scale scores. *Assessment, 8*(4), 391-401. doi:10.1177/107319110100800404
- Yao, G., Chung, C. W., Yu, C. F., & Wang, J. D. (2002). Development and verification of validity and reliability of the WHOQOL-BREF Taiwan version. *Journal of the Formosan Medical Association, 101*, 342-351.
- Youngstrom, E. A., & Green, K. W. (2003). Reliability generalization of self-report of emotions when using the Differential Emotions Scale. *Educational and Psychological Measurement, 63*(2), 279-295. doi:10.1177/0013164403253226

