



國立台灣大學理學院心理學研究所

碩士論文

Graduate Institute of Psychology
College of Science

National Taiwan University

Master Thesis

台灣簡明版世界衛生組織生活品質問卷的信度概化研究

Reliability Generalization

of the WHOQOL-BREF Taiwan Version

蘇意晴

Yi-Ching Su

指導教授：姚開屏博士

Advisor: Kai-Ping Yao, Ph.D.

中華民國 105 年 7 月

July, 2016

摘要



信度是屬於測驗分數結果的一項特質，而不是屬於測驗工具本身(Gronlund & Linn, 1990)。因此每個研究都應該報告自己的研究樣本所估計出的信度係數。但實際上許多研究都沒有報告自己的信度係數。有鑑於此，Vacha-Haase 於 1998 年提出一種信度的整合分析法，稱為信度概化法(reliability generalization, RG)，藉此整合使用同一個測量工具且有報告信度係數的許多研究。在台灣尚未有信度概化研究，因此本研究將針對台灣簡明版世界衛生組織生活品質問卷（台灣版 WHOQOL-BREF）進行信度概化研究，同時也修正一些過去信度概化研究可能有的問題，目的是（一）討論題目、量尺等等各方面完全一樣的測量工具在不同研究中，不同群體的受試者產生的 Cronbach's α 有何變異。（二）這樣的變異可以用哪些和研究參與者有關的因素（例如：參與者性別、年齡）來預測。台灣版 WHOQOL-BREF 包含了生理、心理、社會關係和環境四個範疇。結果發現不管在整份問卷或四個範疇，Cronbach's α 在不同研究的變異程度都不小，但大部分研究都有達到.7，一般認為可接受的信度數值。測驗分數的標準差對整份問卷和四個範疇的 Cronbach's α 都有顯著影響。若去除掉分數標準差的影響，平均年齡是最有影響的變項，平均年齡越大，生理範疇的 Cronbach's α 也會越大，但環境範疇的 Cronbach's α 會越小。

關鍵詞：信度概化、生活品質、Cronbach's α 、台灣簡明版世界衛生組織生活品質問卷

Reliability Generalization of the WHOQOL-BREF Taiwan Version



Yi-Ching Su

Abstract

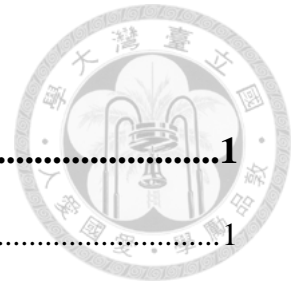
“Reliability refers to the results obtained with an evaluation instrument and not to the instrument itself” (Gronlund & Linn, 1990). Therefore, all researchers should provide reliability from the data actually analyzed in their study. However, in the past many researchers didn’t report it. Hence Vacha-Haase (1998) proposed the reliability generalization (RG) method which is a meta-analytic procedure for synthesizing reliability estimates from different studies. Because there hasn’t been any reliability generalization study in Taiwan, we conduct a reliability generalization study of the WHOQOL-BREF Taiwan version to explore the amount of variability in Cronbach's α across studies, and to examine the factors that may predict the variability in Cronbach's α . The WHOQOL-BREF Taiwan version encompasses four domains: physical health, psychological, social relationships and environment. The results revealed that Cronbach's α for the complete questionnaire and four domains across studies were fairly variable, but most of them were higher

than .7, an acceptable value for a measurement. Cronbach's α were substantially related to standard deviations of scores. After removing the effect of standard deviations of scores, mean age was the primary predictor of Cronbach's α . Higher mean age was associated with higher Cronbach's α for physical health domain, but lower Cronbach's α for environment domain.

Keywords: reliability generalization, quality of life, Cronbach's α ,

WHOQOL-BREF Taiwan version

目次



第一章 緒論	1
第一節 研究動機	1
第二節 古典測驗理論與信度	2
第三節 影響信度的因素	4
第四節 信度概化法	7
第五節 信度概化法的相關研究	10
第二章 研究方法	19
第一節 研究對象	19
第二節 資料蒐集	24
第三節 研究變項說明	25
第四節 資料分析	29
第三章 研究結果	32
第一節 信度係數報告的情形	32
第二節 研究變項之描述統計	36
第三節 Cronbach's α 的預測因子	40
第四節 Cronbach's α 與獨變項兩兩變項間關係	47
第四章 討論與結論	52
第一節 信度係數報告的情形	52
第二節 不同研究中 Cronbach's α 的變異	53
第三節 影響 Cronbach's α 在不同研究中變異的因子	54
第四節 研究貢獻	56
第五節 研究限制與未來研究方向	57
參考文獻	59



表目次



表 1 題數對信度的影響	6
表 2 信度概化法的相關研究整理	12
表 3 台灣版 WHOQOL-BREF 的題目與所屬範疇	23
表 4 教育程度編碼例子	28
表 5 原始學門分類	34
表 6 整份問卷和四個範疇的 Cronbach's α 描述統計	36
表 7 刪除界外值後整份問卷和四個範疇的 Cronbach's α 的描述統計	38
表 8 連續型獨變項的平均數和標準差	39
表 9 類別型獨變項的各組人數和百分比	40
表 10 整份問卷 Cronbach's α 的迴歸分析 ($N = 25$)	41
表 11 生理範疇 Cronbach's α 的迴歸分析 ($N = 53$)	43
表 12 心理範疇 Cronbach's α 的迴歸分析 ($N = 54$)	44
表 13 社會關係範疇 Cronbach's α 的迴歸分析 ($N = 45$)	45
表 14 Cronbach's α 、樣本數、平均年齡、性別和教育程度間皮爾森相關 ($N = 50$)	46
表 15 去除分數標準差和樣本數，年齡、性別、教育程度和 α 的淨相關 ($N = 50$)	46
表 16 環境範疇 Cronbach's α 的迴歸分析 ($N = 50$)	47
表 17 Cronbach's α 與連續型獨變項的皮爾森相關	49
表 18 去掉分數標準差，平均年齡、性別、教育程度和 Cronbach's α 淨相關	50
表 19 「全病人」和「非全病人」研究的 Cronbach's α 比較	50
表 20 各地區研究的 Cronbach's α 比較	51

圖目次

圖 1：全距限制對相關係數的影響	5
圖 2：台灣版 WHOQOL-BREF 的使用單位	20
圖 3：台灣版 WHOQOL-BREF 的研究部門	20
圖 4：資料蒐集流程圖	25
圖 5：使用台灣版 WHOQOL-BREF 的論文數量隨畢業學年度的變化	33
圖 6：有報告自己樣本的 Cronbach's α 確切數值的比例隨畢業學年度的變化....	33
圖 7：各學門使用台灣版 WHOQOL-BREF 的論文數量及信度報告情形	35
圖 8：各學門四類信度報告情形的比例	35
圖 9：整份問卷和四個範疇的 Cronbach's α 箱型圖.....	37
圖 10：社會範疇 Cronbach's α 和分數標準差的散佈圖	55



第一章 緒論



第一節 研究動機

在使用測驗工具的時候，我們會希望假如同樣的受試者在同樣的情形下重複受測，測驗結果應該要是一樣，這種測驗分數的穩定性正是信度所要探討的 (Crocker & Algina, 1986; Cronbach, 1990)。信度除了用在探討時間上的穩定性，也可以用來表示測驗題目的內部一致性、評分者間的一致性、測驗複本間的一致性。因此在發展測驗工具的過程中，信度的研究是不可或缺的部分。

許多研究者使用測驗工具時只有引用別人的信度結果，沒有用自己的研究樣本的測驗結果來估計信度係數。但其實信度是屬於測驗分數結果的一項特質，而不是屬於測驗工具本身 (Gronlund & Linn, 1990)。因此每個研究都應該報告自己的研究樣本所估計出的信度係數。

信度不好最直接的影響就是會弱化 (attenuate) 相關係數，意思是兩個變項實際上的相關會因為測驗分數的信度不好，而使得觀察到的相關係數變低。這會讓研究者對研究結果做出不適當的解釋，也影響到研究結果後續的應用與發展。相關係數弱化也會影響到使用相關係數去估計效度的時候，效度代表一個測驗工具可以測量到它想要測量的東西的程度 (Allen & Yen, 1979)，效度係數的數值和信度係數有關 (Gregory, 2007)，如果信度不好，效度也會跟著不好。換言之，信度是效度的必要非充分條件。

由於信度是測量上的基礎，許多關於信度的研究相繼而生，其中 Vacha-Haase 於 1998 年提出一種信度的整合分析法 (meta-analysis)，稱為信度概化法 (reliability generalization, RG)，目的是 (一) 找出測驗工具在不同研究中的

信度係數。(二) 探討測驗工具在不同研究中信度係數的變異程度。(三) 找出哪些因子可以預測測驗工具在不同研究中信度係數的變異。關於信度概化法及過去有關的研究，後續的章節將會有更詳細的介紹。信度概化法提出後，許多學者紛紛使用此方法對一些常用的測驗工具進行信度概化研究，但在台灣尚未有研究者對任何量表做過信度概化研究。綜觀這些信度概化研究，會發現這些研究可能有些方法上的問題，例如：將不同的問卷或不同種類的信度混在一起分析。因此基於上述理由，本研究將使用此方法進行研究，同時也修正一些可能有的潛在問題。

以下將針對古典測驗理論和信度的概念先做一個說明，並回顧一些傳統測驗書籍中常討論到影響信度的因素，之後再詳細介紹信度概化法，以及過去所做過的信度概化研究。

第二節 古典測驗理論與信度

古典測驗理論

根據古典測驗理論(classical test theory)一個人的測驗（觀察）分數(X)等於真實分數(T)加上測量誤差(e)：

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

其中真實分數是研究者想要測量的特質分數，假設沒有練習效果和疲勞效果，真實分數應該是穩定不變的；測量誤差則是受到研究者想測量的特質以外的東西所影響的部分，可能是受試者有關的因素、施測情境或測量工具本身造成的。

古典測驗理論有以下基本假設(Gulliksen, 1950)：

1. 測量誤差是隨機的。
2. $E(e) = 0$ ，測量誤差的期望值為 0。
3. $r_{Te} = 0$ ，真實分數和測量誤差的相關為 0。

4. $r_{e_1 e_2} = 0$ ，不同次測驗的測量誤差之間相關為 0。

假如現在有一群人接受相同的測驗，由式子(1)和假設 3 可以得到觀察分數的變異數正好等於真實分數的變異數加上測量誤差的變異數：

$$\sigma_X^2 = \sigma_T^2 + \sigma_e^2。 \quad (2)$$

信度係數

信度係數(reliability coefficient)(ρ_{XX})就是真實分數的變異數除以觀察分數的變異數：

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

代表觀察分數的變異中真實分數的變異所占的比例。假如完全沒有測量誤差存在時，信度係數會等於 1；若測量誤差非常大時，信度係數會趨近於 0。由於研究者感興趣的真實分數不能直接被測量到，所以測驗（觀察）分數是否能正確反映出真實分數就變得很重要，這也就是信度所關注的問題，我們會希望得到夠高的信度係數。

不同的測量誤差會有不同的估計信度係數的方法，比方說再測信度(test-retest reliability)反映的是時間所造成的誤差，假設研究者感興趣的真實分數是穩定不變的，同一群受試者兩次測驗分數應該相等，若不相等即是時間間隔所產生的誤差。複本信度(alternate-forms reliability)反映的是同一個測驗不同複本之間所產生的誤差，需讓同一群人接受不同複本的施測，以估計複本產生的測量誤差。評分者間信度(inter-rater reliability)關注的誤差來源就是來自不同評分者之間產生的誤差，有些測驗的分數需要評分者來評分，此時不同的評分者可能會得到不同的分數結果。內部一致性信度(internal consistency reliability)反映出測驗中同個構念不同題目之間的一致性，相較於上述三種信度是最容易估計，只需要一次施測，不需要不同複本或不同評分者，因此是最多研究者使



用的信度係數，其中又以 Cronbach's α 最常見(Hogan, Benjamin, & Brezinski, 2000)，本研究為了盡可能取得最多的研究論文篇數，選擇 Cronbach's α 作為研究變項，以下便針對 Cronbach's α 進一步介紹。



Cronbach's α

由 Cronbach (1951)提出的一種估計內部一致性信度的方法，公式如下：

$$\alpha = \left(\frac{k}{k-1}\right) \left(1 - \frac{\sum \sigma_j^2}{\sigma^2}\right), \quad (4)$$

其中 k 為題目數， σ_j^2 為第 j 題分數的變異數 ($j = 1, 2, \dots, k$)， σ^2 是總分的變異數。總分的變異數 (σ^2) 可以拆解成兩部分，一部分是單一題目分數的變異數總和 ($\sum \sigma_j^2$)，另一部分則是兩兩題目間共變數總和 ($\sum_{i \neq j} \sigma_{ij}$ ，其中 σ_{ij} 表示第 i 題和第 j 題分數的共變數)，因此由式子(4)可以看出 α 越高，代表在總分的變異中，兩兩題間共變數總合占的比例越大。

過去認為 Cronbach's α 高表示這些題目是測量同一個構念，但 Cortina (1993) 和 Schmitt (1996)認為這是對 Cronbach's α 的誤解，題目間相關高不一定代表測量同一個構念，只能說若是測量同一個構念，則題目間相關高，Cronbach's α 便會跟著高。儘管如此，Cronbach's α 依然是最常被拿來使用的內部一致性信度。

第三節 影響信度的因素

一般心理測驗的書籍中最常被提及會影響信度的因素有樣本同質性、題數、題目難度、測驗時間的限制和信度的估計方法(Anastasi & Urbina, 1997; Crocker & Algina, 1986; Gregory, 2007; Murphy & Davidshofer, 1997)。



一、樣本同質性

樣本的同質性會影響信度係數，可將式子(2)帶入式子(3)得到

$$\rho_{XX} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2}, \quad (5)$$

若我們假設 σ_e^2 為定值，則當 σ_T^2 增加時， ρ_{XX} 也會增加。也就是說，如果測量誤差的變異數不變，樣本在所測量的特質上異質性越大，則信度越高。

或者也可以用全距限制(restriction of range)來解釋，在計算相關係數時，若僅選取在所測量的特質上同質性很高的樣本，此時計算出來的相關係數容易偏低。此現象會發生在使用相關係數來估計信度時，如：再測信度。以圖 1 的例子來說明，若考量所有人的時候信度有.90，但如果只選取高分群做為樣本，信度會降到.30。

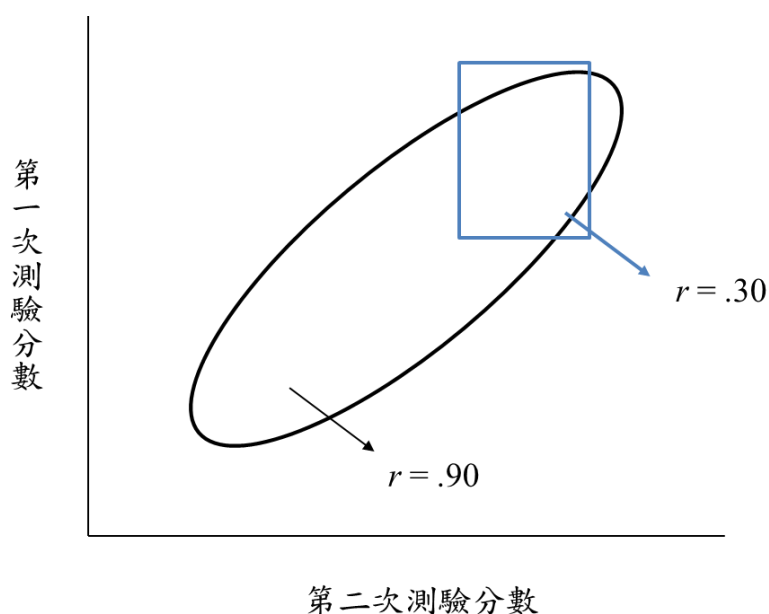


圖 1：全距限制對相關係數的影響

二、題數

假設有 X 和 Y 兩個測驗，若兩測驗為平行測驗，且 X 測驗的題數是 Y 測驗的 k 倍，兩測驗信度的關係可由 Spearman-Brown 公式來表示：

$$\rho_{XX} = \frac{k\rho_{YY}}{1+(k-1)\rho_{YY}}, \quad (6)$$

其中 ρ_{XX} 是 X 測驗的信度係數， ρ_{YY} 是 Y 測驗的信度係數。從式子(6)可以看出當 X 測驗的題數增加時，k 會跟著增加， ρ_{XX} 也會變大。假如有個測驗有 10 題，信度係數為.60，當題目增加到 60 題時，信度係數可以提升到.90（表 1）。

表 1
題數對信度的影響

題數	k	信度係數
10	1	.60
20	2	.75
30	3	.82
40	4	.86
50	5	.88
60	6	.90

三、題目難度

題目難度太低時會有天花板效應(ceiling effect)，使得受試者的分數集中在高分；相反地，若題目難度過高則會產生地板效應(floor effect)，受試者的分數會集中在低分。不管是難度過高或過低都會讓樣本的同質性太高，以至於信度變差。另外，當難度過高時，如果受試者以猜測的方式答題，也會造成誤差變大，信度降低。

四、測驗時間的限制

速度測驗(speed test)有大量難度偏低的題目，有回答幾乎都能答對，不過有作答的時間限制，多數受試者會無法答全部的題目。這類測驗的目的在於

評估受試者是否能夠快速地完成某種作業。此時若將測驗題目分成前後兩半計算折半信度(split-half reliability)，也就是計算前後兩半測驗的相關係數，信度很可能會被低估，因為後面的題目很多會因為寫不完而答錯。若將測驗以奇數題和偶數題分成兩半同樣計算折半信度，這時信度會被高估。因此估計速度測驗的信度時，使用再測信度或複本信度是較為適當的(Crocker & Algina, 1986)。

五、信度的估計方法

如同第二節所述，不同的測量誤差會有不同的估計信度係數的方法，既然是不同的誤差來源，自然地信度可能也會不同。通常來說，內部一致性信度會比再測信度和複本信度高(Murphy & Davidshofer, 1997)。

除了上述五個較常被提及影響信度的因素，其他造成測量誤差的來源也是可能影響信度的因素。在施測的時候主試者會盡可能將施測程序與計分方式標準化，以降低誤差，也會提供舒適的環境，以免因環境因素造成誤差，但受試者本身的狀態所造成的誤差仍然很難避免。另外，估計再測信度時，兩次測驗的時間間隔長短也可能會影響信度，一般來說，時間間隔越長，變數會越多，信度也就越低。

第四節 信度概化法

Crocker 和 Algina(1986)指出「測驗本身不會是『有信度』或『沒有信度的』，確切來說，信度是特定一群受試者在某一個測驗上得分的性質。」許多學者都有相同看法(Gronlund & Linn, 1990; Rowley, 1976; B. Thompson, 1994)。從信度的定義（式子(3)）來看，信度和受試者的測驗分數的變異數有關係，當我們有不同受試者，有不同的測驗分數的變異數，可能會估計出不同的信度係數。也

就是說，同一個測驗工具可能產生有信度的測驗結果，也可能產生信度不佳的測驗結果。因此 Vacha-Haase (1998)提出一種信度的整合分析方法，稱為信度概化法(reliability generalization, RG)，藉由信度概化法我們可以瞭解某個測驗工具在不同研究中產生怎樣的信度結果，以及什麼因素可以預測這些信度係數的變異。

信度概化法是效度概化法(validity generalization, VG)(Schmidt & Hunter, 1977)的延伸，效度概化法假設效度的變異可能是來自於抽樣誤差、測量結果信度的不同、全距限制和其他隨機誤差(Hunter & Schmidt, 1990)，此方法以不同研究中的效度係數當作依變項，研究的特性當成獨變項，使用一般線性模型(general linear model)分析。很類似地，信度概化法則是將依變項改成信度係數，並蒐集可能影響信度係數的因素當成獨變項。一般來說信度概化法的研究程序可分成幾個步驟來說明：

一、選擇測驗

任何能夠估計信度係數測驗都可以作為信度概化研究的對象，但為了有足夠的次級資料進行整合分析，盡量選擇很多研究者使用的測驗，而且只有報告自己研究樣本所估計出的信度係數的研究才可以分析。Vacha-Haase (1998)發現有 13.06%使用 Bem 性別角色量表(Bem Sex Role Inventory, BSRI)的研究有呈現自己的信度係數。Yin 和 Fan(2000)的研究結果則顯示使用貝克憂鬱量表(Beck Depression Inventory, BDI)作為研究工具的研究中，有 7.5%的研究有報告自己的信度係數。這項比例在使用明尼蘇達多相人格測驗(Minnesota Multiphasic Personality Inventory, MMPI)的研究中也只有 7.8%(Vacha-Haase, Kogan, Tani, & Woodall, 2001)。因此更需要選擇已有相當程度使用頻率的測驗來作為研究對象，以免最後分析資料時，可分析的論文數目過小。



二、確立編碼簿

選定測驗後，需發展一份編碼簿來記錄所需要的資訊。除了最重要的信度係數外，各項研究的基本特徵則被當成是可能影響信度係數的變項，例如：樣本數、樣本類型、題數、信度係數的種類、測驗分數的標準差、參與者性別、年齡、種族等等。

三、蒐集資料

在資料庫中搜尋有使用事前選定的測驗的研究，可以變換不同的關鍵字去搜尋，以免遺漏任何可以作為研究對象的論文，並透過任何方式取得全文後，一一的閱讀每篇研究，並將所有研究報告分成三類：

- (一) 文章裡沒有提供信度資訊。
- (二) 文章裡只有引用他人的信度值。
- (三) 文章裡有提供自己蒐集的研究樣本的信度值。

只有第三類的研究會被信度概化研究採用，才需要依據編碼簿去紀錄所需的資訊。

在此階段有需要的話可以訓練多名編碼人員來編碼，編碼人員間的一致性在整合分析中很重要，因此需要不停來回討論以確保編碼者間信度。

四、分析資料

將信度係數視為依變項，研究的基本特徵視為獨變項，有許多統計方法都可以用來探討它們之間的關係，例如：相關分析、變異數分析(ANOVA)、典型相關分析(canonical correlation analysis)、迴歸分析、階層線性模式(hierarchical linear model, HLM)等等。如同 Thompson 和 Vacha-Haase(2000)所說的：「信度概化法不是單指一種分析方法」，不會侷限使用任何一種統計分析方法，研究者可自行選擇適當的分析方法。

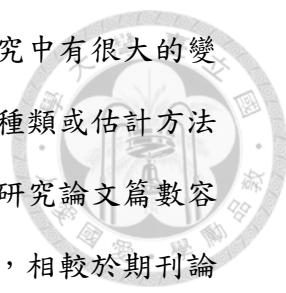
第五節 信度概化法的相關研究



自從 Vacha-Haase 於 1998 年提出信度概化法後，許多研究者紛紛使用此方法針對不同問卷進行信度概化研究。這些研究多發表在 *Educational and Psychological Measurement* 期刊，只有少數發表在 *Psychological Methods* 和 *Assessment*..... 等等期刊。這些研究使用的問卷、變項、統計方法和結果整理於表 2。

從這些信度概化研究可以發現有提供自己資料的信度係數的比例介於 5.6%-73.2% 之間，且大多沒有超過 40%。拿來當依變項的信度係數以 Cronbach's α 為最多。常使用的獨變項中有和研究參與者相關的因素，包括性別、樣本數、年齡、樣本類型、種族和測驗分數的標準差，另外也有和測驗工具本身有關的特性，如：問卷版本和問卷題數。使用的統計方法以相關和迴歸分析為大宗，其他分析方法（如：變異數分析、典型相關分析、階層線性模式）也有人使用。由於信度係數多為負偏的分配，有的研究者會對信度係數進行轉換，以符合統計方法常有的常態假設，最常使用的轉換方式為 Fisher's Z transforms。

回顧過往的信度概化研究，發現由於整合分析的研究對象受限於已有的文獻，為了增加研究論文的篇數，許多信度概化研究的依變項混合了不同類型的信度係數（例如：再測信度、內部一致性信度）(Beretvas & Pastor, 2003; Vacha-Haase, 1998; Vacha-Haase, Tani, Kogan, Woodall, & Thompson, 2001; Yin & Fan, 2000)，或是不同的信度估計方法（例如：Cronbach's α 、折半信度）(Beretvas, Meyers, & Leite, 2002; Reese, Kieffer, & Briggs, 2002; Yin & Fan, 2000)。獨變項的部分除了有和研究參與者相關的因素之外（例如：性別、年齡.....），也包含許多和測量工具本身有關的因素（例如：題目數、語言和量尺點數）(Botella, Suero, & Gambará, 2010; Caruso, Witkiewitz, Belcourt-Dittloff, & Gottlieb, 2001; Gleaves, Pearson, Ambwani, & Morey, 2014; Vacha-Haase, 1998)。但這樣的研究設



計會讓研究問題變得很複雜，若研究結果觀察到信度在不同研究中有很大的變異，除了因為是不同的研究樣本之外，也有可能是出於信度的種類或估計方法不同，還可能是因為測驗工具在不同研究中有差異。為了解決研究論文篇數容易偏小的問題，本研究使用台灣博碩士學位論文當作研究對象，相較於期刊論文，博碩士學位論文篇幅限制較寬鬆，可以提供更多關於樣本特質的描述，以及信度的資訊，也就能提高可分析的論文篇數。有更多資料可以使用時，便可將本研究的目標聚焦到（一）討論題目、量尺等等各方面完全一樣的測量工具在不同研究中，不同群體的受試者產生的 Cronbach's α 有何變異。（二）這樣的變異可以用哪些和研究參與者有關的因素（例如：參與者性別、年齡）來預測。



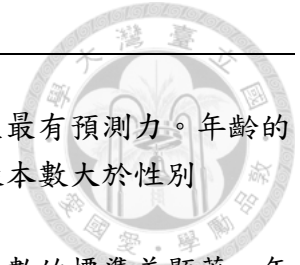
表 2
信度概化法的相關研究整理

作者 (出版年)	問卷	有自己 信度的 篇數 (%)	依變項	獨變項	統計方法	結果
Vacha-Haase (1998)	BSRI	82 (13.1%)	Cronbach's α 、 KR、再測信度	樣本數、信度種類、填答者、 題數、性別、研究類型、語 言、樣本類型、量尺點數	迴歸、典型相 關分析	根據迴歸和典型相關分析的結 果，信度種類影響最大，樣本數 和題目長度對 F 分量表的影響較 M 分量表大
Caruso (2000)	NEO	53 (22.3%)	Cronbach's α 、再 測信度	填答者、量表版本、語言、 樣本類型、性別、信度種類、 樣本數、平均年齡	相關、 ANOVA、 Fisher's z transforms	語言顯著(NEO)；樣本類型顯著 (A)，臨床和其他三組不同；性別 顯著(C)，全男和其他兩組不同； 信度種類顯著(A)；量表版本顯著 (NEOAC)；校正題目長度後，量 表版本顯著(NEAC)
Viswesvaran and Ones (2000)	所有測量 性格的工 具(使用手 冊)		Cronbach's α 、再 測信度(分開分 析)	無	描述統計、信 賴區間	分別比較五個因素的 Cronbach's α 和再測信度，沒有差異

Yin and Fan (2000)	BDI	90 (7.5%)	內部一致性信度(Cronbach's α 、KR、折半信度)、再測信度，有兩個合在一起分析，也有分開分析。SEM	題數、語言、年齡、性別、種族、樣本類型 1(學生/非學生)、樣本類型 2(精神病患/非精神病患)、樣本類型 3(疾病/正常)、樣本類型 4(成癮/沒有成癮)、信度種類	GLM、相關	合在一起：學生/非學生、疾病/正常、成癮/沒有成癮、信度種類(影響最大)顯著； α ：年齡、學生/非學生、成癮/沒有成癮顯著；再測：題目長度、疾病/正常、成癮/沒有成癮顯著
Vacha-Haase, Kogan, et al. (2001)	MMPI	153 (7.8%)	內部一致性信度、再測信度	年齡、樣本類型、版本	迴歸	綜合十個分量表來看，年齡和版本有較大的影響
Henson, Kogan, and Vacha-Haase (2001)	TES、 STEBI、 TLC-RSA		Cronbach's α 、 KR	教師經驗、任教單位、教學類型、性別多數的比例、樣本數、題數、分數的標準差、平均分數	相關	教師經驗、任教單位、性別多數的比例分別和信度有負相關，分數的標準差、題目數和信度有正相關
Caruso et al. (2001)	EPQ	76 (6.2%)	內部一致性信度	題數、分數的標準差、平均分數、平均年齡、年齡的標準差、樣本類型、男性比例、語言、版本	階層式迴歸	每個分量表都發現分數變異數越大，信度越大；P 分量表：年齡的標準差和樣本類型也顯著。
Vacha-Haase, Tani, et al. (2001)	MMPI	153 (7.8%)	Cronbach's α 、再測信度。也有另外分析 SEM	樣本數、信度種類、年齡 1(成人/非成人)、年齡 2(大學生/非大學生)、性別、版本、樣本類型 1(犯人/非犯人)、樣本類型 2(臨床/非臨床)、研究類型、語言	迴歸	信度種類對 L 和 F 分量表都有預測力，年齡 2 在三個分量表都有預測力，樣本類型 2 和研究類型對 F 分量表影響較大，對 L 和 K 較不影響。



Beretvas et al. (2002)	MCSDS	93 (8.7%)	內部一致性信度(Cronbach's α 、KR、折半信度)、再測信度(兩個分開分析)	年齡、題數、男性比例	Fisher's z transformation、HLM、迴歸	女生、題目多內部一致性信度大
Capraro and Capraro (2002)	MBTI	14 (6.7%)	Cronbach's α 、再測信度	無	描述統計	分析的篇數太少，無法做迴歸分析，所以只呈現分量表信度
Barnes, Harp, and Jung (2002)	STAI	46 (5.6%)	Cronbach's α 、再測信度	版本、引發焦慮的情境、研究領域、研究設計、年齡、信度種類、樣本數、平均分數、分數的標準差	t 檢定、相關	年齡(大於 16 歲)、版本、分數標準差最有影響
Reese et al. (2002)	AAS、BORI、IPPA、PAQ、PBI	53 (34.4%)	內部一致性信度(Cronbach's α 、KR、折半信度)、再測信度	版本、量尺、填答者、語言、樣本數、女性比例、白人比例、年齡、樣本類型、信度種類、分數的標準差	相關	信度和研究特性沒有明顯關係
Nilsson, Schmidt, and Meek (2002)	CDMSE	20 (40.8%)	Cronbach's α 、再測信度	樣本數、平均分數、分數的標準差、年齡、女性比例、白人比例	相關	年齡、分數標準差、白人比例和信度係數的相關高
Hanson, Curry, and Bandalos (2002)	WAI、WAI-S	23 (38.3%)	內部一致性信度、評分者間信度	病人樣本數、治療師樣本數、病人性別多數、治療師性別多數	相關	病人和治療師樣本數和信度的相關都顯著，但 n 太小，不見得有意義



Kieffer and Reese (2002)	GDS (4 個版本)	98 (29.0%)	內部一致性信度(Cronbach's α 、KR、折半信度)、再測信度	題數、量尺、填答者、語言、樣本數、女性比例、年齡、樣本類型、信度種類、平均分數、分數的標準差	迴歸	分數標準差最有預測力。年齡的影響大於樣本數大於性別
Shields and Caruso (2003)	AUDIT	23 (22.1%)	Cronbach's α	平均年齡、男性比例、樣本類型、分數的標準差	階層式迴歸、相關	迴歸只有分數的標準差顯著，年齡和信度有顯著負相關
Beretvas and Pastor (2003)	BDI		內部一致性信度、再測信度(分開分析)	題數、年齡、男性比例、學生比例、病人比例、信度種類、再測間隔	HLM	內部一致性：學生比例和年齡顯著(分析的樣本數=106)。再測：不顯著(分析的樣本數=21) 社經地位在最多分量表上相關
Youngstrom and Green (2003)	DES		Cronbach's α	女性比例、平均年齡、白人比例、社經地位	相關、迴歸	顯著，平均年齡和白人比例也有幾個顯著。迴歸結果社經地位影響最大
Shields and Caruso (2004)	CAGE	19 (7.3%)	Cronbach's α	分數的標準差、平均年齡、男性比例、高加索人比例、樣本類型	相關	平均年齡、分數標準差和信度的相關顯著
O'rourke (2004)	CES-D		Cronbach's α	樣本數、平均年齡、年齡的標準差、招募方式、照顧接受者類型、關係、居住、性別、語言、題數	迴歸、Fisher's z transforms	照顧接受者類型、關係、題數顯著迴歸係數，其中題數影響最大

Ryngala, Shields, and Caruso (2005)	RCMAS		Cronbach's α	性別、種族、年齡、平均分數、分數的標準差、樣本數	相關、階層式迴歸、Fisher's z transforms	每個分量表都發現分數變異數越大， α 越大；年齡、種族、平均分數在部分分量表迴歸係數也有顯著。
Ross, Blackburn, and Forbes (2005)	PALS	30 (73.2%)	Cronbach's α	版本、樣本數、改編測驗、分量表、 平均年齡、男性比例、白人比例、國籍、樣本類型、高於切截分數的比例、平均分數、分數的標準差、版本、來源、出版類型	迴歸、相關、ANOVA 單迴歸、ANOVA、Fisher's z transforms	版本和分量表有影響 α ：分數標準差和版本影響最大；ICC：分數標準差和來源影響最大
Campbell, Pulos, Hogan, and Murry (2005)	PCLs		Cronbach's α 、 ICC(分開分析)			
Vassar and Crosby (2008)	UCLA 寂寞量表	80 (37.6%)	Cronbach's α	女性比例、樣本數、平均年齡、分數的標準差、研究類型、語言、學生、青少年、老人、社會支持、量尺點數	相關、迴歸	UCLA-R：分數標準差、社會支持、研究類型、青少年相關顯著；分數標準差、研究類型迴歸係數顯著。UCLA-3 篇數太少，所以沒有做相關和迴歸分析

Graham and Christiansen (2009)	Rubin's Loving and Liking scale、LAS、LAS-SF、PLS、TLS	145 (54.9%)	Cronbach's α	樣本數、白人比例、男性比例、已婚比例、同居比例、異性戀比例、平均年齡、關係長度、大學生比例	相關、迴歸、Fisher's z transforms	白人比例的迴歸係數在最多分量表顯著，男性比例次之(分析樣本數 10-19)。平均年齡在一些分量表上有顯著負相關，白人比例、男性比例、關係長度、同居比例、已婚比例正負相關都有(分析的樣本數 3-57)
Botella et al. (2010)	STAI		Cronbach's α	平均年齡、年齡、女性比例、語言、樣本類型、分數的變異數	迴歸、 Q 檢定、 T transformed	變異數越大， α 越大；大於 65 歲或西班牙文的 α 較低
Bornmann, Mutz, and Daniel (2010)			評分者間信度 (ICC、皮爾森相關)	手稿數、信度估計方法、學門、文章類型、年代、盲評、評分系統	HLM	手稿數和評分系統顯著
Mahapoonyanont, Krahamwong, Kochakornjarupong, and Rachasong (2010)	各種測量批判性思考的工具	14 (56.0%)	Cronbach's α	科目、高中學生、樣本數、嘗試次數、題數、選擇題量尺	相關、迴歸、Fisher's z transforms	選擇題量尺有影響

Aguayo, Pecino, de la Fuente Solana, and Fernández (2011)	MBI	74 (26.1%)	Cronbach's α	平均年齡、年齡的標準差、男性比例、樣本類型、版本、樣本數、改編測驗、語言、平均分數、分數的標準差、抽樣方式、研究類型、出版類型、施測國家	ANOVA、迴歸、 T transformed	EE 分量表：平均年齡和分數標準差單迴歸顯著；版本、改編測驗、語言、抽樣、施測國家 ANOVA 顯著；多元迴歸用分數標準差、抽樣、施測國家。D 分量表：改編測驗、語言、施測國家 ANOVA 顯著；多元迴歸用分數標準差、語言、改編測驗。總分：年齡顯著；body 分量表性別顯著；drive 分量表性別、年齡顯著；bulimia 分量表性別、年齡、樣本類型顯著 EAT-40 總分：性別、年齡顯著；EAT-26 總分：樣本類型顯著；EAT-26 dieting 分量表年齡顯著；EAT-26 bulimia 分量表年齡顯著；EAT-26 oral 分量表語言顯著
Gleaves et al. (2014)	EDI	74 (26.1%)	Cronbach's α 、再測信度(分開分析)	樣本類型、年齡、性別、語言	信賴區間	EAT-40 總分：性別、年齡顯著；EAT-26 總分：樣本類型顯著；EAT-26 dieting 分量表年齡顯著；EAT-26 bulimia 分量表年齡顯著；EAT-26 oral 分量表語言顯著
同上(兩個問卷分開分析)	EAT	78 (36.3%)	Cronbach's α 、再測信度(分開分析)	樣本類型、年齡、性別、語言	信賴區間	EAT-40 總分：性別、年齡顯著；EAT-26 總分：樣本類型顯著；EAT-26 dieting 分量表年齡顯著；EAT-26 bulimia 分量表年齡顯著；EAT-26 oral 分量表語言顯著

第二章 研究方法



第一節 研究對象

本研究蒐集了自 2000 年起到 2015 年 8 月所有使用台灣簡明版世界衛生組織生活品質問卷（台灣版 WHOQOL-BREF）作為研究工具的台灣博碩士學位論文，研究對象另外須符合下列條件：（一）對問卷內容沒有做出任何更動（例如：刪題、更改量尺語詞或題目敘述）。（二）研究需要使用 Cronbach's α （例如：以項目反應理論計算信度的研究就不需要）。

如同先前的提到的，選擇測驗作為信度概化研究的對象時，應選擇許多研究者使用過的測驗。在台灣要使用台灣版 WHOQOL-BREF 須向台灣大學心理系的姚開屏教授取得授權，根據姚開屏教授所提供的資料顯示，截至 2016 年 6 月 2 日止，已有 1342 位研究者提出申請，使用的單位以各級學校最多(60%)，醫院診所次之(34%)，其他單位如政府機關、社會福利機構和公司行號也都有使用（圖 2）。其中各級學校的研究多為博碩士論文，有些醫院的研究也是博碩士論文。研究的部門涵蓋範圍非常廣，以護理相關占最多數(25%)（圖 3）。由此可見，選擇台灣版 WHOQOL-BREF 作為本研究的對象可說是相當適合。而且過去沒有對生活品質問卷進行過信度概化研究，但生活品質是一個廣泛被探討的概念，因此有必要對其進行信度概化研究。以下就台灣版 WHOQOL-BREF 簡略介紹：

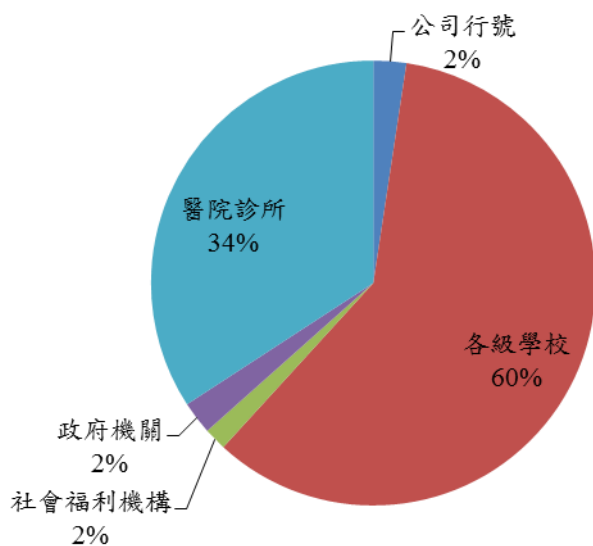


圖 2：台灣版 WHOQOL-BREF 的使用單位

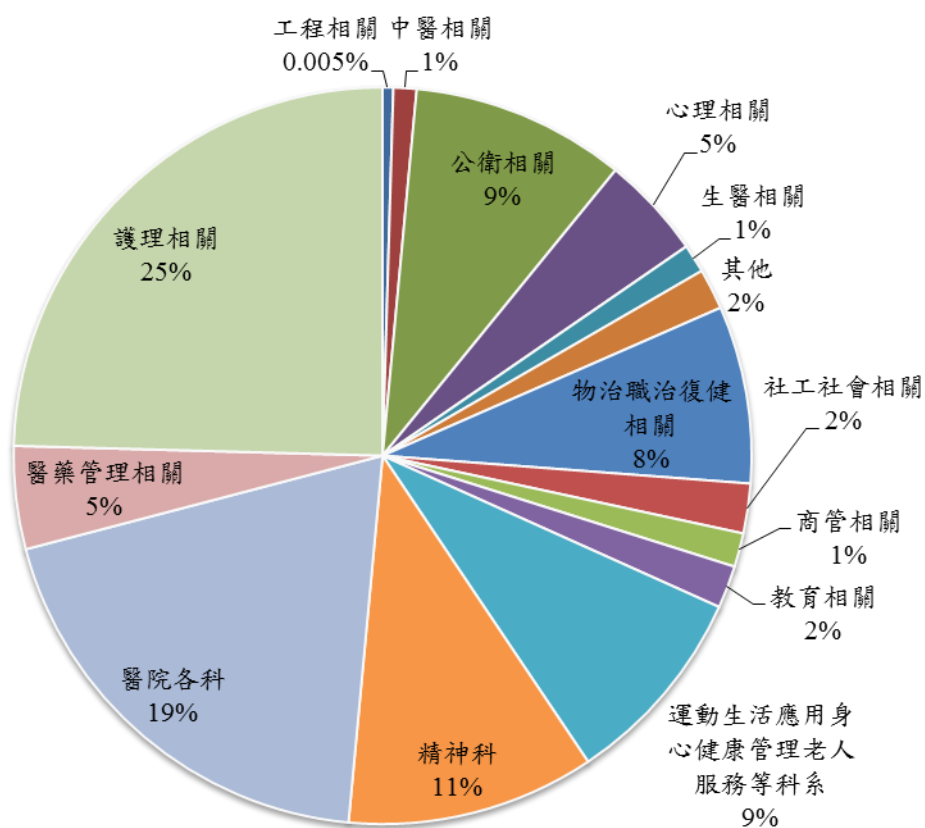


圖 3：台灣版 WHOQOL-BREF 的研究部門

台灣簡明版世界衛生組織生活品質問卷 (台灣版 WHOQOL-BREF)

世界衛生組織(World Health Organization, WHO)為發展一份可供跨文化比較研究的生活品質測量工具，於 1991 年開始生活品質研究計畫，在 1995 年發展出一份健康相關生活品質問卷，即「世界衛生組織生活品質問卷(WHOQOL-100)」。

但這份問卷過長，於是 WHOQOL 研究總部將其簡化成 26 題，稱為「簡明版世界衛生組織生活品質問卷(WHOQOL-BREF)」。

WHOQOL-BREF 分為四大範疇(domains):生理健康範疇(physical domain)7 題、心理範疇(psychological domain)6 題、社會關係範疇(social relationships domain)3 題，和環境範疇(environment domain)8 題。另外再加上一題測量整體生活品質，以及一題測量一般健康的題目。

WHOQOL 台灣版研究小組於 1997 年向 WHOQOL 瑞士日內瓦研究總部取得台灣發展權後，將 WHOQOL-100 翻譯完成，並依照總部的建議加入符合本國文化的本土性題目。至於台灣版 WHOQOL-BREF 是除了原本的 26 題外，在社會關係和環境範疇各加上一題本土題，總共 28 題，題目內容整理於表 3。研究小組在全台灣 16 家區域級以上的醫院以及一家大型診所蒐集了 1068 份有效問卷，有不同疾病的病患和一般健康人，完成信效度研究。台灣版 WHOQOL-BREF 整體問卷的 Cronbach's α 達.91，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(台灣版世界衛生組織生活品質問卷發展小組，2005)。

問卷採自填(self-administered)方式進行，必要時可採用訪談者協助(interviewer-assisted)或訪談者填寫(interviewer-administered)的方式進行，受訪者需為年滿 18 歲的成年人，評估自己最近兩星期的主觀生活品質。每一題都是五點量尺，反向題需反向計分，每題得分最低 1 分，最高 5 分，分數越高表示生活品質越好。各個範疇分數的計分方式為將該範疇內所有題目的分數相加除以題數後成以四，意即計算該範疇平均得分後乘以四得到範疇分數，範圍介在四到二十分之間。另外，WHOQOL 研究總部規定，28 題中若有六題或六題以上

的遺失值，則該受訪者的資料不予採納。生理健康、心理、社會關係範疇最多只能有一題遺失值，環境範疇最多兩題遺失值，遺失值以該範疇內其他題目的平均數取代，但若超過規定的題目數目，該範疇分數不被計算。



表 3

台灣版 WHOQOL-BREF 的題目與所屬範疇

範疇	題號	題目
	1	整體來說您如何評價您的生活品質
	2	整體來說您滿意自己的健康嗎
生理健康	3	您覺得身體疼痛會妨礙您處理需要做的事情嗎*
生理健康	4	您需要靠醫療的幫助應付日常生活嗎*
生理健康	10	您每天的生活有足夠的精力嗎
生理健康	15	您四處行動的能力好嗎
生理健康	16	您滿意自己的睡眠狀況嗎
生理健康	17	您對自己從事日常生活的能力滿意嗎
生理健康	18	您滿意自己的工作能力嗎
心理	5	您享受生活嗎
心理	6	您覺得自己的生命有意義嗎
心理	7	您集中精神的能力有多好
心理	11	您能接受自己的外表嗎
心理	19	您對自己滿意嗎
心理	26	您常有負面的感受嗎（如傷心、緊張、焦慮、憂鬱等）*
社會關係	20	您滿意自己的人際關係嗎
社會關係	21	您滿意自己的性生活嗎
社會關係	22	您滿意朋友給您的支持嗎
社會關係	27	您覺得自己有面子或被尊重嗎
環境	8	在日常生活中您感到安全嗎
環境	9	您所處的環境健康嗎（如污染、噪音、氣候、景觀）
環境	12	您有足夠的金錢應付所需嗎
環境	13	您能方便得到每日生活所需的資訊嗎
環境	14	您有機會從事休閒活動嗎
環境	23	您滿意自己住所的狀況嗎
環境	24	您對醫療保健服務的方便程度滿意嗎
環境	25	您滿意所使用的交通運輸方式嗎
環境	28	您想吃的食物通常都能吃到嗎

註：第 27 和 28 題為本土題。*為反向題。



第二節 資料蒐集



本研究使用國家圖書館開發的「臺灣博碩士論文知識加值系統」，以及「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」作為關鍵字搜尋，在華藝線上圖書館的搜尋欄位有篇名、作者、關鍵字和摘要，臺灣博碩士論文知識加值系統除了提供這四個欄位可以搜尋外，還多包含論文目次和論文參考文獻兩個欄位。去除兩個資料庫重複的部分，總共有 1332 個搜尋結果。其中 170 篇論文既沒有電子全文，在國家圖書館或台大的圖書館裡也沒有紙本論文，故不納入本研究。剩下的 1162 篇論文取得全文後一一讀過，發現 612 篇並沒有使用台灣版 WHOQOL-BREF 作為研究工具，而是在參考文獻中出現，所以包含在本研究的搜尋結果裡。另外剔除 36 篇不符合本研究對象的條件的論文，即對量表有更動，或是研究不需要使用 Cronbach's α 的論文。最後剩下 514 篇使用台灣版 WHOQOL-BREF 的論文納入本研究中（圖 4）。

514 篇論文可以分成四個類別，第一類是完全沒有提到任何關於信度的資訊，有 56 篇(10.9%)；第二類僅引用問卷使用手冊中的信度係數，或是過去其他研究的信度結果，有 206 篇(40.1%)；第三類有 4 篇(0.8%)，這些研究雖然有計算自己資料的內部一致性信度，但把四個範疇的結果合起來報告一個數值範圍（例如：四個範疇的 Cronbach's α 介於 0.XX-0.XX 間），以致於本研究無法更深入的分析；剩下 248 篇(48.2%)屬第四類，用自己的資料計算 Cronbach's α ，且報告出確切的數值。只有第四類的論文符合本研究目的，將會納入後續的分析中。

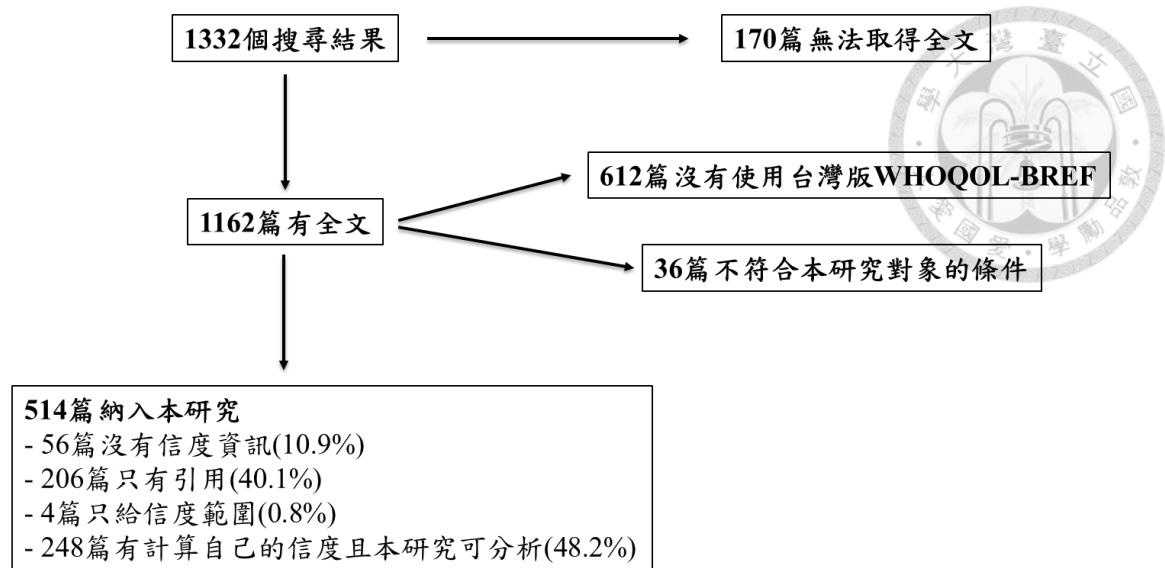


圖 4：資料蒐集流程圖

有些研究除了有報告正式研究的信度，在預試階段也有做信度檢定，或是實驗組和對照組會分開來計算信度，因此一篇論文可能有不止一個 Cronbach's α 。248 篇共產生 291 個整份問卷的 Cronbach's α ，生理健康、心理、社會關係和環境範疇的 Cronbach's α 分別有 259、268、254、258 個。在後面的統計分析中，因每篇論文呈現的資訊完整度不同，不同分析中的 Cronbach's α 個數會有所差異。

第三節 研究變項說明

一、依變項

Cronbach's α

本研究依變項分成整份問卷的 Cronbach's α ，以及生理健康、心理、社會關係和環境四個範疇的 Cronbach's α 。由於在編碼的過程中發現，有些研究會用完整的 28 題去計算整份問卷的 Cronbach's α ，也有的研究會去掉兩題測量整體生

活品質和一般健康狀態的題目，用四範疇的 26 題去計算整份問卷的 Cronbach's α 。考量到 26 題和 28 題的所計算出的 Cronbach's α 應該不會相差太多，所以不論是 26 或 28 題皆採納。但四個範疇的 Cronbach's α 由於題目數目沒那麼多，若少一題對 Cronbach's α 的影響較大，所以必須包含所有題目才能納入本研究中，其中社會關係和環境範疇須含有本土題，也就是生理健康範疇要有完整的 7 題，心理範疇 6 題，社會關係範疇 4 題，環境範疇 9 題。

編碼過程中發現，台灣版 WHOQOL-BREF 的第 21 題「您滿意自己的性生活嗎」是最常被拿掉的題目，研究者常因為認為這個題目和研究目的無關而刪除這題，假如是這種狀況，社會關係範疇的 Cronbach's α 就無法納入本研究，但其他範疇不受影響，還是可以納入，整份問卷的 Cronbach's α 也因為 27 題和 28 題結果應該不會差距太大，所以仍然可以納入本研究中。

二、獨變項

參考過去其他信度概化研究使用過的變項，加上台灣博碩士論文中常報告的人口學變項，本研究蒐集了十一個和研究參與者有關的因素作為獨變項。有些變項雖然在過去的信度概化研究常使用，但在台灣並不適用（例如：種族），所以無法放入本研究中。教育程度、宗教和婚姻狀態這三個變項雖然過去沒有信度概化研究使用過，但許多台灣博碩士論文都有提供這些資訊，因此會當成本研究的獨變項。以下分別說明十一個獨變項在本研究中的定義：

（一）平均分數和分數的標準差

生活品質得分的平均數和標準差，這兩個變項和 Cronbach's α 一樣分成整份問卷，以及生理、心理、社會關係和環境四個範疇。四範疇的平均分數採用問卷使用手冊中範疇分數的計分方式，範圍為四到二十分，標準差為範疇分數的標準差。整份問卷的得分較少研究有提供，所以本研究計算整份問卷的平均

分數時，是將四個範疇分數相加，範圍介在十六到八十分，並計算其標準差。雖然兩題測量整體生活品質和一般健康狀態的題目會因此被排除在外，但為了讓本研究可使用的論文篇數能多一些，只能使用這個折衷的方法。



(二) 樣本數

各個研究中的樣本數，也就是 Cronbach's α 是根據多少筆資料計算出來。

(三) 平均年齡和年齡的標準差

參與者年齡的平均數和標準差。

(四) 性別

本研究採用男性比例，即男性人數除以總人數，介於 0-1。過去有些信度概化研究的性別變項有些是採用男性（女性）的比例(Botella et al., 2010; Caruso et al., 2001; Shields & Caruso, 2003, 2004)；有些則是採用類別變項，分成全男性樣本、全女性樣本或混合樣本(Caruso, 2000; Gleaves et al., 2014; Vacha-Haase, 1998; Yin & Fan, 2000)。由於本研究蒐集到的論文多為混合樣本（約 85%，不同範疇結果略有不同），故使用前者的方式。

(五) 教育程度

在編碼階段時，教育程度分為不識字、未受教育但識字、小學畢（肄）業、國（初）中（肄）業、高中職畢（肄）業、大學或專科畢（肄）業、碩士以上或未知（遺失值），然後記錄下每篇論文參與者的教育程度在這八個分類中組成的比例，但不是每篇論文都有這八個分類，有時會需要把八類中某幾類合併在一起，或是需要把論文裡的類別合併放入本研究八類中的某一類。以表 4 中的例子說明，編碼簿中的不識字、未受教育但識字和小學畢（肄）業這三個分類都被歸在原始論文中的「國小以下」一類，編碼時以顏色標記，將底色標上同

個顏色，代表這三個分類總共占了 5%。而原始論文中的「碩士」、「博士」在編碼簿中都算是碩士以上，因此把 11.1%加上 2.1%，碩士以上總共占了 13.2%。

由於每個論文在教育程度這個變項上使用的分類不盡相同，為了保留較多篇論文，所以本研究分析時將教育程度此變項定義為大專院校（含）以上畢業或肄業的比例，數值介於 0-1。

表 4

教育程度編碼例子

原始論文資料			編碼後	
教育程度	人數	百分比	教育程度	比例
國小以下	14	5.0	不識字	
國中	17	6.1	未受教育但識字	
高中職	61	21.8	小學畢（肄）業	.05
大學或大專	151	53.9	國（初）中（肄）業	.061
碩士	31	11.1	高中職畢（肄）業	.218
博士	6	2.1	大學或專科畢（肄）業	.539
			碩士以上	.132
			未知（遺失值）	.0

資料來源：劉誥洋（2006）。Spa 水療對憂鬱情緒及生活品質之影響—以台南地區 Spa 消費者為例。立德管理學院休閒管理研究所，台南市。

（六）樣本類型

類別變項，1 為研究樣本全部都是病人的研究，0 為非全部是病人的研究，可能是學生、一般民眾等等其他類型的樣本，或是混合了病人與非病人。與研究目的相關的疾病才算是病人。



(七) 宗教

參與者有宗教信仰的比例，介於 0-1。由於每篇論文中包含的宗教信仰不完全相同，所以本研究編碼時僅記錄有宗教信仰和無宗教信仰的比例。

(八) 婚姻狀態

參與者已婚的比例，介於 0-1。編碼時婚姻狀態分成已婚、沒結過婚、離婚、喪偶、未知五類，分別記錄下每篇論文參與者在五個類別的比例，但有許多論文在婚姻狀態此變項上只分成已婚和未婚兩類，其中未婚會包含沒結過婚、離婚、喪偶三種情形，因此本研究在分析時將婚姻狀態定義成已婚的比例。

(九) 地區

類別變項。參與者的居住地區，分成北部、中部、南部、東部、其他共五組。若研究樣本為某學校學生，或是從某醫院選取出來，則以該學校或醫院的地區為準。北部包含臺北市、新北市、基隆市、宜蘭縣、桃園市、新竹縣、新竹市；中部包含苗栗縣、臺中市、彰化縣、南投縣、雲林縣；南部包含嘉義縣、嘉義市、臺南市、高雄市、屏東縣、澎湖縣；東部包含花蓮縣和臺東縣；其他則是混合了兩個以上的地區。

第四節 資料分析

本研究使用統計軟體 SAS 9.4 進行資料分析，部分描述性統計分析以 Microsoft Excel 2010 輔助作圖。

一、使用折線圖檢視台灣版 WHOQOL-BREF 的使用狀況和信度報告情形隨畢業學年度的變化；以及使用長條圖比較在不同學門中台灣版 WHOQOL-BREF 的使用狀況和信度報告情形。

二、以描述性統計初步檢視 Cronbach's α 在不同研究中的變異情形，並且呈現本研究獨變項的描述性統計。

三、如同第一章第三節所述，樣本同質性會影響信度係數，過去的信度概化研究也有相同發現(Botella et al., 2010; Henson et al., 2001; Shields & Caruso, 2003)，因此本研究將以迴歸分析探討在本研究中分數的標準差對 Cronbach's α 是否有影響，還有其他和研究參與者有關的因素在去除分數的標準差影響後，是否能預測 Cronbach's α 。整份問卷、生理健康、心理、社會關係以及環境範疇的 Cronbach's α 將會分開來分析。

(一) 先挑出有報告所有本研究使用的獨變項的論文。

(二) 以分數的標準差預測 Cronbach's α ，進行初步的殘差分析，確定殘差符合常態分配和變異數同質的假設。

1. 用 Kolmogorov-Smirnov 檢定考驗殘差常態性。

2. 計算殘差的絕對值和預測值(predicted value)的 Spearman 相關，以其顯著性檢定來判斷是否符合變異數同質假設。

3. 若不符合假設，則刪除極端值後再次檢驗。

(三) 計算每個獨變項去除掉分數的標準差的影響後，和 Cronbach's α 的淨相關，淨相關達顯著則考慮放進迴歸模型中。其中地區變項是虛擬變項(dummy variable)，所以不能計算淨相關，改以比較完整模型(full model)和簡化模型(reduced model)的淨 F 檢定(partial F test)來決定地區變項是否需要放進模型中。事實上，這兩種統計方法在計算上是等價的。

(四) 上個步驟中若選出不止一個獨變項，則繼續以淨相關或淨 F 檢定的方式篩選出最後模型。

(五) 決定最後模型後再次檢查殘差假設。

四、迴歸分析可以同時考量所有獨變項，但也因為遺失值而刪掉許多篇論文。

因此本研究也會去分析 Cronbach's α 和某一獨變項的兩兩關係。獨變項若

為連續變項便使用皮爾森相關分析（平均分數、分數的標準差、樣本數、平均年齡、年齡的標準差、性別、教育程度、宗教和婚姻狀態）；若為類別變項則使用 t 檢定（樣本類型）或變異數分析(ANOVA)（地區）。



第三章 研究結果



第一節 信度係數報告的情形

本研究共取得 514 篇使用台灣版 WHOQOL-BREF 的台灣博碩士論文全文，由圖 5 可以發現使用台灣版 WHOQOL-BREF 的研究者有逐年上升的趨勢，由於本研究資料只蒐集到 2015 年 8 月，約 103 學年度結束，101 到 103 學年度這三年間的論文明顯有許多尚未公開全文，因此圖中這三年的論文數應該較不準確。

將這 514 篇論文依信度係數報告的情形分成四類：(一) 沒有報告信度的資訊。(二) 只有引用問卷使用手冊中或過去其他研究的信度結果。(三) 有用自己的資料計算 Cronbach's α ，但把四個範疇的結果合起來報告一個數值範圍。(四) 有用自己的資料計算 Cronbach's α ，且報告出確切的數值。結果只有 48.2% (248 篇) 屬於第四類。若將第四類占的比例依畢業學年度分開會發現這項比例逐年下降 (圖 6)。

本研究參考「臺灣博碩士論文知識加值系統」，以及由教育部建置的網站「全國大專校院一覽表線上查詢系統」將 514 篇論文依據系所分成 16 個學門 (表 5)，其中許多學門的論文數量只有個位數，因此把它們合併到「其他學門」中，並同樣依信度報告情形分成四類 (圖 7)。可發現台灣版 WHOQOL-BREF 多用在醫藥衛生學門的研究，56.6% 的論文都來自醫藥衛生學門，其中護理學研究所就占了將近一半 ($135 / 291 = 46.4\%$)。本研究真正關注且會進行相關、變異數、迴歸等分析的第四類論文也是以醫藥衛生學門的研究占最多數。將每個學門分開來看，發現有報告自己的 Cronbach's α 確切數值的論文比例 ($\frac{\text{該學門第四類論文數}}{\text{該學門論文總數}}$)

以民生學門最高(61.5%)，社會服務學門次之(56.3%)，商業及管理學門和醫藥衛生學門也相當接近(53.6%和 51.2%)，最低的則是社會及行為科學學門(26.5%)

(圖 8)。

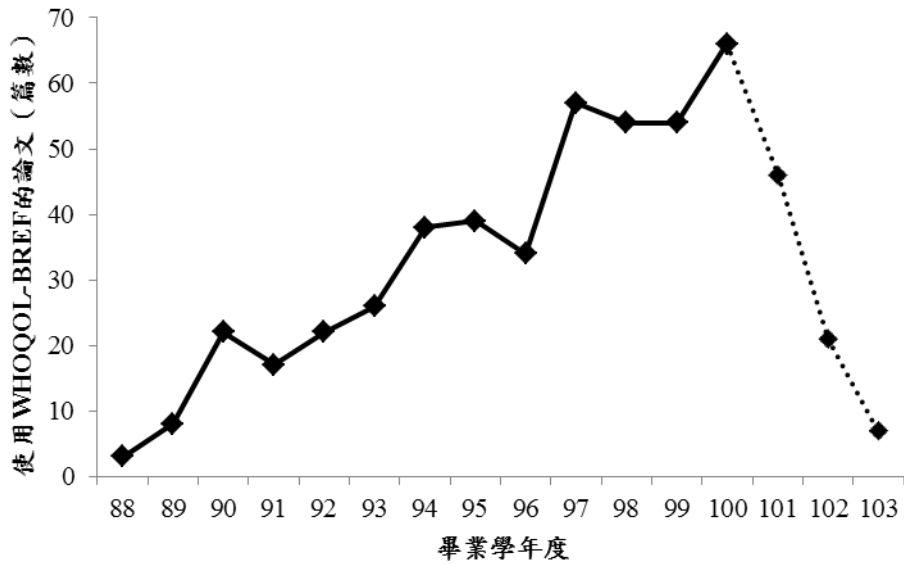


圖 5：使用台灣版 WHOQOL-BREF 的論文數量隨畢業學年度的變化

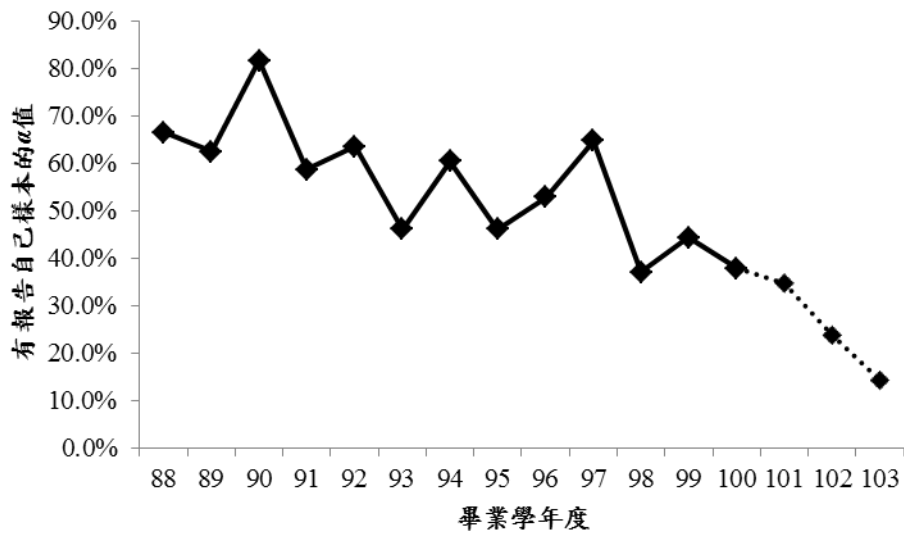


圖 6：有報告自己樣本的 Cronbach's α 確切數值的比例隨畢業學年度的變化

表 5

原始學門分類

學門	系所	論文數	有報告自己的 α 確切數值的論文比例
民生學門	休閒管理研究所、旅遊健康研究所、生活應用科學研究所、.....	39	61.5%
社會及行為科學學門	心理學研究所、行為醫學研究所、社會學系、.....	49	26.5%
社會服務學門	社會工作學研究所、復建諮商研究所、.....	16	56.3%
商業及管理學門	國際企業學系碩士班、管理研究所、醫務管理學研究所、健康產業管理研究所、.....	69	53.6%
教育學門	教育心理與輔導研究所、體育研究所、成人教育研究所、.....	28	32.1%
醫藥衛生學門	公共衛生研究所、職能治療研究所、護理學研究所、長期照護研究所、.....	291	51.2%
人文學門 ^a	生死學研究所、醫學人文研究所	2	0.0%
工程學門 ^a	工業工程與工程管理學系、工業工程與經營資訊學系	8	25.0%
生命科學學門 ^a	生命科學研究所、遺傳學研究所	2	50.0%
建築及都市規劃學門 ^a	建築學系碩博士班	1	100.0%
設計學門 ^a	創新設計研究所	1	0.0%
農業科學學門 ^a	園藝學研究所、農業推廣學研究所	2	0.0%
電算機學門 ^a	資訊管理學系	1	100.0%
數學及統計學門 ^a	應用統計學研究所	2	50.0%
藝術學門 ^a	舞蹈研究所理論組	1	0.0%
其他學門	老年學研究所	2	50.0%

註：^a 合併到「其他學門」。

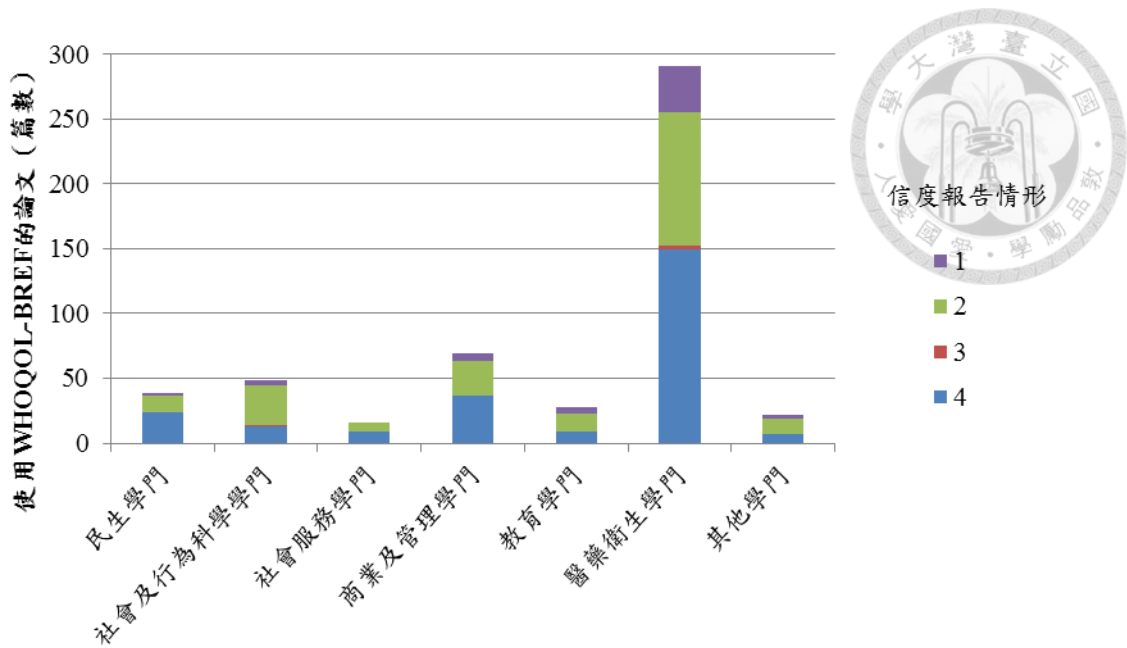


圖 7：各學門使用台灣版 WHOQOL-BREF 的論文數量及信度報告情形

註：信度報告情形 1 = 沒提信度；2 = 只有引用別人的信度；3 = 有自己的信度，但只給範圍；4 = 有自己的信度，且有確切數值。

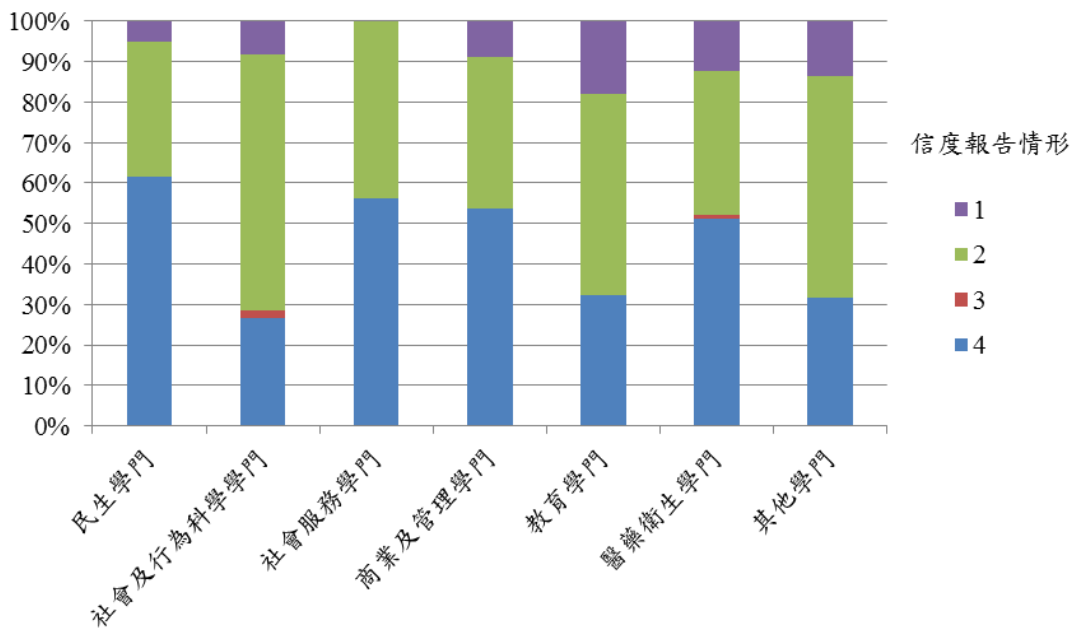


圖 8：各學門四類信度報告情形的比例

註：信度報告情形 1 = 沒提信度；2 = 只有引用別人的信度；3 = 有自己的信度，但只給範圍；4 = 有自己的信度，且有確切數值。

第二節 研究變項之描述統計



一、依變項

如表 6 和圖 9 所示，相較於四個範疇的 Cronbach's α ，本研究蒐集到較多整份問卷的 Cronbach's α ，除了因為較多研究者提供，另一方面是因為四範疇的 Cronbach's α 較容易因為刪題而不納入本研究中。Cronbach's α 的平均值以整份問卷的最高($M = .901$)，其餘四範疇的平均較為接近，其中以社會關係範疇最低($M = .714$)。整份問卷的 Cronbach's α 的最小值也是明顯比四個範疇的高，社會關係範疇還是最低。倒是 Cronbach's α 的最大值雖然也是整份問卷的最高，但和四範疇沒有相差很大。整份問卷的 Cronbach's α 標準差明顯比四個範疇的低($SD = .043$)；生理健康、心理、以及環境範疇的 Cronbach's α 標準差差距不大，介於.071-.075 之間；社會關係的 Cronbach's α 標準差最大($SD = .090$)。

表 6

整份問卷和四個範疇的 Cronbach's α 描述統計

	<i>N</i>	平均數	標準差	最小值	最大值
整份問卷	291	0.901	0.043	0.616	0.976
生理健康	259	0.756	0.075	0.440	0.930
心理	268	0.769	0.074	0.380	0.940
社會關係	254	0.714	0.090	0.270	0.930
環境	258	0.790	0.071	0.330	0.953

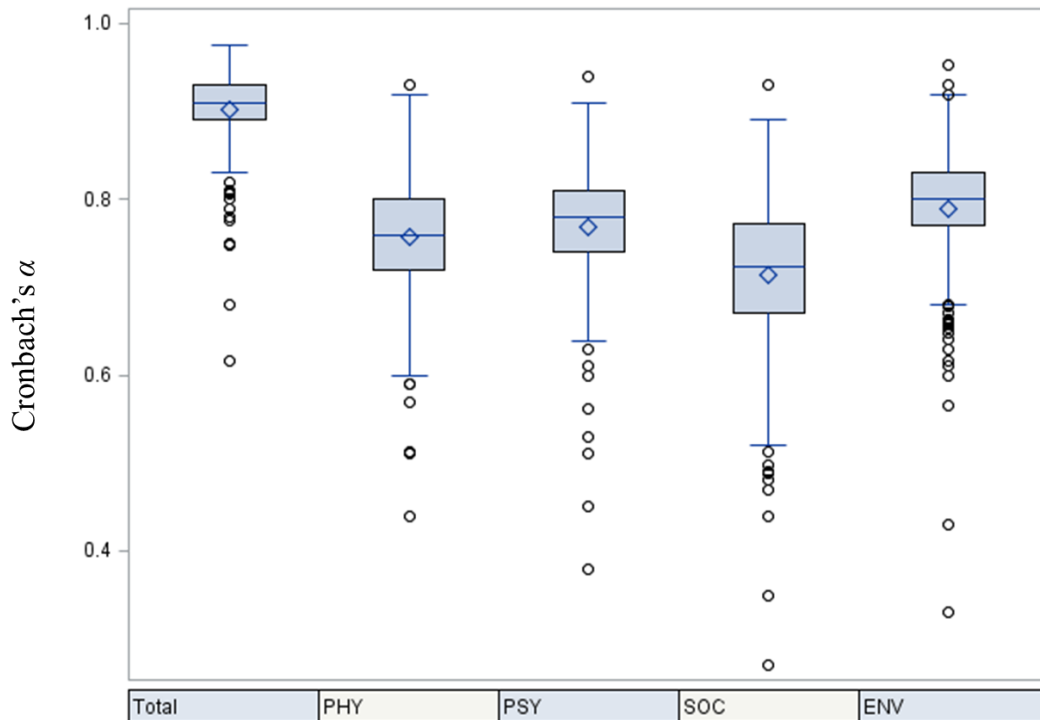


圖 9：整份問卷和四個範疇的 Cronbach's α 箱型圖

註：Total = 整份問卷；PHY = 生理；PSY = 心理；SOC = 社會關係；ENV = 環境。◇ = M ；○ = 大於 $Q_3 + 1.5$ 倍四分位距或小於 $Q_1 - 1.5$ 倍四分位距的觀察值。

從圖 9 可以發現有少數論文的 Cronbach's α 相較於其他論文來說特別低，於是先刪除過低的觀察值，再進行後續的統計分析。整份問卷的 Cronbach's α 刪除兩筆特別低的資料，生理範疇刪除一筆，其餘三個範疇都各自刪除兩筆特別低的資料。總共刪了九個 Cronbach's α ，其中三個來自同一篇論文，其餘六個分別來自不同論文。

整份問卷和四個範疇分別刪除一、兩個界外值後，Cronbach's α 的描述統計呈現於表 7。相比刪除前，Cronbach's α 的平均數在整份問卷和四個範疇上升 0.002-0.003，標準差也都明顯下降，最小值自然是提升許多。

表 7

刪除界外值後整份問卷和四個範疇的 Cronbach's α 的描述統計

	<i>N</i>	平均數	標準差	最小值	最大值
整份問卷	289	0.903	0.037	0.747	0.976
生理健康	258	0.758	0.072	0.510	0.930
心理	266	0.772	0.067	0.510	0.940
社會關係	252	0.717	0.082	0.440	0.930
環境	256	0.793	0.061	0.565	0.953

二、獨變項

表 8 和表 9 呈現了整份問卷和四個範疇的獨變項的描述統計，除了「平均分數」和「分數的標準差」這兩個變項在定義上整份問卷和四個範疇不同，因此結果比較有差異之外，其他每個變項在整份問卷和四個範疇的結果皆類似。許多論文沒有足夠的資訊供本研究當成獨變項使用，因此產生許多遺失值。需注意的是由於多數論文沒有提供整份問卷的得分，所以整份問卷的「分數的標準差」遺失值很多，這會使得後續的分析如果有使用到此變項，研究的論文篇數都會變得比較小。在地區變項上，大約一半的研究是所有參與者都居住在北部，因此之後的迴歸分析有時會因為所選取到的研究剛好沒有某個地區類別，地區變項會變成四組。

表 8

連續型獨變項的平均數和標準差

		整份問卷	生理健康	心理	社會關係	環境
平均分數	<i>N</i>	155	157	164	148	155
	平均數	52.97	13.41	12.83	13.59	13.44
	標準差	3.95	1.28	1.04	0.85	1.04
分數的標準差	<i>N</i>	54	145	150	132	140
	平均數	8.05	2.43	2.48	2.31	2.12
	標準差	1.88	0.49	0.44	0.36	0.35
樣本數	<i>N</i>	248	231	239	225	232
	平均數	200.77	208.95	215.37	216.67	215.97
	標準差	219.51	233.30	232.97	237.73	234.73
平均年齡	<i>N</i>	147	138	141	127	132
	平均數	54.05	53.73	54.14	53.51	54.52
	標準差	16.34	17.63	17.78	17.56	17.96
年齡標準差	<i>N</i>	136	127	129	115	120
	平均數	9.16	8.76	8.72	8.63	8.56
	標準差	3.68	3.68	3.66	3.77	3.71
性別	<i>N</i>	189	179	184	169	176
	平均數	0.42	0.43	0.44	0.44	0.45
	標準差	0.27	0.28	0.28	0.29	0.28
教育程度	<i>N</i>	154	142	147	134	139
	平均數	0.39	0.40	0.41	0.42	0.41
	標準差	0.32	0.33	0.34	0.34	0.34
宗教	<i>N</i>	108	95	97	85	92
	平均數	0.73	0.74	0.74	0.73	0.73
	標準差	0.17	0.17	0.17	0.17	0.17
婚姻狀態	<i>N</i>	163	153	159	146	152
	平均數	0.62	0.62	0.62	0.63	0.61
	標準差	0.25	0.26	0.26	0.25	0.26

表 9

類別型獨變項的各組人數和百分比

		整份問卷	生理健康	心理	社會關係	環境
樣本類型						
全病人	<i>n</i>	80	73	71	68	67
	%	32.3%	31.7%	29.8%	30.2%	28.9%
非全病人	<i>n</i>	168	157	167	157	165
	%	67.7%	68.3%	70.2%	69.8%	71.1%
地區						
北部	<i>n</i>	112	110	114	103	107
	%	52.1%	55.3%	55.3%	53.1%	53.5%
中部	<i>n</i>	38	36	38	39	38
	%	17.7%	18.1%	18.4%	20.1%	19.0%
南部	<i>n</i>	52	39	40	40	43
	%	24.2%	19.6%	19.4%	20.6%	21.5%
東部	<i>n</i>	4	4	4	2	2
	%	1.9%	2.0%	1.9%	1.0%	1.0%
混合	<i>n</i>	9	10	10	10	10
	%	4.2%	5.0%	4.9%	5.2%	5.0%

第三節 Cronbach's α 的預測因子

以下分析將分成五個部分說明，分別以整份問卷及四個範疇的 Cronbach's α 作為迴歸分析的依變項。



一、整份問卷

289 個 Cronbach's α 中只有 27 個包含所有獨變項，所以此部分的分析只會使用這 27 筆資料。由理論和過去研究知道分數的標準差對信度係數會有影響，所以首先使用分數的標準差去預測 Cronbach's α ，得到的殘差以 Kolmogorov-Smirnov 檢定發現不符合常態分配($p < .01$)，殘差的絕對值和預測值的 Spearman 相關未達顯著水準。在刪除兩個極端值後，使用剩下的 25 筆資料，再次以分數的標準差去預測 Cronbach's α ，Kolmogorov-Smirnov 檢定結果殘差符合常態分配($p > .15$)，Spearman 相關也未達顯著水準。迴歸結果顯示分數的標準差對整份問卷的 Cronbach's α 有顯著的預測力($R^2 = .382$ ， $F(1, 23) = 14.22$ ， $p = .001$)。

接著計算去除掉分數的標準差影響後，各別獨變項和 Cronbach's α 的淨相關，結果都不顯著。淨 F 檢定顯示去除分數的標準差影響後，地區變項也未達顯著水準($F(3, 20) = 1.15$ ， $p = .3545$) (地區剩北部、南部、東部、混合四組)。因此最終迴歸模型確定，結果如表 10 所示，當分數的標準差越大，Cronbach's α 也會越大。

表 10

整份問卷 Cronbach's α 的迴歸分析 ($N = 25$)

變項	df	迴歸係數	標準誤	t 值	p 值	標準化迴歸係數
截距	1	0.854	0.016	52.88	< .0001	0
分數的標準差	1	0.007	0.002	3.77	.0010	0.618

$R^2 = .382$ ， $F(1, 23) = 14.22$ ， $p = .001$

二、生理健康範疇

258 個 Cronbach's α 中只有 54 個包含所有獨變項，所以此部分的分析只會

使用這 54 筆資料。首先使用分數的標準差去預測 Cronbach's α ，得到的殘差以 Kolmogorov-Smirnov 檢定發現不符合常態分配($p < .05$)，殘差的絕對值和預測值的 Spearman 相關未達顯著水準。在刪除一個極端值後，使用剩下的 53 筆資料，再次以分數的標準差去預測 Cronbach's α ，Kolmogorov-Smirnov 檢定結果符合殘差常態分配($p > .15$)，Spearman 相關也未達顯著水準。迴歸結果顯示分數的標準差對整份問卷的 Cronbach's α 有顯著的預測力($R^2 = .403$ ， $F(1, 51) = 34.47$ ， $p < .0001$)。

接著計算去除掉分數的標準差影響後，每個獨變項和 Cronbach's α 的淨相關，結果除了宗教和 Cronbach's α 的淨相關為 .282 ($p < .05$)，其餘皆不顯著。淨 F 檢定顯示去除分數的標準差影響後，地區變項達顯著水準($F(3, 48) = 3.39$ ， $p < .05$) (地區剩北部、中部、南部、混合四組)，因此考慮將宗教和地區兩個變項放入模型中。表 11 的模型 3 和模型 2 比較，淨 F 檢定結果發現宗教變項對 Cronbach's α 有顯著的預測力($F(1, 47) = 5.63$ ， $p < .05$)。模型 3 和模型 1 比較，淨 F 檢定結果顯示地區變項對 Cronbach's α 有顯著的預測力($F(3, 47) = 3.90$ ， $p < .05$)，但若細看地區變項的迴歸係數會發現都沒有達顯著水準，只有地區 3 (1 = 南部，0 = 其他) 是邊緣顯著($p = .0631$)，這樣的結果是因為地區 1 (1 = 北部，0 = 其他) 和地區 2 (1 = 中部，0 = 其他) 把部分 Cronbach's α 的變異解釋掉了，導致當地區 1 和地區 2 已經在模型中的時候，地區 3 的迴歸係數就不顯著了。如果將地區變項的分組改成南部一組，北部、中部、和混合是另一組，結果會發現地區的迴歸係數達到顯著水準($p < .05$)，但為了和其他分析的變項定義一致，所以這裡還是維持模型 3 的變項定義方式，且以模型 3 為最後模型，並再次確認殘差符合假設。從模型 3 可預測當分數的標準差或參與者有宗教信仰的比例越大時，Cronbach's α 會越大。或是當研究參與者都來自南部的時候，Cronbach's α 也會越大。

表 11

生理範疇 Cronbach's α 的迴歸分析 ($N = 53$)

變項	df	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
模型 1							
截距	1	0.511	0.044	11.52***	0	.451	20.52***
分數標準差	1	0.073	0.014	5.43***	0.585	(-.429)	
宗教	1	0.097	0.047	2.08*	0.224		
模型 2							
截距	1	0.562	0.039	14.52***	0	.508	12.37***
分數標準差	1	0.070	0.013	5.33***	0.558	(.467)	
地區 1	1	0.033	0.027	1.24	0.233		
地區 2	1	-0.018	0.033	-0.54	-0.079		
地區 3	1	0.065	0.030	2.13*	0.363		
模型 3							
截距	1	0.511	0.043	11.92***	0	.560	11.97***
分數標準差	1	0.064	0.013	4.99***	0.510	(.513)	
宗教	1	0.106	0.045	2.37*	0.243		
地區 1	1	0.019	0.026	0.73	0.134		
地區 2	1	-0.034	0.032	-1.04	-0.146		
地區 3	1	0.056	0.029	1.9	0.312		

註：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)。*** $p < .001$ 。

三、心理範疇

266 個 Cronbach's α 中只有 56 個包含所有獨變項，所以此部分的分析只會

使用這 56 筆資料。首先使用分數的標準差去預測 Cronbach's α ，得到的殘差以 Kolmogorov-Smirnov 檢定發現不符合常態分配($p = .0104$)，殘差的絕對值和預測值的 Spearman 相關未達顯著水準。在刪除兩個極端值後，使用剩下的 54 筆資料，再次以分數的標準差去預測 Cronbach's α ，Kolmogorov-Smirnov 檢定結果殘差符合常態分配($p > .15$)，Spearman 相關也未達顯著水準。迴歸結果顯示分數的標準差對整份問卷的 Cronbach's α 有顯著的預測力($R^2 = .175$ ， $F(1, 52) = 11.00$ ， $p < .01$)。

接著計算去除掉分數的標準差影響後，各別獨變項和 Cronbach's α 的淨相關，結果都不顯著。淨 F 檢定顯示去除分數的標準差影響後，地區變項也未達顯著水準($F(3, 49) = 0.19$ ， $p = .9043$) (地區剩北部、中部、南部、混合四組)。因此最終迴歸模型確定，結果如表 12 所示。

表 12

心理範疇 Cronbach's α 的迴歸分析 ($N = 54$)

變項	df	迴歸係數	標準誤	t 值	p 值	標準化迴歸係數
截距	1	0.631	0.044	14.45	< .0001	0
分數的標準差	1	0.055	0.017	3.32	.0017	0.418

$R^2 = .175$ ， $F(1, 52) = 11.00$ ， $p < .01$ 。

四、社會關係範疇

252 個 Cronbach's α 中只有 45 個包含所有獨變項，所以此部分的分析只會使用這 45 筆資料。首先使用分數的標準差去預測 Cronbach's α ，得到的殘差符合常態分配和變異數同質假設。迴歸結果顯示分數的標準差對整份問卷的 Cronbach's α 有顯著的預測力($R^2 = .156$ ， $F(1, 43) = 7.95$ ， $p < .01$)。

接著計算去除掉分數的標準差影響後，各別獨變項和 Cronbach's α 的淨相關，結果都不顯著。淨 F 檢定顯示去除分數的標準差影響後，地區變項也未達

顯著水準($F(3, 40) = 0.69, p = .5652$) (地區剩北部、中部、南部、混合四組)。
因此最終迴歸模型確定，結果如表 13 所示。



表 13

社會關係範疇 Cronbach's α 的迴歸分析 ($N = 45$)

變項	<i>df</i>	迴歸係數	標準誤	<i>t</i> 值	<i>p</i> 值	標準化迴歸係數
截距	1	0.597	0.048	12.35	< .0001	0
分數的標準差	1	0.057	0.020	2.82	.0072	0.395

$R^2 = .156, F(1, 43) = 7.95, p < .01$ 。

五、環境範疇

256 個 Cronbach's α 中只有 50 個包含所有獨變項，所以此部分的分析只會使用這 50 筆資料。首先使用分數的標準差去預測 Cronbach's α ，得到的殘差符合常態分配和變異數同質假設。迴歸結果顯示分數的標準差對整份問卷的 Cronbach's α 有顯著的預測力($R^2 = .084, F(1, 48) = 4.42, p < .05$)。

接著計算去除掉分數的標準差影響後，各別獨變項和 Cronbach's α 的淨相關，結果樣本數和 Cronbach's α 的淨相關為-.311 ($p < .05$)，平均年齡和 Cronbach's α 的淨相關為-.364 ($p < .05$)，性別和 Cronbach's α 的淨相關為-.367 ($p < .01$)，教育程度和 Cronbach's α 的淨相關為.452 ($p < .01$)，其他變項的淨相關不顯著。淨 *F* 檢定顯示去除分數的標準差影響後，地區變項未達顯著水準($F(3, 45) = 2.01, p = .1261$) (地區剩北部、中部、南部、混合四組)。因此考慮將樣本數、平均年齡、性別以及教育程度這四個變項放入模型中。首先觀察這四個變項和 Cronbach's α 彼此間的皮爾森相關 (表 14)，發現年齡、性別和教育程度彼此間有中度相關，因此推測這三個變項對 Cronbach's α 的預測可能有重疊的部分。雖然樣本數和 Cronbach's α 的相關 (邊緣顯著， $p = .0508$) 較其他三個變項弱，但樣本數和這三個變項的相關都不顯著，因此先將樣本數放入模型中。接著計

算去除掉分數的標準差和樣本數影響後，年齡、性別、教育程度和 Cronbach's α 的淨相關，結果教育程度和 Cronbach's α 的淨相關最高（表 15），所以教育程度也放入模型中。然後再計算去除掉分數的標準差、樣本數和教育程度影響後，年齡、性別和 Cronbach's α 的淨相關，結果都未達顯著水準，因此最終模型的獨變項就是分數的標準差、樣本數和教育程度（表 16），並且再次確認殘差符合假設。

表 14

Cronbach's α 、樣本數、平均年齡、性別和教育程度間皮爾森相關 ($N = 50$)

	α	樣本數	平均年齡	性別	教育程度
α					
樣本數	-.278 ^a				
平均年齡	-.344*	.059			
性別	-.331*	.172	.630***		
教育程度	.447**	-.012	-.729***	-.489***	

註：^a $p = .0508$ 。* $p < .05$ ，** $p < .01$ ，*** $p < .001$ 。

表 15

去除分數標準差和樣本數，年齡、性別、教育程度和 α 的淨相關 ($N = 50$)

	平均年齡	性別	教育程度
Cronbach's α	-0.365*	-0.336*	0.470***

* $p < .05$ ，*** $p < .001$ 。

表 16

環境範疇 Cronbach's α 的迴歸分析 ($N = 50$)

變項	df	迴歸係數	標準誤	t 值	p 值	標準化迴歸係數
截距	1	0.694	0.035	19.59	< .0001	0
分數的標準差	1	0.039	0.016	2.42	.0194	0.288
樣本數	1	-0.0001	0.00004	-2.46	.0176	-0.292
教育程度	1	0.096	0.026	3.62	.0007	0.428

$R^2 = .356$ ，調整後 $R^2 = .314$ ， $F(3, 46) = 8.48$ ， $p = .0001$ 。

第四節 Cronbach's α 與獨變項兩兩變項間關係

上一節的迴歸分析雖然可以同時考慮多個獨變項，缺點是許多論文會因為有遺失值而無法使用，因此本研究也同時進行 Cronbach's α 和某一獨變項兩兩關係的分析，這部分的分析中的 Cronbach's α 數量幾乎都可以增加到一百多，甚至兩百多，比起迴歸分析時的數量增加不少。綜合兩部分的結果更能了解 Cronbach's α 和研究參與者有關的因素（即獨變項）之間是否有所關聯。

計算 Cronbach's α 和連續型獨變項的皮爾森相關(表 17)，發現除了分數的標準差，大部分的獨變項和四個範疇或整份問卷的 Cronbach's α 的相關皆未達顯著水準，只有生理範疇的 Cronbach's α 和平均年齡、性別、教育程度的相關達顯著，以及環境範疇的 Cronbach's α 和平均年齡、教育程度的相關達顯著。若進一步計算去除分數的標準差的影響後，這些變項和 Cronbach's α 的淨相關（刪去沒有呈現分數標準差的資料約二、三十筆）(表 18)，發現平均年齡和生理及環境範疇的 Cronbach's α 都有顯著的淨相關，教育程度和環境範疇的 Cronbach's α 也有顯著淨相關。結果顯示去除分數標準差的影響後，研究參與者的平均年齡越大，生理範疇的 Cronbach's α 也會越大，但環境範疇的 Cronbach's

α 會越小；參與者教育程度為大專院校以上的比例越高，環境範疇的 Cronbach's α 會越大。生理範疇中，性別和教育程度原先和 Cronbach's α 的相關顯著，但去除分數標準差的影響後淨相關則不顯著，因此推測這兩個變項可能是透過分數的標準差去影響 Cronbach's α 的。

另外值得注意的是，社會關係範疇的 Cronbach's α 和分數標準差的皮爾森相關並未如預期達顯著水準。所有研究變項的兩兩相關參見附錄。

樣本類型變項分成全部都是病人和不是全部都是病人兩組， t 檢定結果顯示不管在整份問卷或四個範疇，兩組的 Cronbach's α 平均數都沒有差異(表 19)。地區變項分成北部、中部、南部、東部和混合五組，變異數分析結果也都沒有顯著(表 20)，表示不管在整份問卷或四個範疇，五組地區的 Cronbach's α 平均數都沒有差異。另外如果違反 t 檢定變異數同質的假設時，會使用 Satterthwaite (1946)提出的方式去校正自由度。變異數分析時會使用 Levene 檢定(Levene, 1960)是否符合變異數同質的假設，若違反時則改用 Welch (1951)的方式進行變異數分析。

表 17

Cronbach's α 與連續型獨變項的皮爾森相關

		整份問卷	生理健康	心理	社會關係	環境
平均分數	<i>r</i>	-.078	-.0002	-.069	-.045	-.105
	<i>N</i>	155	157	164	148	155
分數的標準差	<i>r</i>	.320*	.522***	.322***	.040	.368***
	<i>N</i>	54	145	150	132	140
樣本數	<i>r</i>	.070	-.066	-.038	.110	.011
	<i>N</i>	248	231	239	225	232
平均年齡	<i>r</i>	-.030	.400***	.027	.054	-.208*
	<i>N</i>	147	138	141	127	132
年齡標準差	<i>r</i>	-.165	-.005	-.031	-.050	.058
	<i>N</i>	136	127	129	115	120
性別	<i>r</i>	.015	.224**	.039	.145	-.124
	<i>N</i>	189	179	184	169	176
教育程度	<i>r</i>	.124	-.208*	.134	.135	.199*
	<i>N</i>	154	142	147	134	139
宗教	<i>r</i>	.055	.091	.038	-.094	.094
	<i>N</i>	108	95	97	85	92
婚姻狀態	<i>r</i>	-.102	-.084	-.056	.039	-.043
	<i>N</i>	163	153	159	146	152

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

表 18

去掉分數標準差，平均年齡、性別、教育程度和 Cronbach's α 淨相關

		生理健康	環境
平均年齡	<i>r</i>	0.208*	-0.211*
	<i>N</i>	119	112
性別	<i>r</i>	0.113	-
	<i>N</i>	143	-
教育程度	<i>r</i>	0.050	0.279**
	<i>N</i>	118	113

* $p < .05$ ，** $p < .01$ 。

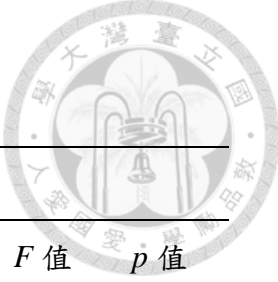
表 19

「全病人」和「非全病人」研究的 Cronbach's α 比較

	樣本類型	<i>n</i>	平均數	標準差	<i>t</i> 值	<i>p</i> 值
整份問卷	非全病人	168	0.906	0.032	1.43 ^a	.1540
	全病人	80	0.899	0.044		
生理健康	非全病人	157	0.755	0.076	-0.37	.7083
	全病人	73	0.759	0.068		
心理	非全病人	167	0.771	0.064	-1.17	.2428
	全病人	71	0.781	0.066		
社會關係	非全病人	157	0.717	0.083	-0.28	.7780
	全病人	68	0.720	0.072		
環境	非全病人	165	0.794	0.063	-0.06	.9521
	全病人	67	0.795	0.053		

註：^a變異數不同質，使用 Satterthwaite 法

表 20

各地區研究的 Cronbach's α 比較


	北部	中部	南部	東部	混合	F 值	p 值
	n	n	n	n	n		
	平均數	平均數	平均數	平均數	平均數		
	標準差	標準差	標準差	標準差	標準差		
整份問卷	112	38	52	4	9	2.43 ^a	.0852
	0.907	0.889	0.911	0.910	0.918		
	0.030	0.050	0.026	0.026	0.014		
生理健康	110	36	39	4	10	2.13	.0783
	0.756	0.759	0.771	0.772	0.695		
	0.080	0.051	0.076	0.038	0.082		
心理	114	38	40	4	10	1.86	.1184
	0.778	0.764	0.782	0.762	0.727		
	0.065	0.054	0.069	0.059	0.091		
社會關係	103	39	40	2	10	1.06 ^a	.4391
	0.729	0.707	0.703	0.673	0.710		
	0.074	0.067	0.112	0.060	0.047		
環境	107	38	43	2	10	1.29	.2735
	0.793	0.787	0.809	0.832	0.816		
	0.060	0.059	0.056	0.016	0.020		

註：^a變異數不同質，改用 Welch 的方法

第四章 討論與結論



第一節 信度係數報告的情形

在本研究蒐集的台灣博碩士論文中，有 48.2%的論文有報告使用自己研究樣本所估計出的信度係數，這個比例相較與過去的信度概化研究來說是滿高的比例，這是一個值得高興的結果。因為許多學者都指出信度是屬於測驗分數結果的一項特質，而不是屬於測驗工具本身(Crocker & Algina, 1986; Gronlund & Linn, 1990; Rowley, 1976; B. Thompson, 1994)，因此每個研究都應該針對自己的研究樣本去估計信度係數，而非直接引用過去研究的信度結果作為依據。但是觀察了隨著時間的演進，使用台灣版 WHOQOL-BREF 的論文數量明顯的逐年上升，遺憾的是在信度報告情形上，有報告自己研究樣本的信度係數的論文比例卻是逐年下降（圖 5、圖 6）。這項結果可能是出於越來越多研究者對信度的有所誤解，因此沒有估計自己樣本的信度係數。不過在編碼的過程中也有發現，整體來說，過去博碩士論文的資料分析方法較現在的論文簡單一些，信效度分析可能會獨立成一個小節探討。現在的論文做了較多複雜的分析，信效度的部分可能因此被忽略了。Graham 和 Christiansen (2009)也曾針對五個常用的愛情量表做過同樣的分析，結果和本研究相反，從 1970 到 2006 年間，有報告自己樣本的信度係數的比例是逐漸上升的，在 2000 到 2006 年之間，比例更是高過七成，不過 Graham 和 Christiansen 的研究對象是期刊文章，本研究使用的是博碩士論文。以學門分類分析信度報告情形的結果有些出乎意料（圖 8），社會及行為科學學門大多是心理學相關的研究所，理論上這些碩博士生應該都修過心理測驗這門課，但似乎並沒有因此了解呈現信度係數的重要性。

另外在編碼的時候也發現很多論文報告信度的方式有些問題。比方說雖然

有 48.2% 的論文有報告自己樣本的信度係數，但其中 24 篇只有提供預試的信度結果，並沒有呈現正式研究的信度。只有引用問卷使用手冊或過去研究的信度結果的論文中，有 30 篇只有說具有良好的信度，沒有說清楚是哪種信度。但不同信度估計方法反映的是不同的誤差來源，所以應該說明清楚，好讓其他研究者作為參考。

第二節 不同研究中 Cronbach's α 的變異

首先本研究想要探討題目、量尺等等各方面完全一樣的測量工具在不同研究中，不同群體的受試者產生的 Cronbach's α 有何變異。由 Cronbach's α 的描述統計量可以發現台灣版 WHOQOL-BREF 在不同研究樣本中 Cronbach's α 的變異相當大（表 7、圖 9），而且呈現負偏。儘管問卷使用手冊中整份問卷的 Cronbach's α 達 .91，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，都是可接受的信度大小（Nunnally, 1978），但在一些博碩士論文中的也出現了相當低的 Cronbach's α ，遠低於一般認為可接受的 .70。比較整份問卷和不同範疇在不同研究樣本中的變異情形，發現使用全部題目去計算出的 Cronbach's α 明顯比四個範疇的穩定，四個範疇中又以題數最少的社會關係範疇（4 題）的變異最大。過去也有模擬研究發現題數越多，Cronbach's α 的標準誤會越小（Duhachek & Iacobucci, 2004）。不過不管是哪個範疇或整份問卷，都至少有接近 70% 以上的論文的 Cronbach's α 達到 .70 以上。但也可能會有整合分析常面臨到的檔案抽屜問題（file-drawer problem），就是研究者可能只呈現較好的統計數據，也就是說，實際上 Cronbach's α 的變異可能會比本研究觀察到的還要大。

在進行迴歸、相關、變異數分析前，刪去了 9 個界外值，這 9 個界外值中有其中 3 個是預試的結果，樣本數只有 6 到 10 人。樣本數雖然不影響 Cronbach's

α 大小，但會影響 Cronbach's α 的標準誤，樣本數越小，標準誤越大(Duhachek & Iacobucci, 2004)。建議大家不管是預試或正式研究都不要使用太小的樣本數進行研究。



第三節 影響 Cronbach's α 在不同研究中變異的因子

接下來探討了 Cronbach's α 在不同研究中的變異是否可以用和研究參與者有關的因素去預測。迴歸分析的部分，在整份問卷和四個範疇中都發現分數的標準差可以有效的預測 Cronbach's α ，和理論預期的也和其他信度概化研究的結果一致，分數的標準差越大，Cronbach's α 就會越大。

相關分析也得到一致的結果的，除了社會關係範疇 Cronbach's α 和分數標準差的皮爾森相關未達顯著水準，但從散佈圖（圖 10）可以發現有兩個界外值（圖中圈起處），若刪除兩個界外值，相關會變成.18 ($p < .05$)。

迴歸分析和兩兩變項的分析中，除了分數的標準差，其他研究參與者有關的因素和 Cronbach's α 大多沒有關係，只有少部分的變項有統計上的顯著。比對兩者分析中顯著的變項，只有教育程度對環境範疇的 Cronbach's α 的影響是一致的。環境範疇中，迴歸分析和兩兩變項分析的結果都顯示參與者教育程度為大專院校以上的比例越高，Cronbach's α 會越大。去除掉分數標準差的影響後，平均年齡和 Cronbach's α 的淨相關也達到顯著水準，但從迴歸分析可以知道平均年齡和教育程度對 Cronbach's α 的解釋變異應該有不少重複的部分，因此在決定模型的過程中只選擇了其中一個。另外在生理範疇，平均年齡和 Cronbach's α 的淨相關也達到了顯著水準，但不同於環境範疇是負相關，生理範疇是正相關。台灣有研究發現不論男性或女性，隨著年齡的增長，生理健康、心理、社會關係和環境範疇的生活品質會越來越低，其中心理、社會關係和環境範疇是略為降低，生理範疇則是明顯下降許多（台灣版世界衛生組織生活品質問卷發

展小組，2005)。過去研究發現的是年齡對於各個範疇的生活品質會有所影響，在本研究中則是發現年齡對生理和環境範疇的內部一致性有影響。以生理範疇來說明，本研究發現相較於年紀輕的人，年齡較大的人在生理範疇中不同層面的生活品質（例如：精力、行動能力、睡眠狀況等等）的得分傾向同時高或同時低，這或許代表了年紀大的人若生理健康較差時，不同層面的生理生活品質都會同時變差；年輕人則可能因為個別狀態的不同，在不同層面的生理生活品質會有好有壞。在環境範疇則是年紀輕的人比較有不同題目同時高或低分的傾向。

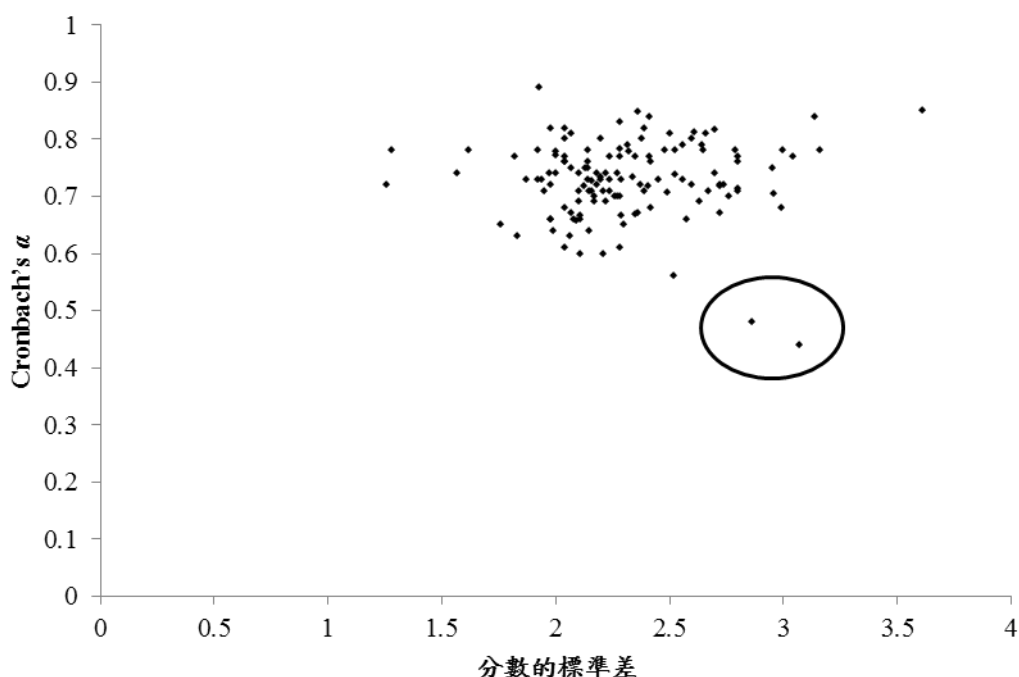
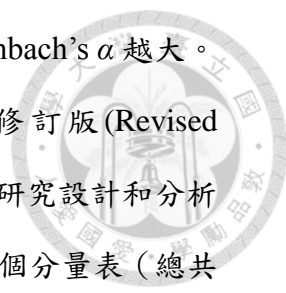


圖 10：社會範疇 Cronbach's α 和分數標準差的散佈圖

將本研究的結果和其他信度概化研究相比較，Shields 和 Caruso(2003)針對酒精使用疾患確認檢測(Alcohol Use Disorders Identification Test, AUDIT)進行信度概化研究，和本研究一樣以 Cronbach's α 作為依變項，使用兩階段的迴歸模型去預測 Cronbach's α ，第一階段只放分數的標準差，第二階段才放入其他獨

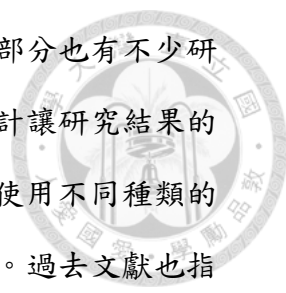


變項，結果發現平均年齡有顯著的預測力，平均年齡越小，Cronbach's α 越大。Ryngala、Shields 和 Caruso(2005)使用兒童顯性焦慮量表修訂版(Revised Children's Manifest Anxiety Scale, RCMAS)進行信度概化研究，研究設計和分析方法和 Shields 和 Caruso(2003)一樣，結果發現在總分和其中三個分量表(總共四個分量表)上，平均年齡都有顯著的預測力，但預測方向恰好相反，都是平均年齡越大，Cronbach's α 越大。由此可見，平均年齡對 Cronbach's α 的影響似乎會隨著問卷內容不一樣，而有不同的方向。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。

另外樣本數、宗教和地區變項在迴歸和相關分析(和變異數分析)中沒有一致的結果，但有其他信度概化研究發現樣本數越大，信度係數會越小(Kieffer & Reese, 2002)，和本研究的結果一樣，不過多數信度概化研究是發現樣本數和信度沒有關係(Caruso, 2000; Henson et al., 2001; Vacha-Haase, Tani, et al., 2001; Vassar & Crosby, 2008)。也有信度概化研究發現施測國家(分成北美、歐洲和其他)的不同，也會影響 Cronbach's α ，北美的 Cronbach's α 會較高(Aguayo et al., 2011)。不過相較國家，台灣的北部、中部、南部是比較小的地區單位。沒有其他信度概化研究使用宗教信仰的有無當作獨變項，但這個變項是台灣博士論文中常見的人口學變項之一，因此被納入本研究中。有研究使用台灣版 WHOQOL-BREF 去比較基督徒和非基督徒的生活品質差異，發現基督徒生理、心理、環境範疇生活品質的高於非基督徒(林楷，2014)。但還沒有研究探討過宗教信仰的有無是否會影響題目的內部一致性。

第四節 研究貢獻

過去的信度概化研究，似乎沒有針對完全一樣的測驗工具去分析，有些研究沒有特別強調，無法確定工具是否一模一樣。大部分的研究可能會把不同語



言或不同版本，甚至完全不同的測驗放在一起分析。依變項的部分也有不少研究把不同估計方法或不同種類的信度係數混在一起。這樣的設計讓研究結果的解釋變得很混淆。Dimitrov (2002)也曾提出類似的想法，認為使用不同種類的信度，不同題數或版本的測驗去探討信度的變異會是一個問題。過去文獻也指出量尺點數(Ooster, 1989)、量尺種類(Cook, Heath, Thompson, & Thompson, 2001)、題數(Crocker & Algina, 1986; Murphy & Davidshofer, 1997)都是會影響信度的因素。因此本研究特別強調工具的一致，只要稍微對量表作出一點更動就會排除在本研究的對象之外。雖然研究的篇數因而少了許多，但也更能了解樣本本身的特性是否能去預測 Cronbach's α 在不同研究中的變異。

本研究發現平均年齡會影響生理和環境範疇的 Cronbach's α ，也就是影響這兩個範疇中不同層面的生活品質的共變程度，這讓我們更了解年齡對生活品質的影響不僅僅是過去發現的會影響整個範疇得分的高低。

最後提醒未來的研究者，由於信度會隨著研究對象在所測量的特質上同質性的不同而有所改變，因此每個研究都應該使用自己的樣本去估計信度係數，不管研究目的是不是心理計量特性的探討，在報告此數值時應該說明清楚是使用什麼方法估計信度係數。對於樣本組成的特性（例如：性別、年齡）也應該有一些基本的說明，讓其他研究可以作為參考。

第五節 研究限制與未來研究方向

為了解決可分析的論文篇數容易偏小的問題，本研究使用博碩士論文當作研究對象，但博碩士論文的品質可能會比較參差不齊。在編碼的時候也發現許多問題，例如 WHOQOL-BREF 中兩題整體性的題目本來是獨立於四個範疇之外，有些論文卻自己把這兩題歸到某個範疇裡。還有論文的生理範疇 Cronbach's α 竟然是-3.13。另外也有樣本數為 143，最大、最小值分別為 83、31，標準差

卻只有 0.96 的狀況。這些問題在編碼時都會盡量挑出來，但可能也會有一些無法察覺到的問題混在當中，影響了本研究的分析結果。

1998 年信度概法化被提出之後，許多研究者紛紛採用這個方法對信度進行整合分析。也有研究者開始針對這些信度概化研究的分析方法進行比較(Mason, Allam, & Brannick, 2007; Sánchez-Meca, López-López, & López-Pina, 2013)，比方說固定效果模型(fixed-effect model)或隨機效果模型(random-effects model)的比較，不同權重的計算方式的比較等等。這個部份本研究並未加以討論。

國家衛生研究院進行的國民健康訪問調查(National Health Interview Survey, NHIS)在 2001 年有使用台灣版 WHOQOL-BREF，未來可以藉由這個資料庫的資料來分析研究參與者有關的因素和 Cronbach's α 之間的關係。由於有原始資料，不像本研究受限於其他論文給的資訊，在探討參與者相關因素和 Cronbach's α 的關係時可以有更細緻的討論，比方說可以把全部人分成不同年齡層去比較他們的 Cronbach's α 有沒有差異，不需要使用平均年齡。或是教育程度也是同樣的道理，可以區分成不同的教育程度來比較。之後將資料庫和整合分析的結果互相比較，以得到更深入的了解。

參考文獻

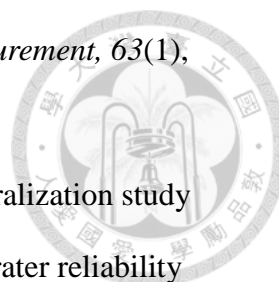


中文文獻

- 台灣版世界衛生組織生活品質問卷發展小組 (2005)。台灣簡明版世界衛生組織生活品質問卷之發展及使用手冊 (第二版)。
- 林楷(2014)。基督徒信仰對於生活品質的影響。《休憩管理研究》, 1(2), 頁 1-12。
- 劉誥洋(2006)。Spa 水療對憂鬱情緒及生活品質之影響—以台南地區 Spa 消費者為例。立德管理學院休閒管理研究所, 台南市。

英文文獻

- 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.
- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, Calif.: Brooks/Cole Pub. Co.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, N.J.: Prentice Hall.
- Barnes, L. L., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger state-trait anxiety inventory. *Educational and psychological measurement, 62*(4), 603-618.
- Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and psychological measurement, 62*(4), 570-589.
- Beretvas, S. N., & Pastor, D. A. (2003). Using mixed-effects models in reliability



generalization studies. *Educational and psychological measurement*, 63(1), 75-95.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PloS one*, 5(12), e14331.

Botella, J., Suero, M., & Gambara, H. (2010). Psychometric inferences from a meta-analysis of reliability and internal consistency coefficients. *Psychological methods*, 15(4), 386.

Campbell, J. S., Pulos, S., Hogan, M., & Murry, F. (2005). Reliability generalization of the Psychopathy Checklist applied in youthful samples. *Educational and psychological measurement*, 65(4), 639-656.

Capraro, R. M., & Capraro, M. M. (2002). Myers-briggs type indicator score reliability across: Studies a meta-analytic reliability generalization study. *Educational and psychological measurement*, 62(4), 590-602.

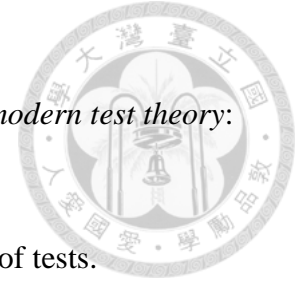
Caruso, J. C. (2000). Reliability generalization of the NEO personality scales. *Educational and psychological measurement*, 60(2), 236-254.

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.

Cook, C., Heath, F., Thompson, R. L., & Thompson, B. (2001). Score reliability in Webor internet-based surveys: unnumbered graphic rating scales versus Likert-type scales. *Educational and psychological measurement*, 61(4), 697-706.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104. doi:

10.1037//0021-9010.78.1.98



Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*:

Holt, Rinehart, and Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests.

Psychometrika, 16(3), 297-334. doi: 10.1007/bf02310555

Cronbach, L. J. (1990). *Essentials of psychological testing*. New York: Harper &

Row.

Dimitrov, D. M. (2002). Reliability: Arguments for multiple perspectives and

potential problems with generalization across studies. *Educational and psychological measurement*, 62(5), 783-801.

Duhachek, A., & Iacobucci, D. (2004). Alpha's standard error (ASE): an accurate

and precise confidence interval estimate. *Journal of Applied Psychology*, 89(5), 792.

Gleaves, D. H., Pearson, C. A., Ambwani, S., & Morey, L. C. (2014). Measuring

eating disorder attitudes and behaviors: a reliability generalization study. *Journal of eating disorders*, 2(1), 1.

Graham, J. M., & Christiansen, K. (2009). The reliability of romantic love: A

reliability generalization meta-analysis. *Personal Relationships*, 16(1), 49-66.

Gregory, R. J. (2007). *Psychological testing : history, principles, and applications*.

Boston: Pearson/Allyn and Bacon.

Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching*.

London: Collier Macmillan.

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley and Sons, Inc.

Hanson, W. E., Curry, K. T., & Bandalos, D. L. (2002). Reliability generalization of

working alliance inventory scale scores. *Educational and psychological*

measurement, 62(4), 659-673.

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.

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.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis : correcting error and bias in research findings*. Newbury Park, Calif.: Sage Publications.

Kieffer, K. M., & Reese, R. J. (2002). A reliability generalization study of the Geriatric Depression Scale. *Educational and psychological measurement*, 62(6), 969-994.

Levene, H. (1960). Robust tests for equality of variances¹. *Contributions to probability and statistics: Essays in honor of Harold Hotelling*, 2, 278-292.

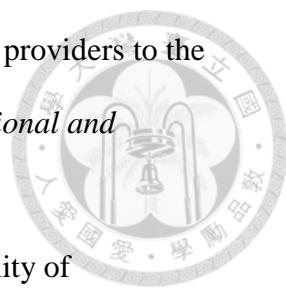
Mahapoonyanont, N., Krahomwong, R., Kochakornjarupong, D., & Rachasong, W. (2010). Critical thinking abilities assessment tools: reliability generalization. *Procedia-Social and Behavioral Sciences*, 2(2), 434-438.

Mason, C., Allam, R., & Brannick, M. T. (2007). How to Meta-Analyze Coefficient-of-Stability Estimates Some Recommendations Based on Monte Carlo Studies. *Educational and psychological measurement*, 67(5), 765-783.

Murphy, K. R., & Davidshofer, C. O. (1997). *Psychological testing : principles and applications* (4th ed ed.): Prentice Hall.

Nilsson, J. E., Schmidt, C. K., & Meek, W. D. (2002). Reliability generalization: An examination of the career decision-making self-efficacy scale. *Educational and psychological measurement*, 62(4), 647-658.

Nunnally, J. C. (1978). *Psychometric theory*: McGraw-Hill.

- 
- O'rourke, N. (2004). Reliability generalization of responses by care providers to the Center for Epidemiologic Studies-Depression Scale. *Educational and psychological measurement, 64*(6), 973-990.
- Oaster, T. (1989). Number of alternatives per choice point and stability of Likert-type scales. *Perceptual and Motor Skills, 68*(2), 549-550.
- 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.
- 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.
- Rowley, G. L. (1976). The reliability of observational measures. *American Educational Research Journal, 13*(1), 51-59.
- 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.
- Sánchez-Meca, J., López-López, J. A., & López-Pina, J. A. (2013). Some recommended statistical analytic practices when reliability generalization studies are conducted. *British Journal of Mathematical and Statistical Psychology, 66*(3), 402-425.
- Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics bulletin, 2*(6), 110-114.
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*(5), 529-540. doi: 10.1037/0021-9010.62.5.529
- Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological assessment,*

8(4), 350.

Shields, A. L., & Caruso, J. C. (2003). Reliability generalization of the alcohol use disorders identification test. *Educational and psychological measurement*, 63(3), 404-413.

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.

Thompson, B. (1994). Guidelines for authors. *Educational and psychological measurement*, 54, 837-847.

Thompson, B., & Vacha-Haase, T. (2000). Psychometrics is datametrics: The test is not reliable. *Educational and psychological measurement*, 60(2), 174-195.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and psychological measurement*, 58(1), 6-20.

Vacha-Haase, T., Kogan, L. R., Tani, C. R., & Woodall, R. A. (2001). Reliability generalization: Exploring variation of reliability coefficients of MMPI clinical scales scores. *Educational and psychological measurement*, 61(1), 45-59.

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.

Vassar, M., & Crosby, J. W. (2008). A reliability generalization study of coefficient alpha for the UCLA Loneliness Scale. *Journal of Personality Assessment*, 90(6), 601-607.

Viswesvaran, C., & Ones, D. S. (2000). Measurement error in "Big Five Factors" personality assessment: Reliability generalization across studies and

measures. *Educational and psychological measurement*, 60(2), 224-235.

Welch, B. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38(3/4), 330-336.

Yin, P., & Fan, X. (2000). Assessing the reliability of Beck Depression Inventory scores: Reliability generalization across studies. *Educational and psychological measurement*, 60(2), 201-223.

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.





附錄 研究變項的兩兩相關表

整份問卷：

	<i>r</i>	<i>α</i>	平均分數	分數標準差	樣本數	平均年齡	年齡標準差	性別	教育程度	宗教	婚姻狀態
平均分數	<i>r</i>	-.078									
	<i>N</i>	155									
分數標準差	<i>r</i>	.320*	.044								
	<i>N</i>	54	54								
樣本數	<i>r</i>	.070	.196*	-.073							
	<i>N</i>	248	155	54							
平均年齡	<i>r</i>	-.030	.087	.104	-.135						
	<i>N</i>	147	125	49	147						
年齡標準差	<i>r</i>	-.165	-.153	.014	-.022	-.083					
	<i>N</i>	136	115	45	136	135					
性別	<i>r</i>	.015	-.043	.345*	.007	.396***	.166				
	<i>N</i>	189	151	52	189	145	135				
教育程度	<i>r</i>	.124	.143	-.103	.253**	-.631***	-.201*	-.276***			
	<i>N</i>	154	128	45	154	120	113	151			
宗教	<i>r</i>	.055	.186	.091	-.179	.057	.156	-.133	-.305***		
	<i>N</i>	108	90	40	108	98	93	107	95		
婚姻狀態	<i>r</i>	-.102	.181*	-.154	-.071	-.064	.222*	-.089	-.159	.317**	
	<i>N</i>	163	135	50	163	132	123	161	139	101	
樣本類型	<i>r</i>	-.102	-.239**	.142	-.245***	-.059	.403***	.242***	-.329***	.104	.144
	<i>N</i>	248	155	54	246	147	136	189	154	108	163

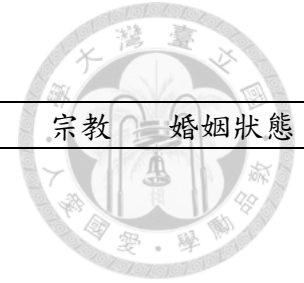
* $p < .05$, ** $p < .01$, *** $p < .001$ 。



生理健康範疇：

		α	平均分數	分數標準差	樣本數	平均年齡	年齡標準差	性別	教育程度	宗教	婚姻狀態
平均分數	<i>r</i>	-.0002									
	<i>N</i>	157									
分數標準差	<i>r</i>	.522***	-.223**								
	<i>N</i>	145	145								
樣本數	<i>r</i>	-.066	.289***	-.229**							
	<i>N</i>	231	157	145							
平均年齡	<i>r</i>	.400***	-.246**	.493***	-.125						
	<i>N</i>	138	125	119	138						
年齡標準差	<i>r</i>	-.005	-.124	.077	-.114	-.060					
	<i>N</i>	127	115	109	127	127					
性別	<i>r</i>	.224**	-.080	.271**	.044	.424***	.101				
	<i>N</i>	179	154	143	179	136	126				
教育程度	<i>r</i>	-.208*	.271**	-.515***	.189*	-.684***	-.198*	-.263**			
	<i>N</i>	142	127	118	142	117	110	140			
宗教	<i>r</i>	.091	-.011	.192	-.241*	-.032	.254*	-.345***	-.201		
	<i>N</i>	95	85	81	95	88	83	95	84		
婚姻狀態	<i>r</i>	-.084	.172*	-.116	-.097	-.203*	.268**	-.233**	-.023	.289**	
	<i>N</i>	153	135	126	153	124	115	151	127	90	
樣本類型	<i>r</i>	.025	-.269***	.139	-.209**	-.073	.473***	.261***	-.316***	.079	.127
	<i>N</i>	230	156	144	227	137	126	178	141	95	153

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



心理範疇：

		α	平均分數	分數標準差	樣本數	平均年齡	年齡標準差	性別	教育程度	宗教	婚姻狀態
平均分數	<i>r</i>	-.069									
	<i>N</i>	164									
分數標準差	<i>r</i>	.322***	-.471***								
	<i>N</i>	150	150								
樣本數	<i>r</i>	-.038	.218**	-.171*							
	<i>N</i>	239	164	150							
平均年齡	<i>r</i>	.027	-.163	.285**	-.122						
	<i>N</i>	141	129	122	141						
年齡標準差	<i>r</i>	-.031	-.077	.237*	-.115	-.067					
	<i>N</i>	129	118	111	129	129					
性別	<i>r</i>	.039	-.056	.243**	.038	.414***	.104				
	<i>N</i>	184	160	148	184	139	128				
教育程度	<i>r</i>	.134	.269**	-.403***	.212*	-.691***	-.190*	-.252**			
	<i>N</i>	147	133	123	147	119	111	144			
宗教	<i>r</i>	.038	.098	.088	-.236*	-.017	.248*	-.341***	-.207		
	<i>N</i>	97	87	83	97	90	85	97	86		
婚姻狀態	<i>r</i>	-.056	.258**	-.180*	-.102	-.206*	.270**	-.224**	-.033	.272**	
	<i>N</i>	159	142	131	159	127	117	156	132	92	
樣本類型	<i>r</i>	.076	-.216**	.328***	-.212**	-.078	.475***	.249***	-.323***	.071	.135
	<i>N</i>	238	163	149	235	140	128	183	146	97	159

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



社會關係範疇：

		α	平均分數	分數標準差	樣本數	平均年齡	年齡標準差	性別	教育程度	宗教	婚姻狀態
平均分數	r	-.045									
	N	148									
分數標準差	r	.040	-.365***								
	N	132	132								
樣本數	r	.110	.069	-.205*							
	N	225	148	132							
平均年齡	r	.054	-.066	.226*	-.105						
	N	127	114	105	127						
年齡標準差	r	-.050	-.014	.133	-.103	-.029					
	N	115	103	94	115	115					
性別	r	.145	-.177*	.265**	.054	.423***	.121				
	N	169	144	130	169	125	114				
教育程度	r	.135	.182*	-.388***	.192*	-.667***	-.232*	-.265**			
	N	134	119	107	134	107	99	131			
宗教	r	-.094	.335**	-.053	-.236*	.007	.219	-.329**	-.200		
	N	85	76	70	85	80	75	85	75		
婚姻狀態	r	.039	.351***	-.182	-.122	-.158	.247*	-.208*	-.062	.261*	
	N	146	129	116	146	115	105	143	121	81	
樣本類型	r	.019	-.256**	.278**	-.221***	-.057	.464***	.259***	-.359***	.059	.123
	N	225	148	132	222	127	115	169	134	85	146

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



環境範疇：

		α	平均分數	分數標準差	樣本數	平均年齡	年齡標準差	性別	教育程度	宗教	婚姻狀態
平均分數	r	-.105									
	N	155									
分數標準差	r	.368***	-.298***								
	N	140	140								
樣本數	r	.011	.047	-.026							
	N	232	155	140							
平均年齡	r	-.208*	.309***	-.015	-.111						
	N	132	120	112	132						
年齡標準差	r	.058	-.094	.154	-.098	-.060					
	N	120	109	101	120	120					
性別	r	-.124	.092	.026	.040	.427***	.107				
	N	176	151	138	176	130	119				
教育程度	r	.199*	-.072	-.110	.202*	-.673***	-.218*	-.273**			
	N	139	124	113	139	110	102	136			
宗教	r	.094	.107	.123	-.221*	-.012	.227*	-.343***	-.209		
	N	92	82	77	92	85	80	92	81		
婚姻狀態	r	-.043	.150	-.230*	-.100	-.228*	.272**	-.226**	-.018	.260*	
	N	152	135	123	152	120	110	149	125	87	
樣本類型	r	.004	-.110	.086	-.212**	-.084	.465***	.260***	-.321***	.049	.144
	N	232	155	140	229	132	120	176	139	92	152

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