

中國醫藥大學

碩士論文

編號：IEH-1712

世界衛生組織生活品質問卷用於

自填模式與面訪模式之比較

**Comparison of self-administration and
face-to-face interview modes on the
WHOQOL-BREF questionnaire**

所別：環境醫學研究所

指導教授：梁文敏

學生：許雅琳 Hsu Ya-Lin

學號：9465012

中華民國九十六年六月

誌謝

本論文承蒙指導教授梁文敏老師於論文寫作期間的細心指導，並不時花費時間進行討論與指導，才得以有今天的成果，梁老師在平時的討論中，有如良師益友般的教導，不僅僅於論文的指導，在平常日常生活當中，也常常提供她的經驗談，實讓我獲益良多。並且也要感謝中醫部陳建仲主任，不論是對於研究執行經費以及人力的支持與指導，本研究收案才得以順利完成，在此謹致本人最深及由衷的感謝。

在兩年的研究所的過程當中，特別感謝所上老師們的細心教導，另外，研究期間，多虧了研究所兼辦公室同學泰進、懿諄學姐、宏偉學長以及哲璋學弟，在學業上互相學習，並在精神上彼此鼓勵與支持，另外，也要感謝在健檢中心幫忙收案的雅慧、寧玗和小杜妹妹，多虧了你們辛苦的收案，這個研究才得以順利完成。

最後，我要感謝背後支持我的好友以及家人，我的好友佳萱、麗杏姐、淑玲姐，真的謝謝妳們的貼心關懷，我要感謝我最愛的父母、哥哥、弟弟和政國，你們對我無限的支持和鼓勵，讓我沒有後顧之憂，並陪伴我一路走來，謝謝！

雅琳
謹致於中國醫藥大學
民國六十九年六月

摘要

研究目的：

世界衛生組織生活品質問卷(WHOQOL-BREF)原先設計用在自我填答施測模式，但實際在臨床應用上，卻有許多限制，尤其在讀或寫不方便的族群，有鑑於此，在臨床上就常常用面對面訪問的方式完成問卷，本研究即在探究 WHOQOL-BREF 實際用在自填模式或面訪模式上的差異。

方法：

本研究在中國醫藥大學附設醫院健檢門診收案，採取交叉研究設計，一共收取212位自願參與者。自願者則先分派為先面訪或先自填兩組，隨後再進行交叉自填或面訪。我們利用古典測驗理論(classic test theory, CTT)及項目反應理論(item response theory, IRT)，看WHOQOL-BREF在使用兩種施測之心理特質及差異性，進而深入探討。

結果：

整體而言，不論施測模式的順序先後，WHOQOL-BREF 全部範疇的面訪得分均高於自填得分。

比較兩組的再現性，結果顯示先面訪組內在等級相關係數(ICC):0.71~0.8，較先自填組 ICC: (0.56~0.77)穩定。再區分年齡、性別來看再現性，發現男性、年齡大於 65 歲者其 ICC 指標較不穩定。不論 CTT 或 IRT 來看問卷各範疇內部一致性，WHOQOL-BREF 之生理、心理與環境範疇信度均不錯，但在社會範疇之中，CTT 與 IRT 信度均較差，結果與 WHOQOL-BREF 其他相關文獻吻合。

項目反應理論的結果顯示：在兩種施測模式下，反向題”需要醫療應付日常生活”、”負面感覺”和”疼痛會妨礙需要做的事”，不符合單一向度的假

設，反向題會降低其單一向度指標。另外，在”性生活”對於老年人不符合單一向度的假設。

不同的施測模式下，題目相對的難度順序亦不同，變化較顯著的題目如：在反向題”負面感覺”男性在自填較面訪容易傾向有負面得分，女性則相反；老年人在”接受自己的外表”，自填較面訪容易傾向有不滿意的感覺；而在字面上較抽象的題目，需要訪員進一步解釋題意，也會造成受訪者對不同施測模式之間的感受不同。

試題差別功能結果顯示在”身體疼痛會妨礙您所需要做的事”、”性生活”和”方便得到資訊”則在不同性別上兩種模式均出現差別，這3題在不同施測模式之間的特質呈現不穩定。

結論：

WHOQOL-BREF用在面訪與自填兩種施測模式之下，面訪分數均高於自填。施測模式的不同其測量結果隨人口特質不同亦有差異。比起古典測驗理論，項目反應理論提供了另一種方式，讓我們進一步從試題層面看到兩種施測模式的不同。

Abstract

Purpose:

The Taiwan version of the WHOQOL-BREF has been designed mainly for using as a self-administered questionnaire inheriting difficulties in clinical application for subjects of limited in reading or writing. The purpose of this study is to test the difference of the face-to-face interview mode and the self-completion mode of administration on the measurement of WHOQOL-BREF.

Methods:

A crossover study was carried out on 212 volunteers from the center of health examination clinic at a teaching hospital in central Taiwan. We divided the sample into two groups, one group used self-completion first, followed by face-to-face interview, another group used face-to-face interview first and then followed by self-completion. The outcome was the difference in WHOQOL-BREF profiles comparing clinical based self-administration with face-to-face interview by using the classical test theory (CTT) and item response theory (IRT).

Results:

All domain scores were higher by interviews than by self-administration. The ICC indices in interview first group (0.71~0.8) were higher than those in self-administration first group (0.56~0.77). In addition, stratifying by age and gender, the reproducibility was lower for the elderly and males. Physical, psychological and environment domains demonstrated satisfactory reliability. However, the social relationships domain exhibited poor reliability.

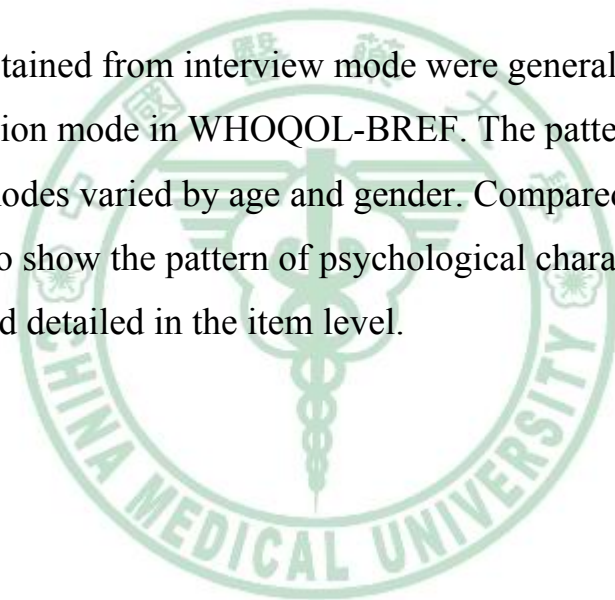
Three items: “need medication” , “negative feeling” and “pain/discomfort” did not show unidimensionality. When we considered the

sequence of item difficulty, for the item “negative feeling” , males in the self-administration mode had lower scores than males in the interview mode; female had opposite result. For the item “accepting appearance” , the elderly in the self-administration mode had lower scores than in the interview mode. For the items, of vague abstract in the wording requiring explanation at interview, the difference of scores between the two modes was more significant.

Significant differential item functioning (DIF) presented in items ” pain and discomfort”, ” sexual activity” and ” opportunities”.

Conclusions:

The scores obtained from interview mode were generally higher than those by self-administration mode in WHOQOL-BREF. The patterns of compared between the two modes varied by age and gender. Compared with CTT, IRT offers a new way to show the pattern of psychological characteristic, which is more structured and detailed in the item level.



目 錄

第一章 緒論.....	1
第一節 研究背景與研究動機.....	1
第二節 研究的重要性.....	4
第三節 研究目的.....	4
第四節 研究問題與研究假設.....	5
第五節 名詞界定.....	6
第二章 文獻查證.....	7
第一節 WHOQOL-BREF.....	7
第二節 Patient Reported Outcome 及相關測量準則.....	8
第三節 自我填答與面對面訪問管理模式.....	16
第四節 不同模式測量偏差的影響因素相關研究.....	20
第五節 項目反應理論看試題的心理特性的應用.....	30
第三章 研究方法.....	33
第一節 研究設計.....	33
第二節 研究對象.....	34
第三節 研究工具的擬定.....	35
第四節 資料收集過程.....	36
第五節 資料統計與分析.....	37
第四章 研究結果.....	41
第一節 基本人口學變項之敘述統計.....	41
第二節 前後測的信度分析.....	43
第三節 組間兩種方式的比較（範疇分析）.....	48
第四節 項目反應理論分析結果.....	53
第五章 討論.....	78
第六章 結論與建議.....	84
第一節 結論.....	84
第二節 研究限制.....	85
第三節 應用與建議.....	86
參考文獻.....	87
附錄一 世界衛生組織生活品質問卷.....	93
附錄二 簡易心智狀態問卷調查表.....	95

Table contents

Table 1. Taxonomy of PROs used in clinical trials.....	10
Table 2. Demographic characteristic of self-administration first group and interview first group.....	42
Table 3. Physician-diagnosed chronic condition distribution in self-administered first and face-to-face interview first.....	42
Table 4. Reproducibility of domain score among two measurements within each group	45
Table 5. Reproducibility of self-administration and interview in domain scores of WHOQOL-BREF for each group stratified by gender, age, education and disease status.....	47
Table 6. Cronbach's alpha, mean score of each item and comparison between self-administration first group and face-to-face interview first group.....	55
Table 7. Cronbach's alpha, mean score of each item and comparison between self-administration first group and face-to-face interview first group for male.....	56
Table 8. Cronbach's alpha, mean score of each item and comparison between self-administration first group and face-to-face interview first group for female.....	57
Table 9. Cronbach's alpha, mean score of each item and comparison between self-administration first group and face-to-face interview first group for age < 65 years.....	58
Table 10. Cronbach's alpha, mean score of each item and comparison between self-administration first group and face-to-face interview first group for age \geq 65 years.....	59
Table 11. Item difficulty, infit statistics, and personal separation reliability of each item among self-administration first group and face-to-face interview first group (arranged by item difficulty within each domain)	62
Table 12. Comparison of internal consistent reliability between CTT and IRT among self-administration first group and face-to-face interview first group.....	63
Table 13. Item difficulty, infit statistics, and personal separation reliability of each item among self-administration first group and face-to-face interview first group for male (arranged by item difficulty within each domain).....	66
Table 14. Item difficulty, infit statistics, and personal separation reliability of each item among self-administration first group and face-to-face interview first group for female (arranged by item difficulty within each domain).	70
Table 15. Item difficulty, infit statistics, and personal separation reliability of each item among self-administration first group and face-to-face interview first group for age < 65 years (arranged by item difficulty within each domain) age 40-64 years, according to the item order of the WHOQOL-BREF	73
Table 16. Item difficulty, infit statistics, and personal separation reliability of each item among self-administration first group and face-to-face interview first group for age \geq 65 years.....	76

Figure contents

Figure 1. The PRO Instrument Development and Modification Process.....	13
Figure 2. Comparison of mean score of each domain in the WHOQOL-BREF among two measurements in self-administrated first group.....	44
Figure 3. Comparison of mean score of each domain in the WHOQOL-BREF among two measurements in face-to-face interview first group.....	44
Figure 4. Comparison of mean score of each domain in the WHOQOL-BREF among the first measurements in both groups.....	50
Figure 5. Comparison of mean score of each domain in the WHOQOL-BREF among the second measurements in both groups.....	50
Figure 6. Comparison of mean score of each domain in the WHOQOL-BREF among the first measurements in both groups for female.....	51
Figure 7. Comparison of mean score of each domain in the WHOQOL-BREF among the first measurements in both groups for male.....	51
Figure 8. Comparison of mean score of each domain in the WHOQOL-BREF among the first measurements in both groups for age < 65 years.....	52
Figure 9. Comparison of mean score of each domain in the WHOQOL-BREF among the first measurements in both groups for age \geq 65 years.....	52
Figure 10. DIF analysis among the mode of self-administration first and the mode of face-to-face interview first in each domain of WHOQOL-BREF.....	63
Figure 11. DIF analysis among the mode of self-administration first and the mode of face-to-face interview first in each domain of WHOQOL-BREF for male.....	67
Figure 12. DIF analysis among the mode of self-administration first and the mode of face-to-face interview first in each domain of WHOQOL-BREF for female.....	71
Figure 13. DIF analysis among the mode of self-administration first and the mode of face-to-face interview first in each domain of WHOQOL-BREF for female age 40-65 years.....	74
Figure 14. DIF analysis among the mode of self-administration first and the mode of face-to-face interview first in each domain of WHOQOL-BREF for female \geq 65 years.....	77

第一章 緒論

第一節 研究背景與研究動機

面訪(face-to-face Interview)方式與自我填答(self-administration)方式是資料收集的兩種主要實施模式(two main common modes of administration of data collections)，針對問卷調查，面訪又可稱為訪談式問卷調查法，自我填答方式則稱為自我填答式問卷調查法。兩者的主要區分可從資料收集過程的管理模式來看，訪談式問卷調查法主要由訪談者來實施問卷的填寫，而自我管理式問卷調查法則主要由受訪者來實施問卷的填寫¹。

面訪是訪問調查者透過口頭交談等模式向被訪問者了解所要的訊息，是一種最古老、最普遍的資料收集方法，其特點在於整個訪談過程是訪問者與被訪問者互相影響、互相作用的過程，訪談者可以創造和諧的調查氣氛，可以獲得較高的應答率，但相對地其結果可能會受到訪談者個人特質、雙方互動的過程等之影響，而影響到測量的品質，其誤差尤其對於測量心理感受的題目特別的敏感，此外面訪花費的人力、物力、財力較大，也比較費時¹。

自我填答(self-administration)方式因由受訪者自己填答問卷，則相對能避免上述所提及的面訪方式所具有的缺點，但由於少了彼此的互動，其缺點是應答率(分為問卷的回收及答題的完整性)降低、尤其是問卷的完整性部份，此外針對不識字的老人、視力狀況不佳者、疾病患者等，因缺少了彼此的互動及關心，造成問卷施測不易實施。故許多生活品質問卷，在資料搜集的管理模式上，原則上均建議採用自我填答方式，而針對不易採用自我填答方式者，則採用面訪或由代理人代為回答²⁻⁴，有關代理人所可能產生的誤差討論常見於文獻⁵⁻⁹。

根據世界衛生組織將生活品質定義為『個人在所生活的文化價值體系中

的感受程度，這種感受程度與個人目標、期望、標準、及關心等方面有關。』。世界衛生組織（World Health Organization, WHO）於 1991 年開始，結合了 15 個國家發展了一份與健康相關生活品質問卷，定名為「世界衛生組織生活品質問卷（WHOQOL-100）」^{10,11}。台灣於 1997 年由姚開屏等人向世界衛生組織取得授權，組成台灣版生活品質問卷發展小組，將 WHOQOL-100 原始問卷翻譯為本國文字，並按其規定先做問卷量尺的發展，進行台灣版生活品質問卷的研究與發展¹²。世界衛生組織生活品質問卷簡明版（WHOQOL-BREF）的問卷題目是由 WHOQOL-100 的 24 個層面中各選出一個題目，並將這 24 題簡明版題目分成四個主要的範疇：生理健康範疇（physical health domain，包含原先的生理及獨立程度範疇）、心理範疇（psychological domain，包含原先的心理及心靈/宗教/個人信念範疇）、社會關係範疇（social relationships domain）以及環境範疇（environmental domain），也從一般性評量中挑選出兩個題目分別為與「整體生活品質」與「一般健康」相關的題目各一題，並且亦提供 2 題針對不同文化特質的本土性題目，此問卷共有 28 題。

WHOQOL-BREF 原先本設計為自我填答(self-administration)方式請患者自行填答問卷，但在實際執行上，尤其是老年人，自我填答的方式不容易執行，其限制在於視力老化、問卷字太小、看不懂題目或時間短促拒絕填答，故針對此族群，多以面訪方式來進行資料的搜集。在過去的經驗談中，面訪方式可能造成系統性偏差(systematic error)，但兩種測量方式所得到的結果差異有多大，所見探討並不多。甚至，有鑑於此，在香港更直接發展 WHOQOL-BREF 的面訪版本，用以對策¹³。而現今在大多數的關於此類研究中，通常直接用面訪方式，或限制研究對象為識字者而採用自我填答方式，或者兩種方式混合使用。這些不同的方式所得到的結果是否代表相同的意

義，鮮少有深入探討，若有也僅止於問卷範疇而非問卷試題。

本研究針對WHOQOL-BREF採取面訪以及自填模式，使用古典測量理論以及項目反應理論分析其差異性，不僅僅從範疇觀點，更微觀至試題層面進行探討。



第二節 研究的重要性

1. 可使研究者更瞭解世界衛生組織生活品質簡明版問卷(WHOQOL-BREF)用於自填模式與面訪模式上的差異，臨床實際使用時之選擇參考。
2. 透過瞭解施測模式所造成差別，對可能造成的偏差，加以調節，對臨床實際應用合併模式的整合提出可能的解決的方案，以提昇測量品質。

第三節 研究目的

本研究目的為探究 WHOQOL-BREF 採用自填模式與面訪模式間的差異，兼用古典測驗理論與項目反應理論的 Rasch 模式分析探討 WHOQOL-BREF 在兩收樣模式下之心理特性差異。

古典測驗理論：

1. 世界衛生組織生活品質簡明版問卷(WHOQOL-BREF)在不同施測模式下，各範疇與各試題平均值、標準差與檢定值。
2. 內在等級相關係數指標(ICC)看同一個人在同一時間點用不同施測模式上的再現性。

項目反應理論：

1. 各範疇試題是否符合單一向度(unidimensionality)的假設。
2. 各範疇試題困難度(item difficulty)在不同施測模式上有何不同。
3. 再各範疇試題看試題差別功能(differential item functioning)。

第四節 研究問題與研究假設

研究問題

- (1)同一個人自填與面訪的組內範疇得分是否有差異？
- (2)同一個人自填與面訪的範疇再測信度是否良好？
- (3)在不同人口特質分層（性別、年齡）下，自填與面訪模式之再測信度是否良好？
- (4) 在同一時間點上自填與面訪範疇得分差異？
- (5) 同一時間點上，在不同人口特質分層（性別、年齡）下，自填與面訪（組間）範疇和試題得分差異？
- (6) 使用項目反應理論看各範疇是否符合單一向度的假設。
- (7) 使用項目反應理論看自填與面訪的心理計量特質？
- (8) 自填與面訪模式下的試題差別功能(DIF)情形。
- (9) 在特定人口族群下不同模式的試題差別功能(DIF)情形。

研究假設

- (1) 使用自填與面訪施測模式在問卷結果中，各範疇得分相同。
- (2) 不同人口特質使用自填與面訪施測模式得分相同
- (3) 針對使用自填與面訪其各試題在量尺上的刻度(item calibration)（或難度）相同。
- (4) 在不同的人口特質下，自填與面訪各試題在量尺上的刻度(item calibration)（或難度）相同。
- (5) 自填與面訪並不會顯示試題差別功能(DIF)情形。

第五節 名詞界定

傳統測量理論：又稱為古典測量理論或真實分數模式 (true score model)，

這個理論是假設任何觀察的測驗分數是由真實分數及隨機誤差兩個成分組成的，亦即觀察分數=真實分數+隨機誤差 ($X=T+E$)。

項目反應理論：在心理計量理論中是較近期發展的一套理論，故又稱為現代測量理論。它是因應古典測量理論的一些限制，而發展出的一系列家族模式，可方便評估模式與資料間之適合度，並常用在教育及心理學上的評估上。項目反應理論主要的假設為一個人對於一個特定的題目答對（期待）機率為人的能力與一個或以上的題目參數的聯合函數。

Rasch 分析：為項目反應理論中的一種。

試題差別功能 (Differential item functioning)：用於決定在不同族群的受測者中測量值是否準確地測量相同的概念。

信度：測量工具中顯出變項誤差的一致性。

第二章 文獻查證

第一節 世界衛生組織生活品質問卷簡明版(WHOQOL-BREF)

本研究使用的健康相關生活品質(health-related quality of life, HRQOL)問卷，為台灣版世界衛生組織生活品質問卷簡明版(WHOQOL-BREF)，其問卷是一個一般性的生活品質測量工具，包含生理、心理、環境與健康等四個向度，具備發展良好之信效度指標¹⁴，版權屬於世界衛生組織及世界衛生組織生活品質問卷台灣版研究小組所有¹⁵。

世界衛生組織生活品質計畫於 1991 年開始，其目標就是為了發展一份國際化且跨文化間可以比較的生活品質量表。世界衛生組織集合 15 國 20 個中心共同發展出 WHOQOL-100 問卷，其內容一共可分為六大範疇 (domains)，再細分成 24 個層面 (facets)；包括：生理範疇 (physical domain)、心理範疇 (psychological domain)、獨立程度 (level of independence)、社會關係 (social relationship)、環境 (environment)、心靈／宗教／個人信念 (spirituality/ religion/ personal beliefs)¹⁶。後來又發展出簡明版問卷 (WHOQOL-BREF)；使獨立程度併入生理範疇，宗教併入心理範疇，各層面一個問題而成。

而 WHOQOL 台灣版問卷發展小組也依照 WHOQOL 研究總部的規定，除了 WHOQOL-BREF 的 26 個題目外，還加上了「飲食」及「面子」等兩個本土性層面的題目，同時也出版了一本「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」¹⁵ 供使用者參考。

第二節 PRO (patient reported outcome)及相關測量準則

本研究為測量WHOQOL-BREF使用兩種施測模式之間的反應差異，問卷的施測方法一直是研究中的重要環節，影響問卷測量的準確性。

在過去，健康相關以及生活品質問卷的建構、評估問卷未予以標準化，學界充斥著信效度參差不齊的問卷，因此，如何得到一個信效度皆佳的評估結果，日益漸形重要，其中包括適當的評估工具、如何使用這些工具，完整詳細的規範是亟需建立的。

有鑑於此，美國食品與藥品管理局FDA(Food and Drug Administration)於2006年2月2日為此提出準則初稿，希望能夠讓醫療界有遵循的標準，包括過去所通稱的「生活品質量表」、「健康生活品質量表」，未來在醫療界都將統一以PRO(患者報告資料：Patient Reported Outcomes)方式稱呼。

「Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims」¹⁷乃因應醫療產業發展而生，其內容包含了引言、背景、PRO的管理觀點、如何評估PRO工具、研究設計與資料分析。

PRO意指「直接」由病人方面所測得到的健康狀況方面的測量，而非經過醫生或其他人解釋所得的結果。臨床研究中，PROs工具(PRO instrument)可以用來測量對於介入一個或多個層面的病人的健康狀況稱之為PRO概念，其概念可分為三種：1.純症狀(purely symptomatic)，例如：單純的頭痛反應；2.較複雜的概念(more complex concepts)，例如：實行一些日常生活的能力；3.極端複雜的概念(extremely complex concepts)，如本研究所測量的健康相關生活品質。

(一) PRO的使用觀點

為什麼醫學相關發展要使用PRO？

1.經常某部分療效只有病人自己知道，例如醫生使用疼痛的藥物治療，要知道病人本身疼痛的強度以及痛苦的減輕有否，答案都在病人本身的感受。

2.病人對療效所提供獨一無二的觀點：PRO可以測量得知病人在他們的治療過程中所想要、期望與最重要的是什麼？要得知病人心中所期待與感覺是非常重要的，例如：對於有氣喘相關症狀的病人，他在臨床上真正有意義的改善應該是在於病人本身因該疾病所帶來給病人本身的衝擊減輕，而不是單單只看肺功能值的改善。

3.正規的評估比起不正規的訪問較為可信：一個結構化的問卷，可以將測量誤差降到最小，並確保其一致性。可以自我填答而完成的問卷，比起經由臨床醫生或其他人來填寫完成來得好，舉例說明，自我填答完成問卷直接反應了病人對於治療之後的反應，沒有第三者的介入觀察再報告，所以相較於一般的觀察報告的測量值更加可信；因為觀察者與觀察者間會有誤差，除非透過規模更大更完整的觀察者（訪員）訓練，才能降低此誤差。

但另一方面來說，PRO也可能因測量工具（問卷）使患者難以理解與填寫，而影響其問卷效度；但儘管如此，如能擁有發展良好且適切效度的PRO工具，它所評估的結果，是可以和專家認定的結果相吻合，所以，這也是PRO在醫學研究中特別適合使用的因素。

(二) PRO的分類(A Taxonomy of PRO Instruments):

PRO 工具所測量的概念非常廣泛，由單一症狀（例如疼痛程度或頻率）到整體的狀況，例如憂鬱，以及較特殊的症狀以及衝擊，像是活動力、以及治療的感覺...等都可以被測量，所以進一步來看，PRO的概念可以是一般普遍性的，也可以是特殊的。

有些屬性較特殊的 PRO 工具會去影響到與其結合的研究，像是療效之類。

PRO 的屬性及其特性如下：

Table 1: Taxonomy of PROs Used in Clinical Trials¹⁷

屬性	型態
欲使用的測量 of the measure Intended use	<ul style="list-style-type: none"> • 定義研究族群 • 評估效益 • 評估不良事件(Adverse events)
測量概念 Concepts measured	<ul style="list-style-type: none"> • 整體健康狀況 • 病徵或疾病的症狀 • 官能狀況，例如生理、心理或社會的狀況 • 自覺健康狀況(Health perceptions) • 療效滿意度或治療的偏好 • 療效評估
試題數 Number of items	<ul style="list-style-type: none"> • 單一題一個概念 • 多題一個概念 • 多題多個層面一個概念
預期測量的族群或狀況 Intended measurement population or condition	<ul style="list-style-type: none"> • 一般狀況 • 特殊狀況 • 族群的特殊狀況
資料收集的模式 Mode of data collection	<ul style="list-style-type: none"> • 訪員面訪 • 自填式 • 電腦自填或電腦輔助自填式
時間與頻率 Timing and frequency of administration	<ul style="list-style-type: none"> • 當事件發生 • 研究中定期 • 療程的開始與結束
計分方式 Types of scores	<ul style="list-style-type: none"> • 某一個概念的程度(e.g., pain severity) • 索引(Index): 單個得分結合了多個相關方面或獨立的概念 • 概況(Profile): 多個分數多個相關層面 • 一系列(Battery): 多個分數多個概念 • 混和的(Composite): 合併索引、概況或一系列的
題目或概念的權重 Weighting of items or concepts	<ul style="list-style-type: none"> • 所有試題和層面(Domains)權重相等 • 試題權重不同 • 層面權重不同
計分法 Response options	<ul style="list-style-type: none"> • 視覺化量表(Visual analog scale ;VAS) • 定錨分類量尺(Anchored categorized VAS) • 李克特尺度量尺(Likert scale)

-
- 等級評價量尺(Rating scale)
 - 狀況紀錄表(Event log)
 - 繪畫量尺(Pictorial scale)
 - 清單(Checklist)
-



(三) 評估 PRO 工具

一份能適切測量的 PRO 工具必須仰賴其發展過程及證明其測量特性。所以在研究初始期間，新的 PRO 工具可以被繼續發展或者當研究者認為其工具不合用、不適切時則繼續修改。下列有五個部分(A-E)分別說明 PRO 評估發展過程：A-概念架構與應用的發展(Development of the Conceptual Framework and Identification of the Intended Application)；B-發展 PRO 工具(Creation of the PRO Instrument)；C-評估測量屬性(Assessment of Measurement Properties)；D-修正現有問卷(Modification of an Existing Instrument)；E-發展特定族群的 PRO 工具(Development of PRO Instruments for Specific Populations)¹⁷

PRO 工具的發展及修改是非現性的一連串的多樣過程事件，有可能同時發生，並且不斷的重複。這個反覆的過程則是一個「循環與對話(wheel and spokes)」的圖示(如圖一所示)，i 為「定義其概念架構的概念與發展」，這個過程對病人來說很重要，決定了欲研究的族群和適用性，所以是假設其相關都在概念之中；ii 為「問卷設計」，涵蓋產生試題、選擇收樣模式、回答選項、問卷的編排、預測試的過程；iii 為「測試其問卷測量特性，其特性包含得分的信度、效度和偵測改變的能力，在這部分當中，也評估收樣方式和回答者的負擔，是否有需要增減或者反向題目，得分範圍的定義，也將定版問卷的版面、計分規則、過程；iv 為「修改問卷」的部分，修改測量的概念、研究的族群、研究的適用性或採樣方式。¹⁷

i. Identify Concepts and Develop Conceptual Framework

Identify concepts and domains that are important to patients.
Determine intended population and research application.
Hypothesize expected relationships among concepts.

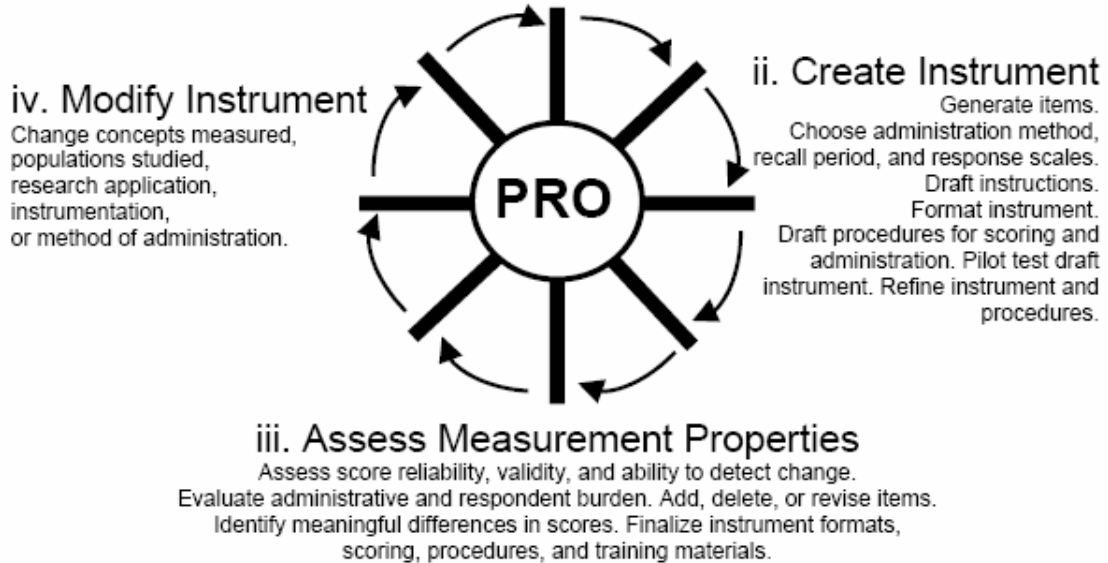


Figure 1. The PRO Instrument Development and Modification Process ¹⁷



(四) 研究設計

研究設計主要集中於對 PRO 獨特性的問題，PRO 指引將其分成六個部分(A-F)¹⁷：

A. 一般研究擬定事項General Protocol Considerations

如果PRO測量的目標是支援所需，我們建議測量PRO概念需要具體的將研究清楚說明：(1) 盲性與隨機分派(Blinding and Randomization)；(2) 臨床試驗的品質控管(Clinical Trial Quality Control)；(3) 防止資料缺失的試驗設計(Designing the Trial to Avoid Data Missing Due to Withdrawal From Exposure)。

B. 測量的頻率(Frequency of Measurements)

PRO評估的頻率取決於疾病自然發生史以及治療的性質，且應該符合其已被證明的測量特性，並且計畫用以資料分析。

C. 研究持續期間(Duration of Study)

研究開始到結束所需要的時間是很重要的評估點，其中要去考慮疾病或狀況的改善時間。一般來說，PRO所評估的時效應比預期測量效性時間長。在對一種會日漸嚴重的疾病研究中，PRO的概念著重在於持續追蹤偵測其效性，而非一時的改變。

D. 多終點(Endpoints)的設計考慮(Design Considerations for Multiple Endpoints)

PRO 的測量終點(endpoint)是根據病人本身的症狀以及能力而評估的，症狀的改善不必然代表功能以及心理狀態的改善；FDA 建議研究計畫應定義其測量終點及準則，以供後續統計分析及結果解釋

E. 規劃結果的解釋(Planning for Study Interpretation)

FDA建議研究者要規劃如何將研究發現做最好的詮釋：詳盡的解釋研究計畫發現。

F. 使用電子式PRO的注意事項(Specific Concerns When Using Electronic PRO Instruments)

電子式的PRO比起過去紙筆填寫的PRO來的容易、簡單，且更受回答者喜愛。但也因為其特性，FDA也建議研究者使用有關電子式問卷注意事項¹⁷。



第三節 自我填答與面對面訪問管理模式

本研究特別將自填與面訪一併探討的原因在於，常常在臨床問卷使用上，遇到病人無法自填，而由訪員面對面訪問完成問卷，究竟會造成什麼誤差？

一般而言，雖然面對面訪問的資料收集方式，比起自填或郵寄方式，受訪者較為喜歡^{18,19}，但是它的花費比起其他方式來說卻較多^{4,18,20,21}，僅管如此，也有研究指出，在有規劃的郵寄問卷或電話訪問條件下，其回收率以及信度也是可以 and 面訪比擬的^{18,20,21}。

自填模式(self-administration)

Ann Bowling將自填模式(self-administration modes)分為三類: (a)將問卷郵寄至受測者，請受測者自行填答(PAPI)後回寄；或是直接將問卷傳遞給受測者自填，填寫完畢後交回; (b)電子化、電腦化輔助自填; (c)電話自動語音系統(ACASI)²²。

自我填答(self-completion)方式因由受訪者自己填答問卷，則相對能避免上述所提及的面訪方式所具有的缺點，但由於少了彼此的互動，其缺點是應答率(分為問卷的回收及答題的完整性)降低、尤其是問卷的完整性部份，此外針對不識字的老人、視力狀況不佳者、疾病患者等，因缺少了彼此的互動及關心，造成問卷施測不易實施。故許多生活品質問卷，在資料搜集的管理模式上，原則上均建議採用自我填答方式，而針對不易採用自我填答方式者，則採用面訪或由代理人填寫。

若問卷內容並非敏感或具威脅性的問題，自填與面訪管理模式所得到的結果應該相差不遠，但是如果問卷內容是較敏感或具威脅性的，則自填方式可能較面訪方式來的好。²²

面訪模式(face-to-face interview)

Ann Bowling將(1)訪問模式分為三類：(a)描述：由訪員面對面訪問，傳統的筆與紙書寫問卷(paper and pencil interview (PAPI); (b) 描述：由訪員面對面訪問，利用電子化(個人電腦(pc)或手提電腦(CAPI))輔助訪問; (c) 描述：由訪員訪問，打電話給受測者訪問訪問或使用問卷或電腦²²。

面對面訪問則是指訪員將設計好的問卷及相關資料，以口述問卷內容的方式當面向受訪者進行訪問。而面對面訪問與其他訪問類型最大的不同在於訪員與受訪者之間的互動模式，訪員可以親自與受訪者實地接觸，受訪者可以看到訪原本人。面訪期間，訪員的表現與受訪者的反應將決定訪問的成敗，經由這樣的互動情境，完成問卷調查的工作。而面訪的優點在於訪問資料的取得較為可信客觀、訪問內容可以深入而追根究底、時間較為寬裕、回收率高且缺失填答者少，故樣本代表性較佳。

但在面訪的缺點有：(1)花費較昂貴。(2)面訪當中也有可能產生偏差(bias)：1. 社會期待性(social desirability)；2. 訪視員偏差(interviewer distortion)：訪視員有可能不經意的會扭曲受訪者的回答，一般而言，訪員在訪視過程若過於主觀，而視對象而定，不去問較困難的問題、或是讓受訪者有不舒服的感覺、不仔細聆聽受訪者的回答或是造成引導式偏差，都是在訪員訓練要注意的。3. 受訪者偏差(false respondents)：訪視過程，受訪者是否真能瞭解問題的意思，而做出正確的回應，若無，則造成隨機性偏差。

另外，決定進行面訪時，訪員的訓練對於面訪相當的重要，也是避免面訪過程中可能會發生的誤差的關鍵。

訪員的訓練過程，有以下重點²²：

(1)對整個研究清楚的描述：訪視原不僅僅只是要做訪視工作，對於研究目的及背景也需能夠瞭解，如此一來，訪員才能知道為何要做訪視工作，也知道訪視工作對於此研究的重要性。

(2)知道誰是研究贊助者：這是訪視員與受訪者的權利，知道為誰工作與誰在做此研究也是很重要的一環，他們有權利知道。

(3)對於此研究有足夠的教導：雖然我們可能沒有足夠的時間將研究方法與問卷架構過程清楚教導說明，但我們還是儘可要讓訪視員知道並且重視研究方法及研究目的，另外，對於問卷本身的結構以及問卷背後的意義，也需要讓訪視員清楚。

(4)清楚解釋抽樣過程的邏輯和過程：訪視員可能不是很能夠瞭解「抽樣」的重要性，所以研究者需將整個邏輯和過程解釋給訪視員知道。

(5)解釋受訪員可能會造成的誤差：訪視員必須要知道，他們可能不經意的會造成結果的偏差，所以他們必須要知道可能會造成問題的所在原因。特別是在研究較敏感的議題上，例如政治立場或道德相關議題，人們有較強的信念，而造成結果的偏斜，這都是要避免的。

(6) 帶領訪員模擬情境：由研究者直接對訪視員演練，將同意書和問卷正式問過，可以讓訪視員直接清楚明白工作。

(7) 教導受訪者篩選過程須知：受訪者是否為研究所需的受測者、是否有符合研究規定、或為研究限制之對象，避免耗時費力。

(8) 訪員實際演練：研究者盡可能的安排幾次演練，讓訪視員練習，從中

可以發現訪問中可能會發現的困難，進一步去討論，藉此訪員可以在真正去訪視前更能熟練。

(9) 管理：為了確保研究資料的品質，研究者的管理是必要的，研究者可以抽選部分資料，再行確認，或是建立訪問後的追蹤機制，都可以確保資料品質。

(10) 解釋研究進度表：訪員必須知道，未在預定時間完成進度會對研究造成影響的重要性，所以讓訪員清楚研究進度也是訓練訪員的一環節²²。



第四節 不同模式測量偏差的影響因素相關研究

接下來，我們針對問卷施測時，可能產生的各種偏差討論。Ann Bowling²²對於各種採樣方式，提出各種的非測量誤差與測量偏差。非測量誤差的產生，是由於研究設計上、抽樣和抽樣架構、受測者的不回應和試題漏答；測量偏差是發生在研究工具上和資料收集的過程所產生的。不管在哪一種收樣模式當中，都可能避免不了這兩種誤差的產生。

非測量誤差(Non-measurement error)

1. 收樣的範圍(coverage for sampling)

所有的方法都需要最新的收樣訊息和在抽樣前將訊息傳達給受測者知道，以確保涵蓋目標族群的完整性和其可能帶來的偏差形式。在屬於訪問和郵寄性質的研究中，相當依賴完整和最新的郵寄資料，若資料不完整或沒有及時更新，可能會造成選樣上的偏差。

電話訪問的研究中通常是隨機選號而播出的，因為在我們的電話簿中並沒有列出號碼使用者的姓名。但這也因電話的回答與否的限制而造成選樣偏差。

電子式的研究則限制受測者需要有個人電腦，電腦必須有網際網路功能和個人電子信箱，這就造成了我們選樣的偏差，因為沒有上述配備的人無法做電子式研究。而這些爭議在方法學的書籍上通常會有解決辦法而這些爭議在方法學的書籍上通常會有解決辦法²³。

2. 回應率(response rates)

在比較不同的收樣方式所造成的回應率在方法學的研究中都有比較，而在大多數的研究中，指出在郵寄研究中特別容易造成回應率偏低²⁴。

對於不回應的主要原因包含回應者不願意參與研究，研究者沒有辦法聯

繫到回應者（例如：做電訪時回應者剛好不在家）和受測者溝通能力上的障礙（例如：閱讀與寫字的能力、認知退化）。進一步探討，受測者不回應受到收樣模式的不同而影響（例如：如果受測者在寫字有困難時，顯然的在郵寄問卷的方式上是行不通的）。

在研究當中如果回應率低的話，尤當不回應族群與回應族群之間，他們具有不同的特徵的時候，潛藏著很大的危險，因為很有可能影響研究的信度，造成研究的偏差，減弱了其研究結果的外推性；即使資料的品質在獲得的過程中是好的，有偏差的樣本族群較難以去預估我們真的想知道的目標族群。

同時也有不少的文獻報告關於不回應與回應的報告，在那些文獻當中，探討回應者與不回應之間造成的不協調性與不確定性^{25,26}。面對面訪問研究比起郵寄問卷自填，長期以來被認為有較高回應率的。一個友善親切的訪員，是可以促進受訪者回答，進而增加回應率的。

但在1990年代之前，也有文獻探討特別指出，在郵寄問卷自填模式當中，能給予受訪者最少兩次的提醒填寫，其回應率幾乎可以達到與面訪時相同(85%)²⁶⁻²⁸。

但在近十年，幾乎所有的方法其回應率都在降低²⁹。有許多的研究都報告過其在電訪、面訪和郵寄問卷時回應率的差異³⁰。Sykes and Collins³¹在1980年在英國時曾做過大樣本的研究：有兩個關於社會態度、一個生活形態的研究，都直接比較電訪和面訪的回應率。

電訪的回應率大致上都比面訪低。也有研究指出，在自填郵寄問卷的方式，其回應率和完成率高於在醫院分給病人自填³²。

不同的收樣方式用不同的順序施測，回應率也會不同。Harris³³嘗試針對住院病人隨機郵寄問卷，並隨後電訪完成其無回應者，比較先用電訪，在

隨後郵寄問卷完成無回應者，報告指出，先電訪者其回應率與答題率均高於先郵寄問卷者。

3. 試題回答率(item response rates)

有較多的報告指出：郵寄較於面訪或郵寄較於電訪³⁴，其試題的缺答率均較來的高³³。

De Leeuw³⁵與 Zouwen的後設分析結果也發現，因為進行面訪訪員面對回答者的時候，相對可以給予回答者些刺激，並且訪員也可以掌控當時的情況，進而增加受測者的試題回答率。

另外，電子、電腦或電話自動語音系統問卷因程式設計防止跳題回答的關係，所以也較不會有試題缺答的現象，但是這個方法並沒有辦法預防填答者中途結束不填寫的現象。

即使這樣，所有的收樣管理模式都沒有辦法避免填答者中途不填寫而離開，雖然這樣的情況，相對於面訪時比較少會發生，但是仍然是避免不了的。

尤在電訪、電子式或自動語音系統³⁶其過程太過冗長以致於中斷受訪者正在做的事，這種情況就特別容易發生。

另外，電子式的回答率仍然高於傳統用紙筆填寫問卷的試題回答率。

Johnson³⁷在英國所做的關於性行為的研究中指出，關於電子式自填問卷比起紙筆自填問卷，電子式的回答率來的較紙筆填寫問卷的方法高。

Tourangeau³⁸也支持電子式的填答較傳統紙筆的填答有較高的試題回應率這個結論。

測量誤差(Measurement error)

1. 社會環境以及偏見的影響 (the social setting and bias)

標準化測量的目的，除了有好的研究設計之外，並且希望可以達到問題

對不同的受測者、文化差異、社會階層、不同的語言，都具有相同的意義，並且對於這些造成的影響，可以合理化的解釋。

在實際資料收集的過程當中，也會受到問卷和回答者交互作用(interaction)的影響，甚至電訪或面訪中，也會受到訪員的影響。

這樣的交互作用不管在訪問或是自行填寫，其實是很自然卻也不可避免的。所以在不同的管理模式之下所造成的資料結果呈現的差異，我們是可以預先假設並盡量避免的。雖然不是所有的研究調查者都可以去發現其回答者在回應上的差別，但是理論上，我們知道各種管理模式差別的所在之後，在臨床實際應用上就可以合併使用的。Tourangeau³⁹等人認為，對於所採取的模式（包括在訪問在內），對於其所可能造成的偏差，加以調節，也可以掌握到問題的狀況和維持資料的品質^{22,40-43}。

關於填寫問卷的步調和填寫問題因順序而產生的效應，如何控制與處理²²？

1. 利用電腦或其他自填問卷的方式會比起面訪更能將填寫時間的步調放慢，因而給予回應者更多的時間思考，因此產生的回應更具準確性。
2. 面訪的優點在於問問題時，訪員可以控制、掌握問題的次序及步調。
3. 電子式和自動語音系統訪問方法的流程設計，為防止跳題回答的情況，程式設計上通常填寫一題完畢後，下一題才會出現，但也因如此，填答者無法檢查剛才填寫過的答案是否正確與否。
4. 紙本的自填問卷因為沒有像訪員/程式控制問卷的填答，所以填答者可以瀏覽過整份問卷，調整他們的回答。

2. 社會期待性偏差 (Social desirability bias)

面對面訪問過程當中，人與人之間的互動產生社會相互作用，導致回答者在回答問題時，考慮社會規範與社會期待，導致社會期待性偏差(回答者的回答為符合社會期待的回答)，社會期待性的回答會使研究問題的結果高

估，同時也會低估反社會性的行為，不論是社會期待性的回答或是反社會性的回答，都屬於社會期待性偏差。

為了降低社會期待性偏差所產生的問題，除了保證資料的機密性和匿名之外，（雖然敏感的題目會引起受訪者的防備心進而降低其答覆率）⁴⁴；對於關於有可能造成社會期待性的題目，我們可以檢查所回應的答案有沒有違反事實、有技巧性的間接詢問問題。

對於後者這樣的一個技巧，回答者將被紀錄在一對的題目，一題是較敏感的題目，另一題則不是，而答題時則隨機先回答其中一個問題，再將另一題完成，整個過程當中，訪員是看不到其回答的結果。

為了要能推斷其反應是否會造成社會期待的影響，我們必須要知道其反（非）社會期待所回答的另一個來源族群的分佈，要一個真實且隨機的過程需要篩選過題目，並且再從其中抽樣比對⁴⁵。

比起自填（例如郵寄自填），回應者通常在面訪和電訪時，在於回答中表現出更多的正面與社會期待性的回答^{38,46,47}，甚至於面訪時的環境也會影響造成偏差(contextual effects)^{40,48}。

因此，評估帶有正面性的健康狀態、健康相關生活品質議題時，都是帶有社會期望的行為與活動的內容，在評估反社會期待性行為時（例如抽煙），採取面訪法或電話訪談時比起自填法，更容易低估了其反社會性的結果^{19,20,40,48-52}。

所以敏感的健康相關問題在面訪或電訪中是較容易被低估的（例如前列腺疾病、泌尿道症狀）^{53,54}。

但是也有小部分相關研究報告指出⁵⁵⁻⁶⁰，採用隨機抽樣的方式進行面訪法與自填法，其結果並無異。

Williams⁶¹在進行HIV的相關危險行為問題研究時，隨機分派的交叉研

究設計用於有聲電子輔助訪問(audio-CASI)和面對面訪談兩種方法，隨後再進行交換方法。他們的報告指出，收樣模式的不同並未造成其效應。但這份文獻也指出，採取相同的模式比起不同的模式^{35,48,61}，其結果上的差異會來的較少，這結果建議需特別注意在混用模式的研究設計上（例如先以面訪為基準，再郵寄追蹤）。也有其他非隨機化的研究報告也指出，電訪與面訪進行選票和酗酒的相關議題，其結果並無差異^{62,63}。

再者，Leeuw³⁵的後設分析結果發現，有小部分的差異在於電訪和面對面訪談中，而造成社會期待性偏差。

但是Evans⁶⁴卻指出不一樣的想法，交替採用面對面訪談和電化訪問於同一組人，發現年齡大於60歲者在電訪中更易顯的焦慮或憂鬱。

Fouladi⁶⁵也比較自填模式中的紙筆自填和電子式自填，問卷中的情感功能和少部分的回應有發現到差異處。

類似的情況也在不同的自填研究中（電子郵件、郵寄、直接分派給病人填寫）報告中，發現對於健康相關的滿意度與評估的研究，其結果並無明顯差異^{32,66}。

以上的研究結果並非全部一致，讓人不禁聯想到非實驗設計所帶來的效應影響。對於回答形式所造成的差異可能歸因於選擇性偏差和回應偏差，或者是兩群人真實存在著不同造成⁶⁷。

3. 默許偏差(Acquiescence bias)

任何訪問的過程中，回答正向的回應，比起自填的狀況，要來的多出許多，這可能起因於回答多次的「是」或者是默許偏差：從文化基礎與他人同意一致的趨勢，因感覺填答「同意」比較填答「不同意」來的容易。

雖然這樣的情形，也有可能也有可能出現在自填模式當中，但是這樣的情況比起訪談法還是較來的少，在古典心理學試驗中有證據顯示在自填問卷

當中因「容易」，而潛在造成其偏差，因回應者有傾向選擇最靠近問題的選項。

通常我們會嘗試的去將其回答選項轉向以供受測者回答（例如：從強烈的同意-強烈不同意轉成由從強烈的不同意-強烈同意）⁶⁸，因在古典實驗 classic experiments 證據顯示，在不同的模式中，其回答產生「yes」的情形是不一致的。Nicholaas⁶²指出尤在面訪及電訪的方法，證據發現比起其他的方法，更容易產生默許偏差。

De Leeuw³⁵的後設分析中並沒有發現在郵寄、電訪和面訪的情況之中有其差異，但在De Leeuw更早期比較電訪與面訪研究中，後設分析中結果發現了默認偏差的證據，其中較多許多模擬兩可（回答「不知道」或沒有回答）和許多極端的回應出現在電訪當中。

4. 訪員偏差 (Interviewer bias)

訪員的存在會讓回應者分心。如果在訪問的情況下產生過多的社會期待性偏差的答案，這有可能是因為訪員所造成的^{23,69}。

除此外，訪員可以在他們的能力可及之下，盡可能的維持中立的聲調、形象，使用此法可以幫助回應者回想、以及記錄其回應。仔細的訪員訓練和定期監督管理也可以降低訪員偏差，使用訪員間信度分析(inter-rater reliability)可以檢測是否有訪員偏差的存在。而自填模式則沒有訪員偏差。

5. 選擇回應的次序 (Response-choice order: primacy and recency effects)

收樣模式中有關回應選擇所造成的偏差目前已經被報告研究。

回應者在訪談中聆聽問題後，如能在心中已有答案，並且他們能夠由他們自行決定其選擇這是最理想。但其實這是有難度的任務，尤其是在電訪中因看不到對方，而使得訪問過程中感到倉促而特別有時間壓力，Knauper⁷⁰報告指出電訪時間在一個問題的問與答的完成不超出1分鐘。

造成偏差的原因在於，當問卷試題呈現視覺化時（例如自填問卷），回答者傾向於勾選他眼前第一個出現的選項，這個就是primacy effects。會發生這種情況的原因在於回答者選擇選項的慣性，當他看到第一個選項，而如果這個選項也是他所同意的話，他就會去選擇第一個選項，而不會在進一步去看其他的選項。

再者，當問題的呈現是呈口語化的情況時（例如面對面訪問或電話訪問時），回應者則會傾向於回答剛剛訪員所提供的最後唸出的選項（因為最先想到的），並且最後一個選項他們也同意的。這個情況我們可以說是「最近的效應」導致^{46,71-73}。

這樣的情況也有研究證據見於電訪比較面訪的情況下，受測者在電訪中更容易選擇較極端的答案，例如「非常滿意」相對於「極滿意」，他們則選擇訪員最後念出的「極滿意」^{62,73}。

有大樣本的研究在後設分析中(meta-analysis)發現，在年齡大於65歲以上的老人^{70,74}，有顯著的「最近的效應」發生，有可能是因為認知程度隨著老化影響所導致⁷⁵。

其他的研究報告也顯示，老年人和年輕人的回答會因研究問卷的特徵影響而導致差異⁷⁶。

6. 回憶效應 (Recall effects)

回應者在答題時需要時間回想相關的訊息，所以，管理模式的相同也會造成影響。

在視覺化的紙本自填問卷當中，當回應者對於問題有所疑惑時，訪員可以立即協助回應者，客觀描述問題，誘使回應者回想相關訊息，而後自行判斷填答問題，相反的，在自行填寫模式當中，因沒有訪員立即協助，所以填寫問卷時若有疑惑時，回應者可能用其他的方式來填答問卷²³，但在這過

程，我們是無法確定回應者是真的回想到我們要問的概念。

在自填模式當中，由回應者自行判斷關於問題的相關訊息，和哪一個是最好的回應，並不一定是指同一件事。

然而在非實驗研究設計中結果指出²⁰，其採樣模式的不同，的確會造成回應者在回想時的不一致^{20,77}。

7. 言語的長度 (Length of verbal response)

電話訪問的時間比起面訪所需的溝通時間，因情境較倉促的關係，所以需要更短的時間來完成問卷³⁵，電訪法比起面訪法，問題的開始與回答的結束，都較為簡短³¹。

De Leeuw³⁵認為在面訪當中比起電訪會得到更多的訊息，原因可能是因為電訪法中只有言語的交流往返，互動較少，導致回應者回答問題更加快速，而完成問題的時間則較少許多，所以，在面訪當中，是較可以接受較長時間的過程。

8. 敏感的問題 (Sensitive information)

問卷採取自填模式比起面對面訪問和電訪，更可以增加回答者對於回答敏感题目的意願。而為了削弱其社會期待性效應，匿名的方式多用在郵寄問卷研究當中，因過去有報告指出，對於較敏感的健康和行為相關議題，其敏感議題的回答是容易高估其正向行為的^{4,39,78,79}。

因此，首要需要能知道何種是較敏感的訊息和察覺偏差的方法是很重要的，這其中和降低社會期待性的方法有些雷同，所以，最好一開始就能夠降低其潛在的偏差⁷⁸。

敏感的題目最好是用較客觀的方式來問，自填法是大多研究在用的⁸⁰。

其中較好的自填法通常是用電子聽力自填系統或電腦輔助自填系統，因這樣的系統可以讓回應者直接將答案輸入在電腦中，保有更多的隱密性。

由於文獻中並沒有確定在這些模式當中他們的差異與敏感問題的效度²²。也有一些研究報告顯示，他們在比較電訪和面訪之間，電訪較能誘使回應者更坦白的回答^{31,82}。

但Groves⁸³報告顯示比起電訪，個人的面對面訪談更能夠讓回應者更愿意回答關於競爭、收入或者禁藥使用的問題。

但Smith⁸⁴ 和Tourangeau³⁹他們的觀點卻與這個結果不符。

9. 回應者的喜好 (Respondent preferences)

在過去，回應者的偏好效應仍未知的時候，就有研究結果報告回應者喜好面訪甚於電訪^{62,67}，而喜好電子式自填甚於紙筆自填⁸⁴。

在國外，電子式問卷的填答上，對於年齡65歲以上的老人⁸⁵的接受度也是很高的。



第五節 項目反應理論看試題的心理特性的應用

項目反應理論 (item response theory, IRT) 乃因應過去古典測量理論 (classical test theory, CTT) 的限制而發展出來⁸⁶⁻⁸⁸。過去CTT的缺點在於(i)無法將受測者表現特徵和題目的表現特徵分開看,受測者的能力就是等於題目的得分,而在IRT之中,題目參數與潛在能力間是獨立的。(ii)過去我們知道,同一個人,做不同份難度不一的問題,得分當然可能有所不同,表面上的得分看似程度有差異,但實際上都是同一個人的程度,而在CTT無法解決因問卷不同,而造成表面上程度的差異,IRT模式的表現是從觀察題目反應層級,而不是從測驗分數反應(加總後的分數)的層級,(iii)經由IRT所提供的訊息可以用來評估每一個題目對於所測量概念的貢獻,(iv)可用在偵測不同次族群(subgroups)間的試題差別功能(differential item functioning, DIF)或題目偏差,(v)可用來產生簡短,複本及特定形式的問卷,並且(vi)儘管研究對象回答不同的問卷也可以將研究對象的分數等化(equating)⁸⁹。

IRT分析可以提供的檢驗量表的心理特性和量表的品質,IRT提供一個數學的基礎來代替傳統的方法,將量表的序位資料轉換為等距資料,並且提供有關這些資料是否有達到測量被接受的指標。

這樣的分析可以提供量表的訊息量有:這個活動力是否適配這個潛在特性?訪員如何使用這個量表?這一系列的活動(試題)是否可以定義在這個潛在特性之中

所以項目反應理論不僅提供一個數學的基礎來代替傳統的方法,並且當其他非心理學者努力地將他們的研究概念化成可操作的測量時,他們可以很容易地了解,但是實證調查需要有一個明確的理論架構作為基礎並據以檢驗找出適當模式整合,而不是任意找一個進階的模式來處理就可以了⁴⁰,因應

以上的優點項目反應理論現在已被廣泛地用在主觀的健康測量上⁹⁰。

Wang 等利用多向度 Rasch 模式調整 WHOQOL-BREF 各範疇的相關性分析台灣國民健康局之國民健康調查資料庫，來評估建構效度並且改善問卷的信度及估計的精確度⁹¹。DIF 分析被用來評估模式-資料的適合度，在不同性別及年齡分層間題目難度是否會有顯著的不同，當不同分群中題目難度的估計值最大差距大於 0.5 時表示有 DIF⁹¹。

本研究，使用項目反應理論中的試題困難度來看不同收樣模式間的試題困難度表現與適配度指標，另外，我們也用試題差別功能分析看不同收樣模式的造成同一族群差異的心理特質。

在過去，也有研究做過問卷跨文化時，比較其心理特性差異，其中就使用 Rasch 分析中的 DIF 功能，比較兩組在同一族群的人，使用荷蘭板和英文版的差異⁹²。

近年來，有相當多的研究報告使用 IRT 的方式來測試跨文化的問卷試題，不在拘泥在問卷傳統信、效度的特質，而是將試題等化，各個題目擁有不同的困難度，以往古典測驗理論是沒有辦法將試題區別開來。

即使對什麼樣的 DIF 的定義才較適當？也有很多爭辯存在。目前有個關於測驗公平性問題的看法認為：「在某個試題上，如果多數族群和少數族群的平均表現有所不同的話，該試題便顯示出具有 DIF 的現象。」其實，這種看法也有個缺失，那就是未考慮其他影響變項的可能性，如：原本這兩個族群的能力就有所不同，因此才導致他們在某個試題（或某份測驗）上表現不同。

目前，比較被心理計量學者所接受的 DIF 的定義為：「來自不同族群，但能力相同的個人，如果在答對某個試題上的機率有所不同的話，則該試題便顯現出 DIF 的現象。」有了這項定義，試題反應理論(IRT)很自然的提供

一個研究 DIF 的架構，因為試題特徵函數正可以說明答對某個試題的機率，是與受試者的潛在能力和試題的潛在特徵有某種關聯存在。因此，DIF 的定義可以被寫成下列的操作型定義：「某個試題特徵函數如果對不同的族群而言都不相同的話，則該試題便顯現出 DIF；反之，如果跨越不同族群的試題特徵函數都相同的話，則該試題便不具有 DIF。」本文即談論試題反應理論對診斷試題偏差（或說試題 DIF）的各種方法，並舉例說明它的用法。

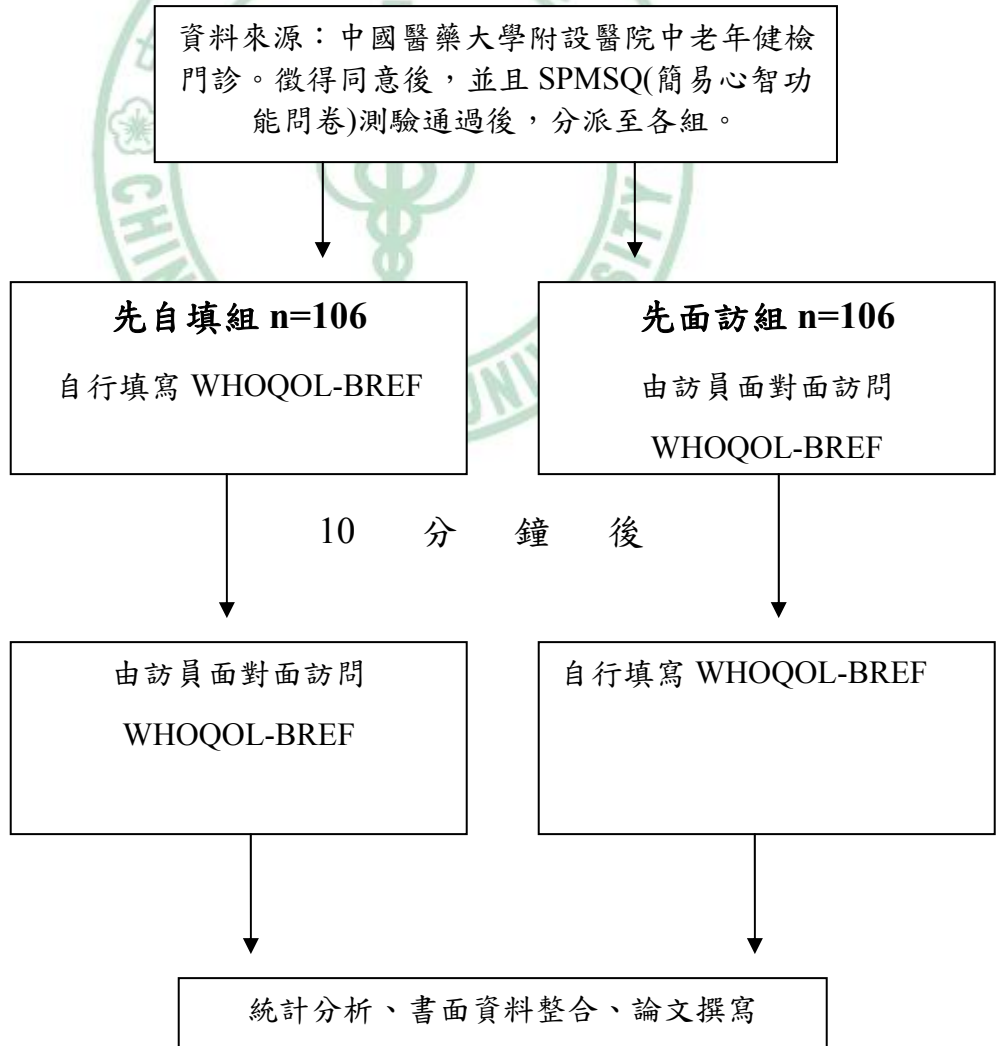
許多的跨文化研究更應用項目反應理論來看其相等性。Roorda將英文版的 Western Ontario and McMaster Universities osteoarthritis index (WOMAC) 問卷翻譯成荷蘭版後，除了用傳統方法看其信效度外，也使用項目反應理論看其跨文化轉譯後的問卷，看荷蘭版試題的適合度指標是否仍符合原版問卷單一向度的假設，並進一步再比較英文版與荷蘭版的 DIF 情形，來檢查在不同的語言對同一族群，題目的重要性是否有顯著不同⁹²。

現今 DIF 功能大多用在跨文化的量表研究比較相等性當中，鮮少用 DIF 功能來看收樣管理模式之間的差異，在過去，多用傳統理論 (cronbach alpha, internal consistent reliability, factor analysis)，未曾有過用探討心理特質的項目反應理論來比較收樣模式間的差異。

第三章 研究方法

第一節 研究設計

本研究為一橫斷式研究，採用交叉設計並便利取樣，將受測者分成兩組，第一組先自填 WHOQOL-BREF，而後過 10 分鐘再由訪員面訪 WHOQOL-BREF，則算是完成；第二組先由訪員面對面訪問 WHOQOL-BREF，而後過 10 分鐘再由受測者自填 WHOQOL-BREF，兩組在填寫 WHOQOL-BREF 之前，都會由訪員測驗簡易心智功能問卷(short portable mental state questionnaire; SPMSQ)，符合標準後再納入研中，訪視期間一併記錄受訪者基本資料以及過去病史，以便研究未來分析，本研究架構如下：



第二節 研究對象

本研究對象為中國醫藥大學附設醫院健檢門診之成年人與老年人。健檢門診為中央健保局健保給付的項目，並又分為兩個部分，一是成年人健康檢查，接受服務對象為四十歲以上至未滿六十五歲，每三年給付乙次，六十五歲以上，每年給付乙次，二是老人健檢，接受服務對象則需設籍台中市(住滿6個月)並年齡需滿70歲(民國21年12月31日以前出生者)，方可參加。對象為參與本研究的所有成年人與老年人均有加入全民健康保險。因此，本研究之基本資料與個人病史資料均與健康檢查登記資料同步。

本研究收案其間為2006年三月至2007年四月，篩選年齡為40歲以上之對象，收案個數一共為212位，男性108名及女性104名。而研究期間，如因不識字、視力模糊、聽力減弱...等因素無法完成兩份問卷者，或填寫SPMSQ錯誤達兩題以上，以及不同意參與本研究者，均不納入本研究中。

在研究過程中，一共有三位訪員，訪員訓練時間長達3天，並訓練其問卷問題台語口語問法的標準，使其過程盡其可能標準化。

第三節 研究工具的擬定

本研究使用SPMSQ (Short Portable Mental Status Questionnaire) 及 WHOQOL-BREF 臺灣簡明版，2000 版問卷進行施測，分別介紹如下：

(一) SPMSQ (Short Portable Mental Status Questionnaire)。由Pfeiffer^{93,94}發展，簡便的評估認知功能，本研究用以測量老人認知狀況是否正常。此問卷共10 題，內包含定向力、個人史、最近記憶及計算力，答錯2題以上則視為異常^{93,95}。

(二) WHOQOL-BREF 台灣簡明版

「世界衛生組織生活品質-簡明版問卷：台灣版」WHOQOL-BREF 台灣版問卷¹⁵ 共由28題目所組成，其中有2題是屬於測量整體生活品質及一般健康的題目。其餘26題主要分為四個範疇：生理健康範疇（physical health domain）、心理範疇（psychological domain）、社會關係範疇（social relationships domain）以及環境範疇（environment domain），而在問卷計分方面，問卷中所有題目皆是採用五點式量尺計分。本研究在分析及結果的呈現上，所有反向題目皆經轉向，分數愈高，代表患者的生活品質愈好。範疇分數的計算為範疇中所有題目的平均值乘上4，範疇的得分範圍為4~20分⁹⁶。

第四節 資料收集過程

針對健檢中年、老年人，徵求其同意後填寫同意書，先由訪員以面對面訪談的方式進行SPMSQ再分派到先自填組或先面訪組。先自填組則先完成WHOQOL-BREF自填式問卷，再由訪員進行面對面訪談的方式進行WHOQOL-BREF 的施測。先面訪組則反之。



第五節 資料統計與分析

以Excel將資料建檔，並以SAS 8.2版統計分析軟體進行資料分析，以雙樣本t檢定先自填組與先面訪組比較。然而，對於敏感議題相關的研究，最好要有一正式評估代表性的樣本需要考量到人口基本變項（例如年齡、性別）。過去對於一些較敏感的議題，我們在對性別、年齡分層評估其關係⁹⁷。

由古典測驗理論的結果觀察到面訪範疇得分與自填平均得分的差異，再細切至看單題平均值的差異。再用項目反應理論分析其試題，看在兩種管理模式間的試題難度順序是否有差異，分別去看兩種模式在試題之中的感受性差異。

內在等級相關係數分析指標

所以我們將在考量收案的人口基本變項，進行內在等級相關係數（intraclass correlation coefficient, ICC）分析指標，進而評估組間每個範疇的再現性(reproducibility)⁹⁸，進行面訪與自填問卷結果之比較。

本研究採用 SPSS10.0 英文版來分析組內相關係數 ICC 用再評估再現性⁹⁹，ICC 透過把同一測量個體的不同的等級的變化性和在全部等級和全部主題的總變動程度以評價其可信度。

ICC 的原理公式如下：

$$\text{ICC coefficient: } \rho = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2}$$

σ_e^2 ：組內變異量

σ_s^2 ：組間變異量

$\sigma_s^2 + \sigma_e^2$ ：組內變異量 + 組間變異量 = 總變異量

內在等級相關係數則根據 Shrout & Fleiss¹⁰⁰, ICC 值小於 0.4 則代表低度再現性(poor reproducibility), ICC 值介於 0.4~0.7 則代表中等再現性(fair to good reproducibility), ICC 值介於 0.7~1 則代表再現性非常好(excellent reproducibility)¹⁰¹。

項目反應理論看不同管理模式下之心理特質

以 IRT 的 Rasch 系列模式中的 Partial credit model 來分析範疇及題目的心理計量特質。分析項目包括：(1)以適合度指標(infit statistic)來檢驗各範疇單一向度的假設。(2)IRT 的信度評估，受測者潛在特質的信度指標與傳統的測量信度的 Cronbach's alpha 相同，用以比較其內在一致性。(3)題目難度的測定值來檢驗題目難度的範圍及順序。(4)以試題差別功能 (differential item functioning, DIF)分析來檢驗性別及年齡間題目難度的順序是否相同。

(1)以適合度指標來檢驗各範疇單一向度的假設

一個範疇中的所有題目應該是要同一個方向或概念。例如，我們如果想要測量生理健康，這個範疇的題目就不應該包含測量心理的題目。假如一個範疇中包含許多不同概念的題目，所得分數的解釋將是困難的。在 Rasch 分析中，題目若缺少單一向度會反應在適合度檢定 (misfit statistic)，即當同一範疇的題目間所測量的並非同一個概念時，misfit statistic 指標將會大於判別標準，本研究中評估單一向度採用 infit 統計量，並以 1.4 作為判別標準，大於此值則代表此題在此範疇不符合單一向度的假設。

(2)IRT 的信度評估

Person separation reliability 是真實的變異數與獲得的誤差的比值。當誤差愈小，這個比值就愈大。

數值的範圍為 0~1 並且結果解釋同 Cronbach's alpha。係數為 0.7（可接受），係數為 0.8（良好），係數為 0.9（極好）。

(3)以題目難度的測定值來檢驗題目難度的範圍及順序

利用 Rasch 模式將人能力及題目難度放在同一個尺度上。當 logit 愈大，表示題目愈難；logit 愈小，表示愈簡單。題目難度範圍愈廣愈好，並且要準確的估計不同能力下的人需要有均勻分佈的題目難度，即盡量減少在測量尺度上題目分佈的間隔（gap）及重疊性（redundancy）。在評估範疇中的題目有無間隔時，我們是利用題目分佈的圖示來看每個範疇中兩個題目的間距若大於一個 logit，即表示有間隔存在。判別一組題目間難度的重疊性時，可利用範疇平均難度（ $\hat{\beta}$ ）的 2 個題目測量值的標準誤（2 standard error of measurement）。分別將兩種管理模式先依題目的難度排序，比較兩種管理模式間的難度順序差異，若是難度順序有差異，則表示受測者填答試題時在這兩種管理模式的感受性不同。

(4) 試題差別功能的評估

試題差別功能（DIF）主要對於不同管理模式間的估計值，來檢查在不同的族群間題目是否有顯著不同的重要性。藉由比較兩個或以上樣本的題目特質，評估 DIF 所要呈現的是題目難度是否具有不變性。這個評估過程中，需要分開估計每一個樣本的題目難度，並且以不同樣本題目的測量值來畫散佈圖。在本分析中我們比較自填及面訪是否有 DIF 存在。

以本研究為例，要評估 WHOQOL-BREF 的題目難度會不會因為管理模式的不同而不同。所以需要以管理模式分群分別估計在自填與在面訪的題目

難度。以自填及面訪題目難度畫散佈圖，若是題目的難度具有的不變性，此兩組人相對的難度估計值將會落在散佈圖中的 45 度角附近。本研究中用 0.5-logit 作為評估的標準¹⁰²以 45 度角的對角線往上下平移 0.5-logit，若是兩組人相對的難度估計值落在此範圍外，就表示此題目有 DIF。



第四章 研究結果

第一節 基本人口學變項之敘述統計

表二為本研究基本人口學特質。總樣本數為 212 人，先自填組男性有 54 人，女性有 52 人；先面訪組男性有 55 人，女性有 51 人；年齡為歲，標準差為 6.04，範圍介於 40~93 歲，範圍介於 40~90 歲。表三為兩組之中，各種疾病的分佈情況，登記的疾病，包含：高血壓、腎病、牙周病、心血管疾病、糖尿病、肺病、B 型肝炎、消化性潰瘍、高血脂病、中風、攝護腺癌、攝護腺肥大與其他。



Table 2. Demographic characteristic of self-administration first group and interview first group.

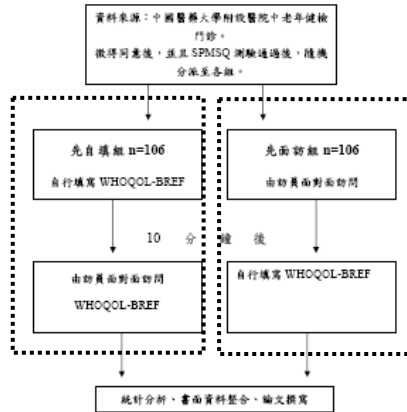
	Self-administration first		Face-to-face interview first		<i>p</i> -value ^a
	106		106		
	n	%	n	%	
Gender					0.89
Male	54	50.9	55	51.9	
Female	52	49.1	51	48.1	
Age					0.89
40-64	49	23.2	50	13.7	
≥ 65	57	26.9	56	10.9	
Marriage					0.21
Married	88	41.5	80	37.8	
Not married	17	8.1	24	11.3	
Education (years)					0.93
≤ 9	37	17.5	36	17	
> 9	68	32.1	68	32.1	
Physician-diagnosed chronic condition					0.34
0	51	48.1	42	39.6	
1	30	28.3	39	36.8	
2	15	14.2	11	10.4	
>3	10	9.4	14	13.2	

based on chi-square test

Table 3. Physician-diagnosed chronic condition distribution in self-administrated first and face-to-face interview first.

Disease	Self-administration first		Face-to-face interview first	
	n	%	n	%
1. Hypertension	21	7.7	32	11.7
2. Nephrotic Syndrome	3	1.1	5	1.8
3. Periodontal disease	19	7	18	6.6
4. Heart disease, Cardiovascular Dysfunction	8	2.9	14	5.1
5. Diabetes Mellitus	7	2.6	7	2.6
6. Pulmonary disease	1	0.4	1	0.4
7. Hepatitis-B	4	1.5	7	2.6
8. Peptic ulcer	7	2.6	5	1.8
9. Hyperlipidemin	5	1.8	11	4.0
10. Stroke	1	0.4	3	1.1
11. Prostate cancer	1	0.4	0	0
12. Benign Prostatic, Hypertrophy	12	4.4	17	6.2
13. Others	20	7.3	8	2.9

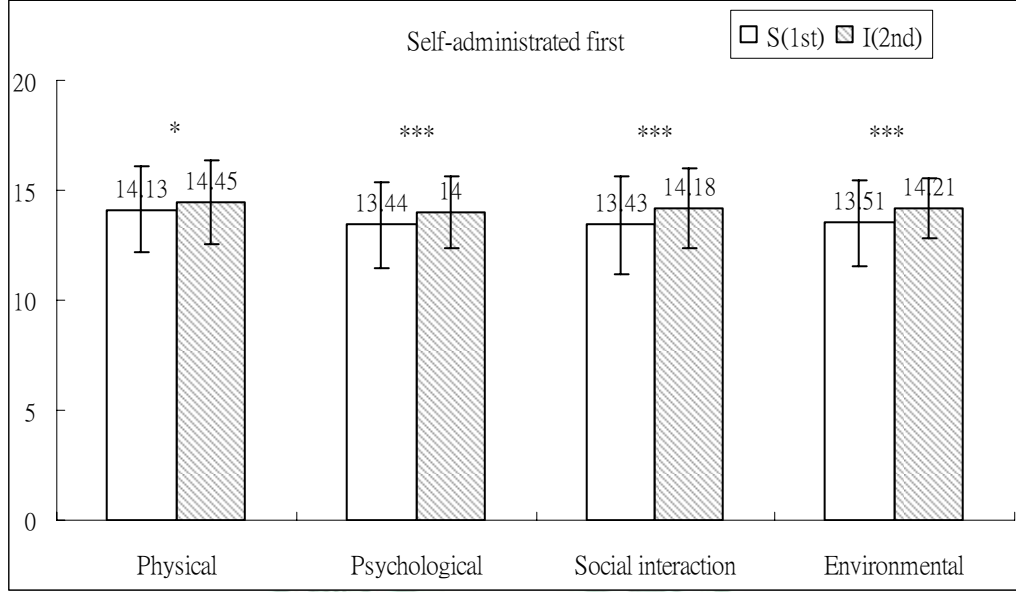
第二節 前後測的信度分析



圖二為先自填組(n=106)之中自填與面訪 WHOQOL-BREF 的分數，自填組的平均得分最低至最高為心理範疇(13.44)~生理範疇(14.13)；面訪組則為心理範疇(14)~生理範疇(14.45)。使用配對 t 檢定比較其平均分數之差異，可以很清楚的看到，先自填組中的面訪組全部 4 個範疇的分數都顯著高於自填組(生理 $p<0.05$ ；心理 $p<0.001$ ；社會 $p<0.001$ ；環境 $p<0.001$)；圖三為先面訪組(n=106)之中面訪與自填的分數，自填組的平均得分最低至最高為心理範疇(13.8)~生理範疇(14.84)；面訪組則為心理範疇(14.27)~生理範疇(15.07)，而先面訪組內則以面訪的心理範疇($p<0.01$)和環境範疇($p<0.01$)顯著高於自填，其他在生理範疇及社會範疇雖無顯著差異，但面訪分數仍比自填分數高。

不同的施測順序和人口特質來看內在等級相關係數(intra-class correlation coefficient: ICC)之信度差異，ICC在這裡做同一組人前後的比較(自己和自己的比較，組內信度)。

表三為 WHOQOL-BREF 在四個範疇之中，先面訪與先自填兩組組內不同收樣模式的 ICC 指數。先自填組之中 ICC 值最差至最好的範圍為社會範疇 0.56-生理範疇 0.78，先面訪組之中則為環境範疇 0.71-生理範疇 0.8。



*0.01<P<0.5; ** 0.001<P<0.01; *** P<0.001; based on paired t test

Figure 2. Comparison of mean score of each domain in the WHOQOL-BREF among two measurements in self-administrated first group



*0.01<P<0.5; ** 0.001<P<0.01; *** P<0.001; based on paired t test

Figure 3. Comparison of mean score of each domain in the WHOQOL-BREF among two measurements in face-to-face interview first group

Table 4. Reproductivity of domain score among two measurements within each group

	Self-administration	→	Interview	ICC
Physical	14.13(1.93)		14.45(1.9)	0.78
Psychological	13.44(1.95)		14(1.67)	0.68
Social	13.43(2.24)		14.18(1.81)	0.56
Environmental	13.51(1.99)		14.21(1.38)	0.66
	Interview	→	Self-administration	ICC
Physical	15.07(2.17)		14.84(2.24)	0.8
Psychological	14.27(2.15)		13.8(2.21)	0.71
Social	14.4(2.08)		14.27(2.17)	0.76
Environmental	14.52(1.84)		14.07(2.08)	0.71

表五承接表四，為兩組組內依照人口特質分組ICC，用性別分層，分成男生、女生組，在男生先自填組的ICC指數範圍表現為社會範疇0.5-生理範疇0.73，先面訪組則ICC指數範圍為心理範疇0.57-生理範疇0.57。

用年齡分層，分成65歲以上和65歲以下兩組，在65歲以上先自填組之中，ICC指數範圍為社會範疇0.4-生理範疇0.71，先面訪組的ICC指數範圍為生理範疇0.47-生理範疇0.71；在65歲以下先自填組之中，ICC指數範圍為社會範疇0.73-生理範疇0.8，先面訪組的ICC指數範圍為生理範疇0.87-社會範疇0.91。

用年齡性別分層後，可分為65歲以上男、女組和40-65歲男、女組，在65歲以上男生先自填組，ICC指數範圍為社會範疇0.38-生理範疇0.7；先面訪組之ICC指數範圍為環境0.43-生理0.71；在65歲以上女生先自填組，ICC指數範圍為社會範疇0.53-生理範疇0.75；先面訪組之ICC指數範圍為心理0.53-生理0.72。在40-65歲男生先自填組，ICC指數範圍為生理範疇0.62-環境範疇0.84；先面訪組之ICC指數範圍為心理0.78-社會0.91；在40-64歲女生先自填組，ICC指數範圍為環境範疇0.71-生理範疇0.85；先面訪組之ICC指數範圍為心理0.9-生理0.94。

用教育年數程度分層，教育程度九年以下在自填組的ICC範圍為社會0.49-生理0.79，在面訪組的ICC範圍為心理0.75-環境0.8；九年以上在自填組的ICC範圍為社會0.59-生理0.77，在面訪組中ICC範圍為環境0.64-心理0.81。

用疾病分層，無共病症在自填組的ICC範圍為社會0.61-生理0.84，在面訪組中ICC範圍為生理0.79-社會、環境0.87；有共病症在自填組的ICC範圍為社會0.53-生理0.7，在面訪組中ICC範圍為環境0.64-生理0.77。

在先自填組中，WHOQOL-BREF四個範疇之中，以社會範疇表現最差，生理範疇表現最好；在表二列出的13個分層中，其中有11個分層（全部、男性、65歲以上、65歲以上男性、65歲以上女性、65歲以下、教育程度國中以下、教育程度高中以上、無共病症和有共病症）其ICC值在社會範疇表現最差；其中又有12個分層（全部、男性、女性、65歲以上、65歲以上男性、40-64歲女性、40-64歲男性、65歲以上女性、教育程度國中以下、教育程度高中以上、無共病症和有共病症）其ICC值在生理範疇表現最好。

在先面訪組中，以心理和環境範疇表現最差，生理表現最好。13個分層中，其中有5個分層（男性、65歲以上、65歲以上女性、65歲以下男性、教育程度國中以下）其ICC值在心理範疇表現最差；另5個分層（全部、女性、65歲以上男性、教育程度高中以上、有共病症）其ICC值在環境範疇表現最差；有10個分層（全部、男性、女性、65歲以上、65歲以上男性、65歲以上女性、65歲以下、65歲以下女性、教育程度9年以上、有共病症）其ICC值在生理範疇表現最好。

在範疇觀點來看，生理範疇表現最好，社會範疇的表現最差。在組別觀點來看，先面訪的表現較先自填來的好。在性別觀點來看，女生表現較男生好。

在年齡的觀點來看，40-64歲組較65歲以上來的佳。

在共病症有無的觀點來看，無共病症較有共病症來的佳

Table 5 . Reproductivity of self-administration and interview in domain scores of WHOQOL-BREF for each group stratified by gender, age, education and disease status.

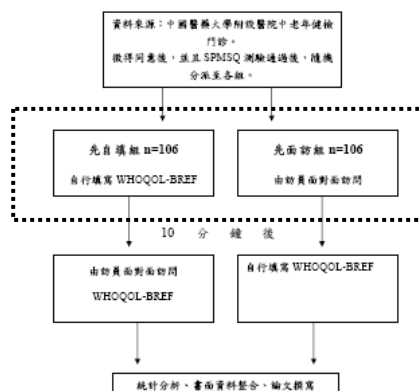
	Physical domain		Psychological domain		Social domain		Environmental domain	
	S(1st)vs I(2nd)	I(1st)vs S(2nd)	S(1st)vs I(2nd)	I(1st)vs S(2nd)	S(1st)vs I(2nd)	I(1st)vs S(2nd)	S(1st)vs I(2nd)	I(1st)vs S(2nd)
Gender								
Male	0.73	0.75	0.57	0.57	0.5	0.66	0.65	0.59
Female	0.83	0.84	0.78	0.81	0.65	0.82	0.68	0.8
Age								
65-	0.80	0.87	0.76	0.9	0.73	0.91	0.77	0.8
65+	0.71	0.71	0.61	0.47	0.4	0.6	0.58	0.49
Education years								
≤9	0.8	0.78	0.58	0.75	0.49	0.79	0.57	0.8
>9	0.77	0.81	0.71	0.7	0.59	0.74	0.7	0.64
Disease								
0	0.84	0.79	0.8	0.84	0.614	0.87	0.73	0.87
≥1	0.7	0.77	0.59	0.66	0.531	0.71	0.59	0.64
Interaction								
65+_male	0.7	0.71	0.54	0.44	0.38	0.6	0.56	0.43
65-_male	0.71	0.8	0.62	0.78	0.74	0.91	0.84	0.88
65+_female	0.75	0.72	0.72	0.53	0.53	0.61	0.62	0.54
65-_female	0.85	0.9	0.83	0.94	0.73	0.91	0.71	0.91

ICC: One way random

S(1st)vs I(2nd): Self-administration first vs. Face-to-face interview second

I(1st)vs S(2nd): Face-to-face interview first vs. Self-administration second

第三節 同一時間點（組間）兩種方式的比較（範疇分析）



由圖四為先自填(n=106)與先面訪(n=106)分數比較，先自填的平均得分最低至最高為心理(13.44)~生理(14.13)；先面訪則為心理(14.27)~生理(15.07)，可從圖中看出全部四個範疇面訪分數顯著高於自填分數(生理 $p<0.001$ ；心理 $p<0.01$ ；社會 $p<0.01$ ；環境 $p<0.001$)；圖五為後自填與後面訪分數比較，後自填的平均得分最低至最高為心理(13.8)~生理(14.84)；後面訪則為心理(14)~生理(14.45)，兩組檢定比較，均沒有顯著差異。

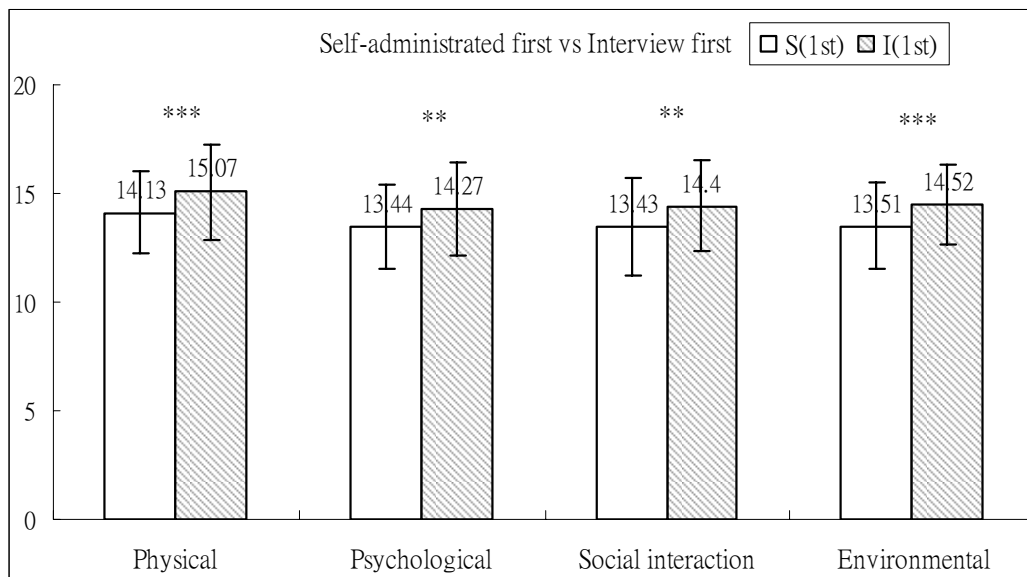
接下來，比較先自填與先面訪兩組人，將分層性別、年齡、教育程度和有無共病症之次族群，看其得分差異。

圖六為女性組的比較，自填分數由最低到最高為心理(13.42)~生理(14.18)；面訪組則為心理(13.97)~生理(15.11)，雙樣本 t 檢定女性在生理 ($p<0.05$)以及環境層面($p<0.01$)有顯著差異；圖七男性組的比較，自填分數由最低到最高為社會(13.13)~生理(14.08)；面訪組則為環境(13.97)~生理(15.08)，兩組雙樣本 t 檢定比較，則是面訪全部的範疇高於自填，並社會範疇相差最為顯著($p<0.001$)。

圖八為年齡 65 歲以下組的比較，自填分數由最低到最高為環境(13.69)~生理(14.83)；面訪組則為心理(13.95)~生理(15.25)，檢定比較後可以看到每

個範疇得分完全沒有顯著差異；圖九為年齡 65 歲以上組的比較，自填分數由最低到最高為社會(12.96)~生理(13.54)；面訪組為環境(14.55)~生理(15.08)，檢定比較可以很清楚看到，每一個範疇面訪分數顯著大於自填分數(生理 $p < 0.05$ ；心理 $p < 0.001$ ；社會 $p < 0.001$ ；環境 $p < 0.001$)。





* $0.01 < p < 0.05$; ** $0.001 < p < 0.01$; *** $p < 0.001$; Two sample t test

Figure 4 . Comparison of mean score of each domain in the WHOQOL-BREF among the first measurements in both groups

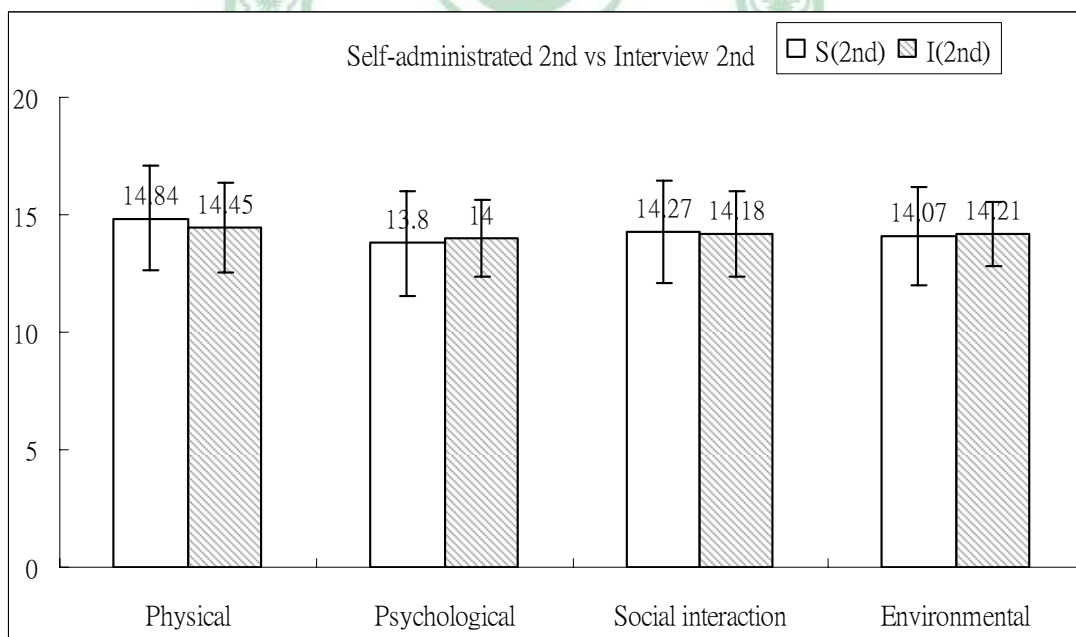
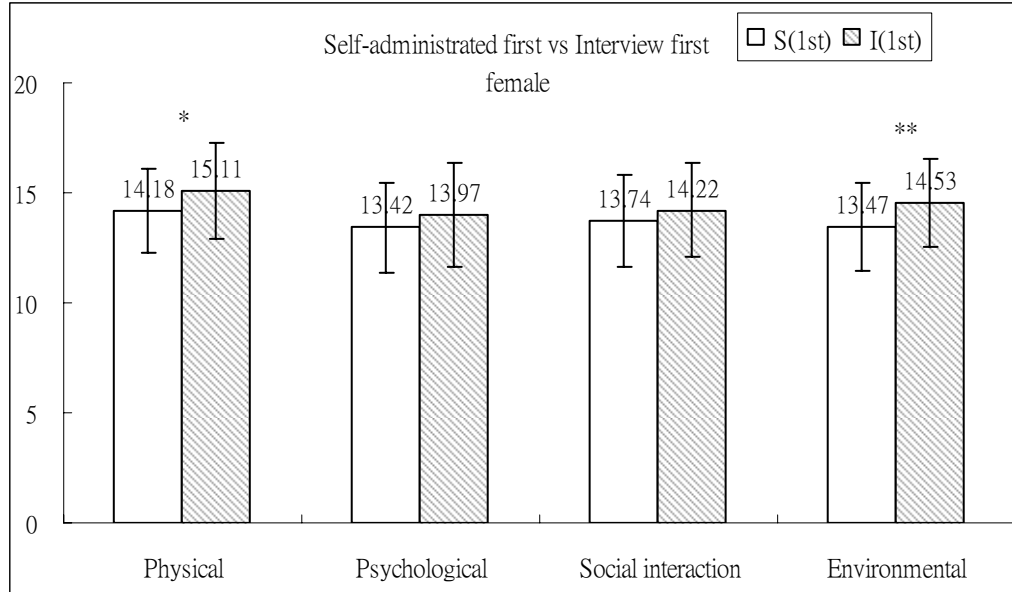


Figure 5. Comparison of mean score of each domain in the WHOQOL-BREF among the second measurements in both groups



*0.01<p<0.5; ** 0.001<p<0.01; *** p<0.001; Two sample t test

Figure 6 . Comparison of mean score of each domain in the WHOQOL-BREF among the first measurements in both groups for female



*0.01<p<0.5; ** 0.001<p<0.01; *** p<0.001; Two sample t test

Figure 7. Comparison of mean score of each domain in the WHOQOL-BREF among the first measurements in both groups for male

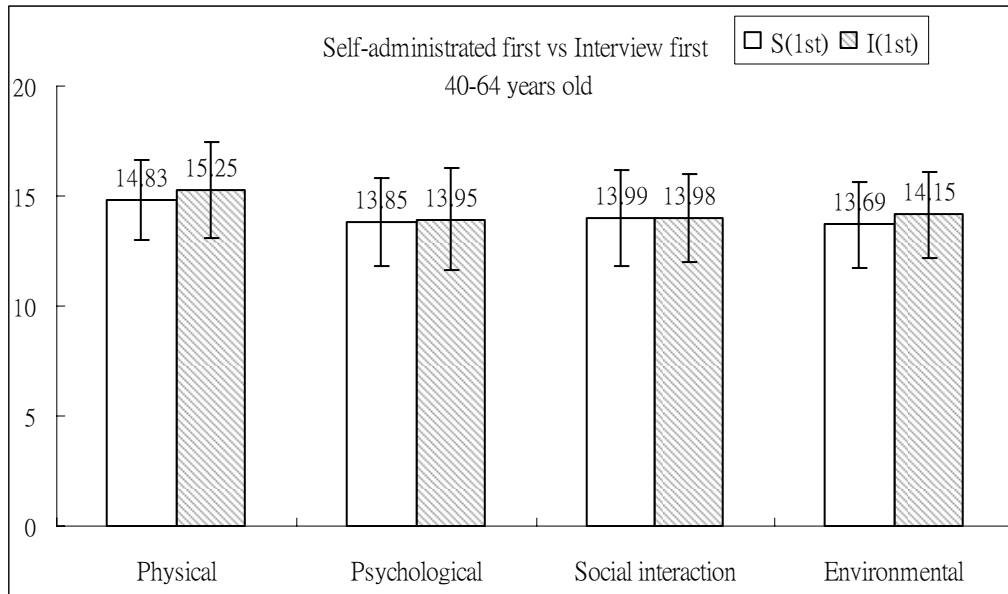


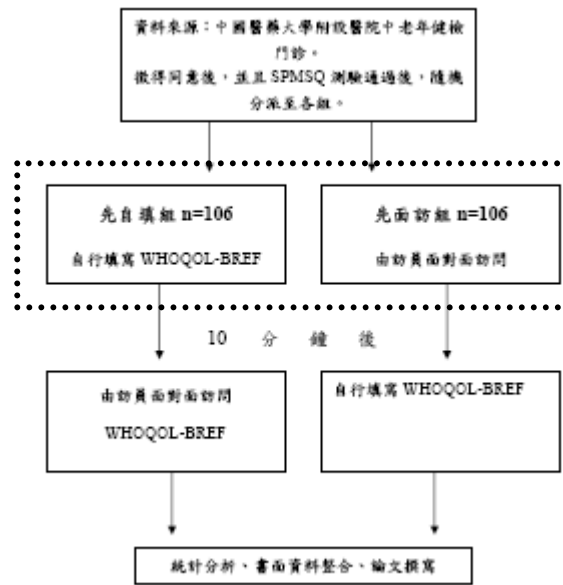
Figure 8. Comparison of mean score of each domain in the WHOQOL-BREF among the first measurements in both groups for age < 65 years



*** $p < 0.001$; Two sample t test

Figure 9. Comparison of mean score of each domain in the WHOQOL-BREF among the first measurements in both groups for age ≥ 65 years

第四節 項目反應理論分析結果



表六為各組試題的平均值(標準差)、各組平均差(標準差)以及 p 值，可以從表 1 看出在平均差的部分，除了 PN16、YN6 之外，其餘的試題面訪分數均高於自填分數，並且在生理範疇(PN3, PN4, PN10, PN15)，心理範疇(YN5, YN11, YN26)，社會範疇(SN21, SN27)和環境範疇(EN8, EN12, EN13, EN14, EN25, EN28)其面訪分數均顯著大於自填分數。

表七為分男性組的得分與兩種模式的得分差異，和表五相同，除了 PN16、YN6 之外，其餘的試題面訪分數均高於自填分數，男性組在生理範疇(PN3, PN10, PN15)，心理範疇(YN5, YN11, YN26)，社會範疇(SN21, SN27)和環境範疇(EN12, EN13, EN14, EN28)其面訪分數均顯著大於自填分數。

表八為分女性組的得分與兩種模式的得分差異，除了 PN16、EN12 之外，其餘的試題面訪分數均高於自填分數，女性組在生理範疇(PN3, PN4, PN15)，心理範疇(YN26)，社會範疇(SN27)和環境範疇(EN8, EN13, EN25, EN28)其面訪分數均顯著大於自填分數。

試題差異男性多於女性。

表九為年齡 65 歲以上組的得分與兩種模式的得分差異，除了 PN16 之外，其餘的試題面訪分數均高於自填分數，65 歲以上組在生理範疇(PN3, PN10, PN15)，心理範疇(YN 5, YN 11, YN 19, YN26)，社會範疇(SN21, SN22, SN27)和環境範疇(EN8, EN9, EN12, EN13, EN14, EN25, EN28)其面訪分數均顯著大於自填分數。

表十為年齡 40-64 歲以下組的得分與兩種模式的得分差異，其分佈未若之前分層（全部、男性、女性、年齡 65 歲以上）面訪大於自填，僅只有 6 題在生理範疇(PN3, PN4, PN15)，心理範疇(YN26)和環境範疇(EN14, EN28)其面訪分數均顯著大於自填分數。



Table 6. Cronbach's alpha, mean score of each item and comparison between self-administration first group and face-to-face interview first group.

Domains/Item	Self-administration Mean (SD)	Face-to-face interview Mean (SD)	S-I Mean (SD)	p-value
Overall				
WN1. Overall quality of life	3.37(0.62)	3.42(0.64)	-0.05(0.63)	0.5737
WN2. Overall health	3.09(0.77)	3.07(0.93)	0.02(0.86)	0.8635
Physical	$\alpha_s=0.73$	$\alpha_t=0.76$		
PN3. Pain and discomfort	3.66 (0.85)	4.32 (0.93)	-0.65 (0.89)	<.0001
PN4. Dependence on medication or treatment	3.79 (1.09)	4.16 (1.12)	-0.37 (1.1)	0.0154
PN10. Energy and fatigue	3.16 (0.67)	3.41 (0.84)	-0.25 (0.76)	0.0182
PN15. Mobility	3.62 (0.79)	3.98 (0.67)	-0.37 (0.73)	0.0004
PN16. Sleep and rest	3.25 (0.84)	3.15 (1.08)	0.1 (0.97)	0.4585
PN17. Activities of daily living	3.59 (0.68)	3.68 (0.68)	-0.09 (0.68)	0.3247
PN18. Working capacity	3.65 (0.66)	3.69 (0.71)	-0.04 (0.68)	0.6855
Psychological	$\alpha_s=0.67$	$\alpha_t=0.73$		
YN5. Enjoyment of life	2.84 (0.8)	3.14 (0.99)	-0.3 (0.9)	0.0155
YN6. Spirituality, religion and personal beliefs	3.57 (0.82)	3.57 (0.85)	0 (0.83)	0.9907
YN7. Thinking, learning, memory & concentration	3.33 (0.77)	3.46 (0.77)	-0.13 (0.77)	0.2389
YN11. Body image and appearance	3.53 (0.78)	3.73 (0.65)	-0.2 (0.72)	0.0436
YN19. Self-satisfaction	3.6 (0.73)	3.71 (0.76)	-0.11 (0.75)	0.2694
YN26. Negative feeling	3.3 (0.9)	3.79 (0.91)	-0.49 (0.9)	0.0001
Social interaction	$\alpha_s=0.77$	$\alpha_t=0.66$		
SN20. Personal relationship	3.52 (0.68)	3.53 (0.73)	-0.01 (0.71)	0.8901
SN21. Sexual activity	3.11 (0.89)	3.53 (0.71)	-0.42 (0.81)	0.0005
SN22. Friends' support	3.57 (0.65)	3.68 (0.67)	-0.11 (0.66)	0.2353
SN27. Esteemed and respected	3.22 (0.72)	3.69 (0.73)	-0.47 (0.73)	<.0001
Environmental	$\alpha_s=0.81$	$\alpha_t=0.73$		
EN8. Physical safety and security	3.12 (0.82)	3.44 (0.89)	-0.32 (0.86)	0.0065
EN9. Physical environments	2.9 (0.95)	3.12 (1.02)	-0.22 (0.98)	0.1098
EN12. Financial environment	3.09 (0.77)	3.34 (0.96)	-0.25 (0.87)	0.0389
EN13. Opportunities	3.44 (0.74)	3.79 (0.75)	-0.36 (0.74)	0.0006
EN14. Participation & support of leisure activities	3.32 (0.88)	3.64 (0.86)	-0.33 (0.87)	0.0069
EN23. Home environment	3.63 (0.7)	3.7 (0.7)	-0.07 (0.7)	0.4925
EN24. Health & social care: availability & quality	3.59 (0.65)	3.72 (0.64)	-0.13 (0.64)	0.1352
EN25. Transportation	3.52 (0.68)	3.73 (0.61)	-0.2 (0.65)	0.0238
EN28. Eating food	3.75 (0.85)	4.18 (0.87)	-0.43 (0.86)	0.0004

α =(internal consistent reliability) cronbach's alpha

Table 7 : Cronbach's alpha, mean score of each item and comparison between self-administration first group and face-to-face interview first group for male.

Domains/Item	Self-administration Mean (SD)	Face-to-face interview Mean (SD)	S-I Mean (SD)	p-value
Overall				
WN1. Overall quality of life	3.34(0.59)	3.47(0.69)	-0.13(0.64)	0.2834
WN2. Overall health	3.13(0.79)	3.04(0.92)	0.1(0.86)	0.5634
Physical domain	$\alpha_s = 0.73$	$\alpha_t = 0.77$		
PN3. Pain and discomfort	3.75 (0.79)	4.33 (0.94)	-0.58 (0.87)	0.0009
PN4. Dependence on medication or treatment	3.56 (1.18)	3.93 (1.23)	-0.37 (1.21)	0.1159
PN10. Energy and fatigue	3.08 (0.68)	3.51 (0.77)	-0.43 (0.73)	0.0027
PN15. Mobility	3.62 (0.84)	3.96 (0.61)	-0.35 (0.73)	0.0167
PN16. Sleep and rest	3.37 (0.79)	3.24 (1.08)	0.12 (0.95)	0.4988
PN17. Activities of daily living	3.58 (0.7)	3.7 (0.6)	-0.13 (0.65)	0.3178
PN18. Working capacity	3.69 (0.62)	3.73 (0.76)	-0.04 (0.7)	0.7736
Psychological domain	$\alpha_s = 0.65$	$\alpha_t = 0.66$		
YN5. Enjoyment of life	2.83 (0.83)	3.2 (1.06)	-0.37 (0.95)	0.0463
YN6. Spirituality, religion and personal beliefs	3.58 (0.84)	3.58 (0.85)	0 (0.85)	0.9849
YN7. Thinking, learning, memory & concentration	3.4 (0.72)	3.55 (0.74)	-0.15 (0.73)	0.2899
YN11. Body image and appearance	3.5 (0.8)	3.84 (0.5)	-0.34 (0.67)	0.0116
YN19. Self-satisfaction	3.58 (0.8)	3.78 (0.63)	-0.21 (0.72)	0.1429
YN26. Negative feeling	3.32 (0.73)	3.93 (0.96)	-0.61 (0.85)	0.0003
Social interaction domain	$\alpha_s = 0.76$	$\alpha_t = 0.68$		
SN20. Personal relationship	3.44 (0.73)	3.55 (0.66)	-0.1 (0.69)	0.4435
SN21. Sexual activity	3 (0.93)	3.6 (0.68)	-0.6 (0.82)	0.0003
SN22. Friends' support	3.52 (0.7)	3.7 (0.66)	-0.18 (0.68)	0.1663
SN27. Esteemed and respected	3.17 (0.76)	3.8 (0.7)	-0.63 (0.73)	<.0001
Environmental domain	$\alpha_s = 0.85$	$\alpha_t = 0.67$		
EN8. Physical safety and security	3.17 (0.73)	3.42 (0.9)	-0.25 (0.82)	0.1259
EN9. Physical environments	3 (0.79)	3.09 (0.97)	-0.09 (0.89)	0.5972
EN12. Financial environment	3.08 (0.76)	3.55 (0.81)	-0.47 (0.79)	0.0027
EN13. Opportunities	3.48 (0.7)	3.84 (0.63)	-0.36 (0.67)	0.0068
EN14. Participation & support of leisure activities	3.33 (0.88)	3.75 (0.84)	-0.42 (0.86)	0.0135
EN23. Home environment	3.63 (0.69)	3.71 (0.74)	-0.07 (0.71)	0.5904
EN24. Health & social care: availability & quality	3.62 (0.63)	3.69 (0.63)	-0.08 (0.63)	0.5386
EN25. Transportation	3.42 (0.72)	3.59 (0.63)	-0.18 (0.68)	0.177
EN28. Eating food	3.72 (0.89)	4.09 (0.87)	-0.38 (0.88)	0.0294

α =(internal consistent reliability) cronbach's alpha

Table 8 : Cronbach's alpha, mean score of each item and comparison between self-administration first group and face-to-face interview first group for female.

Domains/Item	Self-administration Mean (SD)	Face-to-face interview Mean (SD)	S-I Mean (SD)	p-value
Overall				
WN1. Overall quality of life	3.4(0.66)	3.37(0.6)	0.03(0.63)	0.8023
WN2. Overall health	3.03(0.77)	3.12(0.95)	-0.08(0.86)	0.6425
Physical domain	$\alpha_s = 0.74$	$\alpha_t = 0.75$		
PN3. Pain and discomfort	3.58 (0.91)	4.31 (0.93)	-0.74 (0.92)	<.0001
PN4. Dependence on medication or treatment	4.02 (0.94)	4.43 (0.92)	-0.41 (0.93)	0.0268
PN10. Energy and fatigue	3.25 (0.65)	3.31 (0.91)	-0.06 (0.79)	0.6835
PN15. Mobility	3.62 (0.75)	4.02 (0.73)	-0.4 (0.74)	0.0066
PN16. Sleep and rest	3.13 (0.89)	3.06 (1.08)	0.08 (0.99)	0.6983
PN17. Activities of daily living	3.6 (0.66)	3.67 (0.77)	-0.07 (0.72)	0.6186
PN18. Working capacity	3.62 (0.69)	3.66 (0.66)	-0.05 (0.67)	0.7392
Psychological domain	$\alpha_s = 0.69$	$\alpha_t = 0.79$		
YN5. Enjoyment of life	2.85 (0.78)	3.08 (0.93)	-0.23 (0.86)	0.1729
YN6. Spirituality, religion and personal beliefs	3.56 (0.8)	3.57 (0.85)	-0.01 (0.83)	0.9467
YN7. Thinking, learning, memory & concentration	3.27 (0.82)	3.37 (0.8)	-0.1 (0.81)	0.5186
YN11. Body image and appearance	3.56 (0.75)	3.63 (0.77)	-0.07 (0.76)	0.6436
YN19. Self-satisfaction	3.62 (0.66)	3.65 (0.89)	-0.03 (0.78)	0.8383
YN26. Negative feeling	3.29 (1.05)	3.67 (0.84)	-0.38 (0.95)	0.047
Social interaction domain	$\alpha_s = 0.77$	$\alpha_t = 0.62$		
SN20. Personal relationship	3.6 (0.63)	3.53 (0.81)	0.07 (0.73)	0.6418
SN21. Sexual activity	3.24 (0.82)	3.45 (0.76)	-0.21 (0.79)	0.2419
SN22. Friends' support	3.62 (0.6)	3.66 (0.69)	-0.05 (0.64)	0.7274
SN27. Esteemed and respected	3.27 (0.69)	3.59 (0.75)	-0.32 (0.72)	0.0271
Environmental domain	$\alpha_s = 0.78$	$\alpha_t = 0.78$		
EN8. Physical safety and security	3.06 (0.89)	3.47 (0.9)	-0.41 (0.9)	0.0217
EN9. Physical environments	2.81 (1.09)	3.16 (1.08)	-0.35 (1.08)	0.1055
EN12. Financial environment	3.1 (0.78)	3.1 (1.06)	0 (0.93)	1
EN13. Opportunities	3.39 (0.78)	3.76 (0.86)	-0.37 (0.82)	0.024
EN14. Participation & support of leisure activities	3.31 (0.9)	3.55 (0.88)	-0.24 (0.89)	0.1711
EN23. Home environment	3.63 (0.71)	3.71 (0.67)	-0.07 (0.69)	0.6035
EN24. Health & social care: availability & quality	3.56 (0.67)	3.76 (0.65)	-0.21 (0.66)	0.1146
EN25. Transportation	3.63 (0.63)	3.88 (0.55)	-0.25 (0.59)	0.036
EN28. Eating food	3.79 (0.82)	4.29 (0.86)	-0.51 (0.84)	0.0029

α =(internal consistent reliability) cronbach's alpha

Table 9. Cronbach's alpha, mean score of each item and comparison between self-administration first group and face-to-face interview first group for age < 65 years.

Domains/Item	Self-administration Mean (SD)	Face-to-face interview Mean (SD)	S-I Mean (SD)	p-value
Overall				
WN1. Overall quality of life	3.33(0.56)	3.38(0.66)	-0.05(0.61)	0.6775
WN2. Overall health	2.98(0.84)	3.01(0.92)	-0.04(0.88)	0.8206
Physical domain	$\alpha_s = 0.72$	$\alpha_I = 0.78$		
PN3. Pain and discomfort	3.66 (0.94)	4.35 (1.02)	-0.69 (0.98)	0.0006
PN4. Dependence on medication or treatment	4.22 (0.9)	4.6 (0.78)	-0.38 (0.84)	0.0208
PN10. Energy and fatigue	3.33 (0.62)	3.2 (0.91)	0.13 (0.79)	0.378
PN15. Mobility	3.8 (0.75)	4.16 (0.5)	-0.36 (0.63)	0.0049
PN16. Sleep and rest	3.27 (0.87)	3.11 (1.03)	0.17 (0.96)	0.3766
PN17. Activities of daily living	3.69 (0.65)	3.67 (0.77)	0.01 (0.71)	0.9225
PN18. Working capacity	3.8 (0.61)	3.65 (0.75)	0.15 (0.69)	0.2802
Psychological domain	$\alpha_s = 0.7$	$\alpha_I = 0.78$		
YN5. Enjoyment of life	2.88 (0.79)	2.96 (1.09)	-0.08 (0.96)	0.6594
YN6. Spirituality, religion and personal beliefs	3.78 (0.7)	3.65 (0.7)	0.13 (0.7)	0.343
YN7. Thinking, learning, memory & concentration	3.37 (0.8)	3.44 (0.69)	-0.06 (0.74)	0.6596
YN11. Body image and appearance	3.67 (0.82)	3.62 (0.73)	0.05 (0.77)	0.7479
YN19. Self-satisfaction	3.69 (0.71)	3.56 (0.9)	0.12 (0.81)	0.4387
YN26. Negative feeling	3.25 (0.93)	3.78 (0.9)	-0.53 (0.92)	0.0038
Social interaction domain	$\alpha_s = 0.82$	$\alpha_I = 0.67$		
SN20. Personal relationship	3.53 (0.67)	3.36 (0.7)	0.17 (0.69)	0.2189
SN21. Sexual activity	3.35 (0.75)	3.53 (0.72)	-0.19 (0.74)	0.2209
SN22. Friends' support	3.63 (0.66)	3.58 (0.69)	0.05 (0.67)	0.7285
SN27. Esteemed and respected	3.43 (0.64)	3.58 (0.71)	-0.15 (0.68)	0.2567
Environmental domain	$\alpha_s = 0.82$	$\alpha_I = 0.77$		
EN8. Physical safety and security	3.12 (0.86)	3.35 (0.95)	-0.23 (0.91)	0.1997
EN9. Physical environments	2.92 (0.96)	3 (0.98)	-0.08 (0.97)	0.6781
EN12. Financial environment	3.08 (0.72)	3.05 (1.04)	0.03 (0.91)	0.884
EN13. Opportunities	3.45 (0.76)	3.71 (0.9)	-0.26 (0.83)	0.1135
EN14. Participation & support of leisure activities	3.25 (0.84)	3.6 (0.89)	-0.35 (0.87)	0.0441
EN23. Home environment	3.75 (0.66)	3.69 (0.72)	0.05 (0.69)	0.6868
EN24. Health & social care: availability & quality	3.53 (0.67)	3.69 (0.66)	-0.16 (0.67)	0.2166
EN25. Transportation	3.61 (0.7)	3.73 (0.62)	-0.12 (0.66)	0.3528
EN28. Eating food	3.84 (0.76)	4.16 (0.86)	-0.32 (0.81)	0.0444

α =(internal consistent reliability) cronbach's alpha

Table 10. Cronbach's alpha, mean score of each item and comparison between self-administration first group and face-to-face interview first group for age ≥ 65 years.

Domains/Item	Self-administration	Face-to-face interview	S-I	<i>p</i> -value
	Mean (SD)	Mean (SD)	Mean (SD)	
Overall				
WN1. Overall quality of life	3.4(0.68)	3.45(0.63)	-0.05(0.66)	0.6815
WN2. Overall health	3.18(0.71)	3.1(0.96)	0.07(0.84)	0.6768
Physical domain	$\alpha_s=0.69$	$\alpha_i=0.78$		
PN3. Pain and discomfort	3.65 (0.77)	4.31 (0.81)	-0.66 (0.79)	<.0001
PN4. Dependence on medication or treatment	3.38 (1.07)	3.73 (1.22)	-0.35 (1.15)	0.1094
PN10. Energy and fatigue	3.02 (0.7)	3.65 (0.7)	-0.64 (0.7)	<.0001
PN15. Mobility	3.45 (0.78)	3.82 (0.8)	-0.37 (0.79)	0.0147
PN16. Sleep and rest	3.21 (0.82)	3.19 (1.12)	0.03 (0.98)	0.8772
PN17. Activities of daily living	3.48 (0.69)	3.69 (0.58)	-0.2 (0.64)	0.0968
PN18. Working capacity	3.49 (0.67)	3.72 (0.66)	-0.23 (0.66)	0.0734
Psychological domain	$\alpha_s=0.63$	$\alpha_i=0.69$		
YN5. Enjoyment of life	2.81 (0.79)	3.33 (0.84)	-0.52 (0.81)	0.001
YN6. Spirituality, religion and personal beliefs	3.39 (0.86)	3.51 (0.98)	-0.12 (0.92)	0.4808
YN7. Thinking, learning, memory & concentration	3.26 (0.77)	3.51 (0.86)	-0.25 (0.81)	0.1126
YN11. Body image and appearance	3.39 (0.71)	3.87 (0.55)	-0.48 (0.63)	0.0001
YN19. Self-satisfaction	3.48 (0.74)	3.85 (0.62)	-0.37 (0.68)	0.0049
YN26. Negative feeling	3.33 (0.91)	3.78 (0.94)	-0.45 (0.92)	0.0116
Social interaction domain	$\alpha_s=0.71$	$\alpha_i=0.64$		
SN20. Personal relationship	3.48 (0.69)	3.69 (0.74)	-0.21 (0.72)	0.127
SN21. Sexual activity	2.85 (0.92)	3.54 (0.71)	-0.68 (0.83)	0.0002
SN22. Friends' support	3.5 (0.63)	3.75 (0.68)	-0.26 (0.65)	0.0447
SN27. Esteemed and respected	3.05 (0.75)	3.84 (0.74)	-0.78 (0.74)	<.0001
Environmental domain	$\alpha_s=0.81$	$\alpha_i=0.66$		
EN8. Physical safety and security	3.11 (0.76)	3.55 (0.86)	-0.44 (0.81)	0.0051
EN9. Physical environments	2.86 (0.94)	3.27 (1.03)	-0.42 (0.99)	0.0283
EN12. Financial environment	3.07 (0.81)	3.6 (0.85)	-0.53 (0.83)	0.0011
EN13. Opportunities	3.4 (0.71)	3.84 (0.66)	-0.44 (0.69)	0.0012
EN14. Participation & support of leisure activities	3.38 (0.91)	3.73 (0.83)	-0.35 (0.87)	0.0347
EN23. Home environment	3.52 (0.71)	3.73 (0.71)	-0.21 (0.71)	0.123
EN24. Health & social care: availability & quality	3.61 (0.65)	3.76 (0.61)	-0.16 (0.63)	0.1936
EN25. Transportation	3.39 (0.7)	3.74 (0.59)	-0.36 (0.65)	0.0048
EN28. Eating food	3.67 (0.91)	4.24 (0.89)	-0.57 (0.9)	0.0011

α =internal consistent reliability cronbach's alpha

項目反應理論分析面訪與自填模式

表十一為在面訪與自填的信度(reliability)、難度(item difficulty)與適配指標(infit mean square)比較。

適合度指標(infit statistic)在WHOQOL-BREF的四個範疇中，兩組均符合單一向度的假設，但在反向題PN4“*依賴藥物*”(1.31)、YN26“*負面感覺*”(1.34)不符合和PN3“*疼痛影響需要做的事*”的指標值稍微較高。

在信度的評估中，受測者潛在特質的信度指標與傳統的測量信度的Cronbach's alpha 相同，在自填組信度指標生理、心理、社會及環境範疇的信度指標分別為 0.68、0.6、0.61 及 0.79。在面訪組信度指標生理、心理、社會及環境範疇的信度指標分別為 0.71、0.69、0.45 及 0.71，其中面訪組社會範疇的信度指標最差為 0.45，內部一致性最好的為環境範疇 0.79。

在自填組生理範疇難度範圍為(-0.79~0.39)，最難的是 PN17“*滿意從事日常活的能力*”(0.39)，最簡單的是 PN15“*四處行動的能力*”(-0.79)；面訪組難度範圍為(-0.97~0.79)，最難的是 PN16“*滿意睡眠狀況*”(0.79)，最簡單的是 PN3“*疼痛是否會妨礙需要處理的事*”(-0.97)，其中在自填組之中 PN4“*依賴藥物*”；在心理範疇之中，自填組難度範圍為(-0.28~0.48)，最難的 YN11“*接受自己的外表*”(0.48)，最簡單 YN19“*對自己滿意*”(-0.28)，面訪組難度範圍為(-0.41~0.72)，最難的 YN5“*享受生活*”(0.72)，最簡單 YN11“*接受自己的外表*”(-0.41)；在社會範疇之中，自填組難度範圍為(-1.25~1.68)，最難的 SN22“*滿意朋友給您的支持*”(1.68)，最簡單 SN27“*覺得自己有面子或被尊重*”(-1.25)，面訪組難度範圍為(-1.52~1.31)，最難的 SN22“*滿意朋友支持*”(1.31)，最簡單 SN21“*滿意自己的性生活*”(-1.52)；在環境範疇之中，自填組難度範圍為(-1.44~0.9)，最難的 EN9“*所處的環境健康嗎*”(0.9)，最簡單 EN28“*想吃的食物能吃*”

到” (-1.44)，面訪組難度範圍為(-1.12~0.64)，最難的 EN9”*所處的環境健康嗎*” (0.64)，最簡單 EN28”*想吃的食物能吃到*” (-1.12)

我們可以看到，兩組的困難度順序(item hierarchy)並不一致，這代表受測者在填答試題時，面對這兩種管理模式的反應性並不同，我們評估了兩組的難度順序與難度值看其差異。生理範疇中，PN3”*疼痛是否會妨礙需要處理的事*”，在面訪組中顯得較簡單；心理範疇中，YN5”*享受生活*”，自填組較簡單（題目較抽象）；社會範疇中，SN21”*滿意自己的性生活*”在面訪明顯較簡單，而 SN27”*覺得自己有面子或被尊重*”則在自填組較簡單；環境範疇中，EN13”*方便得到生活所需資訊*”在面訪組明顯較簡單。

圖十為自填與面訪的 DIF(differential item functioning)情形，一共有 8/26 題產生 DIF。生理範疇有 PN3”*身體疼痛妨礙需要處理的事*”、PN16”*滿意自己的睡眠狀況*”，心理範疇有 YN5”*享受生活*”、YN11”*接受自己的外表*”，社會範疇有 SN21”*滿意自己的性生活*”、SN27”*覺得有面子或被尊重*”，環境範疇有 EN13”*方便得到生活所需資訊*”、EN23”*滿意所住的狀況*”，顯示受測者在使用自填與面訪兩種施測模式填答題目反應不穩定。

Table 11. Item difficulty, infit statistics, and personal separation reliability of each item among self-administration first group and face-to-face interview first group (arranged by item difficulty within each domain)

Domains / Item	Self-administration		Face-to-face interview	
	Item difficulty	Infit Mean Square	Item difficulty	Infit Mean Square
Physical domain	Reliability: 0.68		Reliability: 0.71	
PN15. Mobility	-0.79	0.91	-0.44	1.02
PN4. Dependence on medication or treatment	-0.59	1.31	-0.84	1.23
PN16. Sleep and rest	0.2	0.86	0.79	1.28
PN18. Working capacity	0.23	0.87	0.15	0.83
PN3. Pain and discomfort	0.27	1.29	-0.97	1.01
PN10. Energy and fatigue	0.29	0.93	0.77	0.97
PN17. Activities of daily living	0.39	0.82	0.54	0.7
Psychological domain	Reliability: 0.6		Reliability: 0.7	
YN19. Self-satisfaction	-0.28	0.84	-0.4	0.68
YN5. Enjoyment of life	-0.27	0.97	0.72	0.9
YN6. Spirituality, religion and personal beliefs	-0.23	0.84	0.03	1.16
YN7. Thinking, learning, memory & concentration	0.11	1.04	0.07	1.01
YN26. Negative feeling	0.19	1.34	0	1.23
YN11. Body image and appearance	0.48	0.9	-0.41	0.94
Social domain	Reliability: 0.61		Reliability: 0.45	
SN27. Esteemed and respected	-1.25	1.07	0.04	1.14
SN21. Sexual activity	-0.5	1.05	-1.52	1.23
SN20. Personal relationship	0.07	0.96	0.17	0.82
SN22. Friends' support	1.68	0.87	1.31	0.83
Environmental domain	Reliability: 0.79		Reliability: 0.71	
EN28. Eating food	-1.44	1.07	-1.12	0.97
EN14. Participation & support of leisure activities	-0.54	1.18	-0.63	0.96
EN12. Financial environment	-0.31	0.86	0.27	0.98
EN23. Home environment	-0.23	0.9	0.29	0.95
EN24. Health & social care: availability & quality	0.02	1.12	0.31	0.96
EN8. Physical safety and security	0.38	0.78	0.31	0.97
EN13. Opportunities	0.59	0.82	-0.34	0.98
EN25. Transportation	0.63	1.24	0.27	0.9
EN9. Physical environments	0.9	1.04	0.64	1.21

Table 12. Comparison of internal consistent reliability between CTT and IRT among self-administration first group and face-to-face interview first group

Domain\mode of administrated	Self-administration		Face-to-face interview	
	CTT	IRT	CTT	IRT
Physical	0.73	0.68	0.76	0.71
Psychological	0.67	0.6	0.73	0.69
Social	0.77	0.61	0.66	0.45
Environmental	0.81	0.79	0.73	0.71

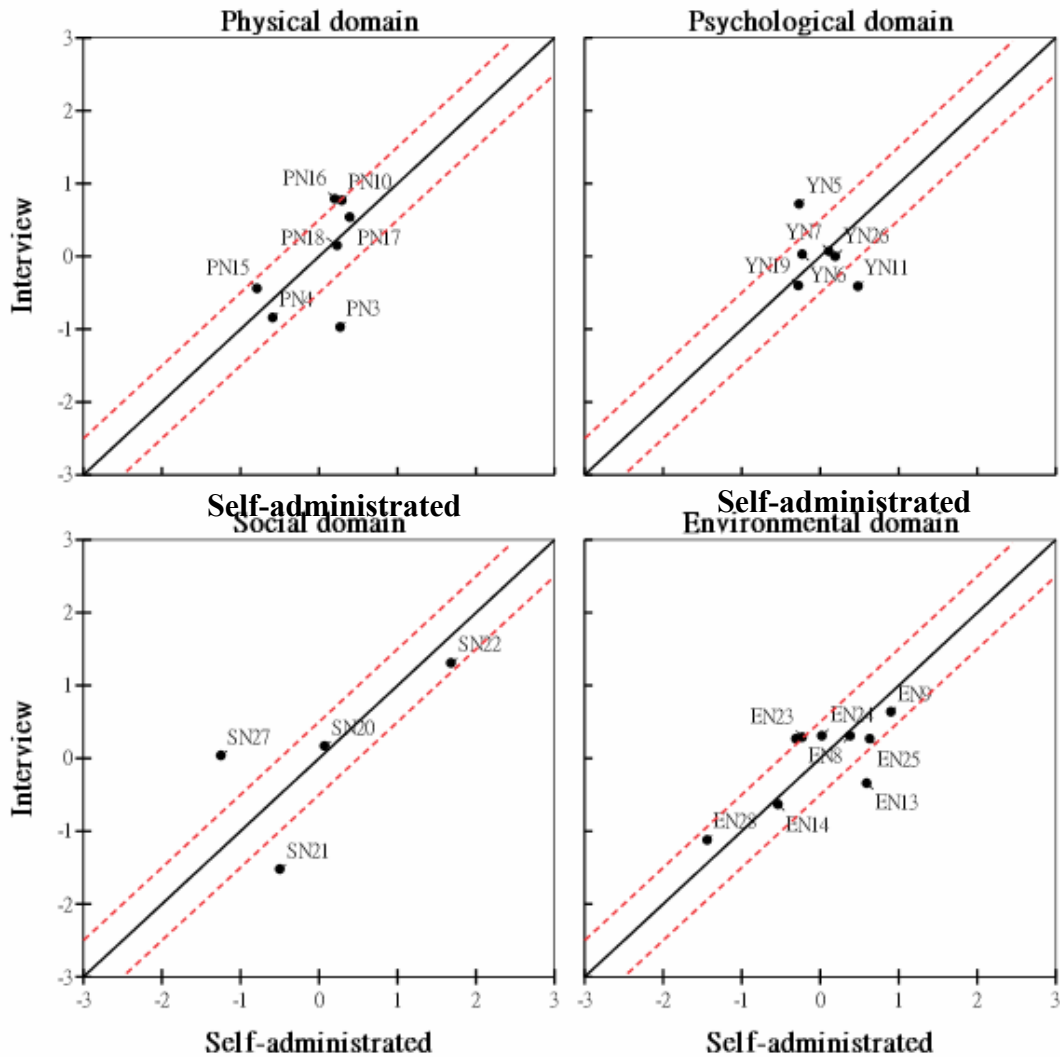


Figure10. DIF analysis among the mode of self-administration first and the mode of face-to-face interview first in each domain of WHOQOL-BREF. Self-administration calibrations on the x axis, and face-to-face interview on the y axis. Calibrations are expressed in logits. Negative calibrations indicate easier items. Positive calibrations indicate more difficult items. An identity line is drawn through the origin, with a slope of 1. Statistical control lines (95% CI) are drawn around the identity line (dotted lines). The area between the dotted lines depicts acceptable item deviation. Items outside the control lines demonstrate DIF.

項目反應理論分析男性在自填與面訪心理特質

表十三為男性在自填與面訪的試題難度與相關指標，其中適合度指標 (infit statistic) 在 WHOQOL-BREF 的四個範疇中，在自填組有 PN3”疼痛會妨礙需要做的事”(1.45)、PN4”需要醫療應付日常生活”(1.42)、YN26”負面感覺”(1.44)，都是反向題。

在男性的信度分析當中，最差的仍是在面訪組的社會範疇，信度只有 0.37。

男性生理範疇中，自填組範圍為(-0.85~0.73)，最難的 PN17”滿意從事日常活動的能力”(0.73)、最簡單 PN10”足夠的精力”(-0.85)，面訪組的範圍(-0.86~1.13)，最難的 PN17”滿意從事日常活動的能力”(1.13)，最簡單 PN4”需要醫療應付日常生活”(-0.86)；男性心理範疇中，自填組範圍為(-0.41~0.7)，最難 YN26”負面感覺”(0.7)、最簡單 YN19”對自己滿意”(-0.41)，面訪組範圍(-0.31~0.31)，最難 YN5”享受生活”(0.31)、最簡單 YN26”負面感覺”(-0.31)；男性社會範疇中，自填難度範圍為(-0.99~1.35)，最難 SN22”滿意朋友支持”、最簡單 SN27”覺得自己有面子或被尊重”，面訪難度範圍為(-1.31~1.05)，最難 SN20”滿意人際關係”(1.05)、最簡單 SN21”滿意自己的性生活”(-1.31)；男性在環境範疇之中，自填組難度範圍為(-1.35~0.97)，最難的 EN9”所處的環境健康嗎”(0.9)，最簡單 EN8”感到安全”(-1.35)，面訪組難度範圍為(-1.1~0.52)，最難的 EN25”滿意交通運輸方式”(0.52)，最簡單 EN28”想吃的食物能吃到”(-1.12)。

男性的兩組的困難度順序在生理範疇中，PN3”疼痛會妨礙需要做的事”、PN4”需要醫療應付日常生活”在面訪組則明顯較簡單，心理範疇則在 YN26”負面感受”在面訪較簡單，社會範疇以 SN21”滿意性生活”明顯在面

訪組較簡單，SN20”滿意人際關係”則在面訪組較簡單，環境範疇 EN8”感到安全”、EN25”交通運輸方式”在自填組較簡單，EN13”方便得到資訊”在面訪組較簡單。

圖十一為在男性自填與面訪的 DIF 情形，有 10/26 題有 DIF 的情況，生理範疇有 PN3”疼痛會妨礙需要做的事”、PN4”需要醫療應付日常生活”和 PN10”足夠的精力”；心理範疇有 YN26”負面感受”，而 YN5”享受生活”有輕微的 DIF；在社會範疇之中，SN20”滿意人際關係”、SN21”滿意性生活”；環境範疇 EN8”感到安全”、EN25”交通運輸方式”、EN12”金錢足夠應付所需”、EN13”方便得到資訊”。



Table 13. Item difficulty, infit statistics, and personal separation reliability of each item among self-administration first group and face-to-face interview first group for male (arranged by item difficulty within each domain)

Domains / Item	Self-administration		Face-to-face interview	
	Item difficulty	Infit Mean Square	Item difficulty	Infit Mean Square
Physical domain	Reliability: 0.65		Reliability: 0.71	
PN10. Energy and fatigue	-0.85	0.92	0.22	0.93
PN15. Mobility	-0.43	0.78	-0.45	0.92
PN16. Sleep and rest	0	0.74	0.51	0.8
PN18. Working capacity	0.14	0.78	0.25	1.38
PN4. Dependence on medication or treatment	0.2	1.42	-0.86	1.22
PN3. Pain and discomfort	0.21	1.45	-0.78	1.06
PN17. Activities of daily living	0.73	0.85	1.13	0.71
Psychological domain	Reliability: 0.63		Reliability: 0.65	
YN19. Self-satisfaction	-0.41	0.8	0.07	0.75
YN6. Spirituality, religion and personal beliefs	-0.27	0.83	-0.08	1.31
YN5. Enjoyment of life	-0.24	1	0.31	0.87
YN7. Thinking, learning, memory & concentration	-0.02	0.89	-0.12	0.77
YN11. Body image and appearance	0.23	0.79	0.13	0.89
YN26. Negative feeling	0.7	1.44	-0.31	1.36
Social domain	Reliability: 0.69		Reliability: 0.37	
SN27. Esteemed and respected	-0.99	1.35	-0.67	1.08
SN21. Sexual activity	-0.24	0.95	-1.31	1.27
SN20. Personal relationship	-0.11	0.76	1.05	0.92
SN22. Friends' support	1.35	0.82	0.93	0.76
Environmental domain	Reliability: 0.79		Reliability: 0.76	
EN8. Physical safety and security	-1.35	0.9	0.15	0.94
EN25. Transportation	-0.72	1.16	-1.1	1.03
EN28. Eating food	-0.18	0.73	0.17	0.94
EN23. Home environment	-0.07	1.04	-0.02	0.89
EN14. Participation & support of leisure activities	0.53	1.28	0.35	1.06
EN24. Health & social care: availability & quality	0.66	0.87	-0.02	0.89
EN12. Financial environment	0.78	0.99	0.52	0.88
EN13. Opportunities	0.93	0.88	-0.48	1.09
EN9. Physical environments	0.97	1.1	0.43	1.26

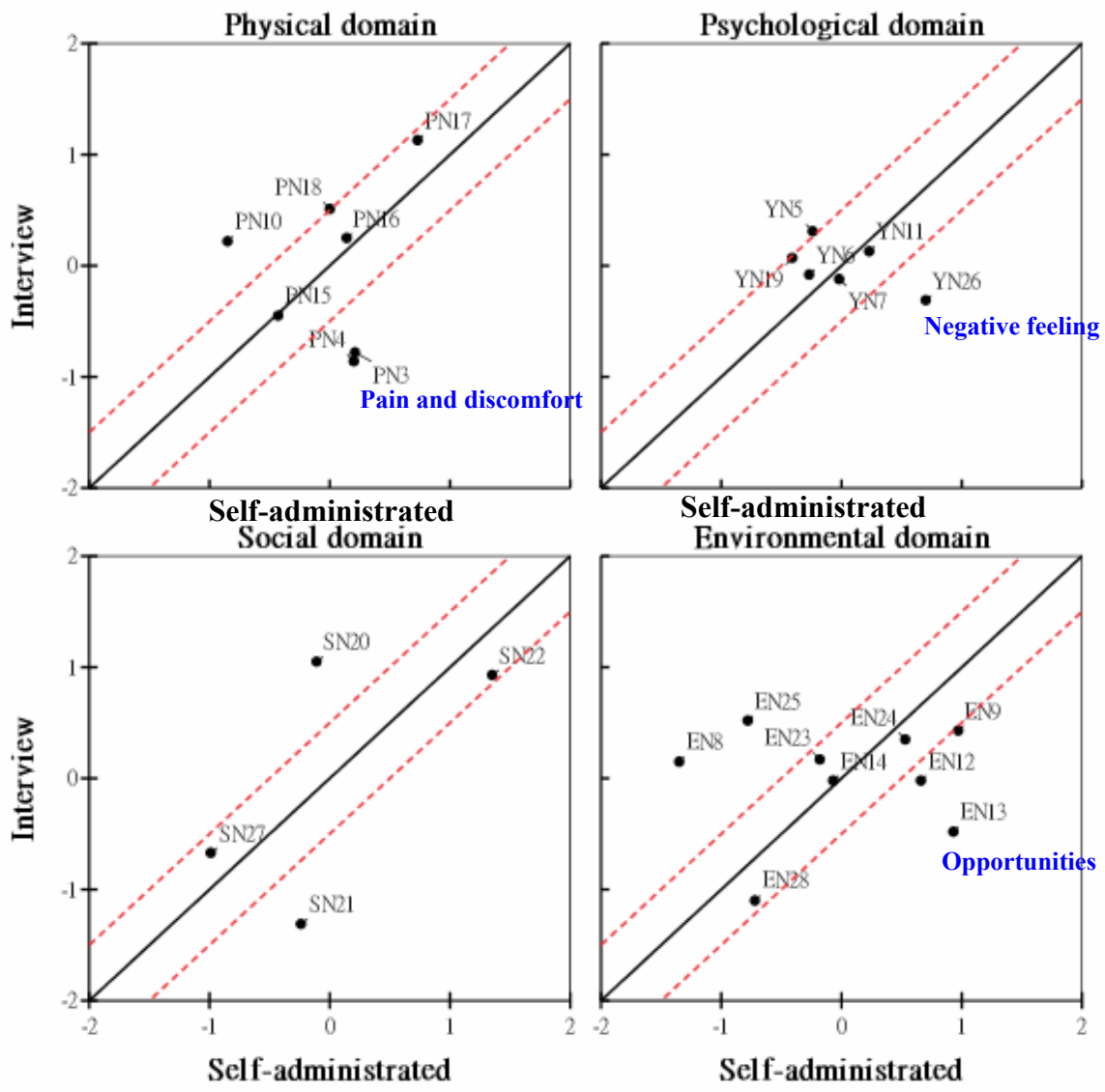


Figure11. DIF analysis among the mode of self-administration first and the mode of face-to-face interview first in each domain of WHOQOL-BREF for male.

項目反應理論分析女性在自填與面訪心理特質

表十四為女性在自填與面訪的試題難度與相關指標，其中適合度指標 (infit statistic) 在 WHOQOL-BREF 的四個範疇中，全部題目均符合單一向度的假設。在女性的信度分析當中，最差的仍是在社會範疇，自填與面訪信度只有 0.47 與 0.5。

在女性組試題難度上，在生理範疇中，自填組範圍為(-1~1.04)，最難的 PN4”*依賴藥物*”(1.04)、最簡單 PN10”*足夠的精力*”(-1)；面訪組的範圍 (-1.03~0.78)，最難的 PN16”*睡眠狀況*”(0.78)，最簡單 PN3”*疼痛是否會妨礙需要處理的事*”(-1.03)；女性在心理範疇中，自填難度範圍為(-0.47~0.71)，最難 YN7”*集中精神的能力*”(0.71)，最簡單 YN26”*負面感覺*”(-0.47)；面訪難度範圍為(-0.48~0.37)，最難 YN7”*集中精神的能力*”(0.37)，最簡單 YN5”*享受生活*”(-0.48)；女性社會範疇中，自填難度範圍為(-0.76~1.15)，最難 SN27”*有面子或被尊重*”(1.15)、最簡單 SN22”*滿意朋友支持*”(-0.76)，面訪難度範圍為(-1.19~1.04)，最難 SN22”*滿意朋友支持*”(1.04)、最簡單 SN21”*滿意自己的性生活*”(-1.19)；女性在環境範疇之中，自填組難度範圍為 (-0.7~0.88)，最難的 EN13”*日常生活所需資訊*”(0.9)，最簡單 EN28”*想吃的食物能吃到*”(-0.7)，面訪組難度範圍為(-1.35~0.75)，最難的 EN9”*所處環境健康*”(0.75)，最簡單 EN28”*想吃的食物能吃到*”(-1.35)。

難度順序上，生理範疇有 PN3”*疼痛會妨礙需要做的事*”在面訪較簡單和 PN16”*滿意睡眠狀況*”在自填較簡單，心理範疇有 YN26”*負面感受*”和 YN5”*享受生活*”，環境範疇 EN9”*所處環境健康*”自填較簡單、EN13”*方便得到資訊*”、EN28”*想吃的食物能吃到*”面訪較簡單。

圖十二為在女性自填與面訪的 DIF 情形，一共有 9/26 題 DIF，生理範疇有 PN3”*疼痛會妨礙需要做的事*”和 PN16”*滿意睡眠狀況*”；心理範疇有 YN26”

負面感受”在自填較簡單和 YN5”享受生活”在面訪較簡單；在社會範疇之中， SN27”有面子或被尊重”、”SN21”滿意性生活”在面訪較簡單；環境範疇 EN9”所處環境健康”、EN13”方便得到資訊”、EN28”想吃的食物能吃到”。



Table 14. Item difficulty, infit statistics, and personal separation reliability of each item among self-administration first group and face-to-face interview first group for female (arranged by item difficulty within each domain)

Domains / Item	Self-administration		Face-to-face interview	
	Item difficulty	Infit Mean Square	Item difficulty	Infit Mean Square
Physical domain	Reliability:0.75		Reliability:0.72	
PN4. Dependence on medication or treatment	-1	1.1	-0.73	1.13
PN17. Activities of daily living	-0.1	0.8	0.39	0.69
PN15. Mobility	-0.03	1.04	-0.56	1.15
PN18. Working capacity	0.01	0.98	0.45	0.85
PN3. Pain and discomfort	0.04	1.2	-1.03	1
PN16. Sleep and rest	0.04	0.93	0.78	1.12
PN10. Energy and fatigue	1.04	0.98	0.71	1.02
Psychological domain	Reliability:0.6		Reliability:0.73	
YN26. Negative feeling	-0.47	1.05	0.33	0.99
YN6. Spirituality, religion and personal beliefs	-0.25	0.88	0.12	0.92
YN11. Body image and appearance	-0.21	1	-0.16	1.07
YN19. Self-satisfaction	0.04	0.88	-0.18	0.68
YN5. Enjoyment of life	0.19	0.97	-0.48	0.84
YN7. Thinking, learning, memory & concentration	0.71	1.18	0.37	1.27
Social domain	Reliability:0.47		Reliability:0.5	
SN22. Friends' support	-0.76	0.9	1.04	0.84
SN20. Personal relationship	-0.47	1.13	0.07	0.74
SN21. Sexual activity	0.08	1.18	-1.19	1.23
SN27. Esteemed and respected	1.15	0.81	0.08	1.18
Environmental domain	Reliability:0.79		Reliability:0.66	
EN28. Eating food	-0.7	1.02	-1.35	0.91
EN9. Physical environments	-0.67	0.95	0.75	1.24
EN14. Participation & support of leisure activities	-0.48	1.32	-0.27	1.06
EN25. Transportation	-0.17	1.32	-0.06	0.85
EN24. Health & social care: availability & quality	0.02	0.99	0.19	0.86
EN23. Home environment	0.12	1.03	0.24	0.98
EN8. Physical safety and security	0.21	0.66	0.35	1.04
EN12. Financial environment	0.8	0.85	0.47	0.92
EN13. Opportunities	0.88	0.79	-0.32	0.95

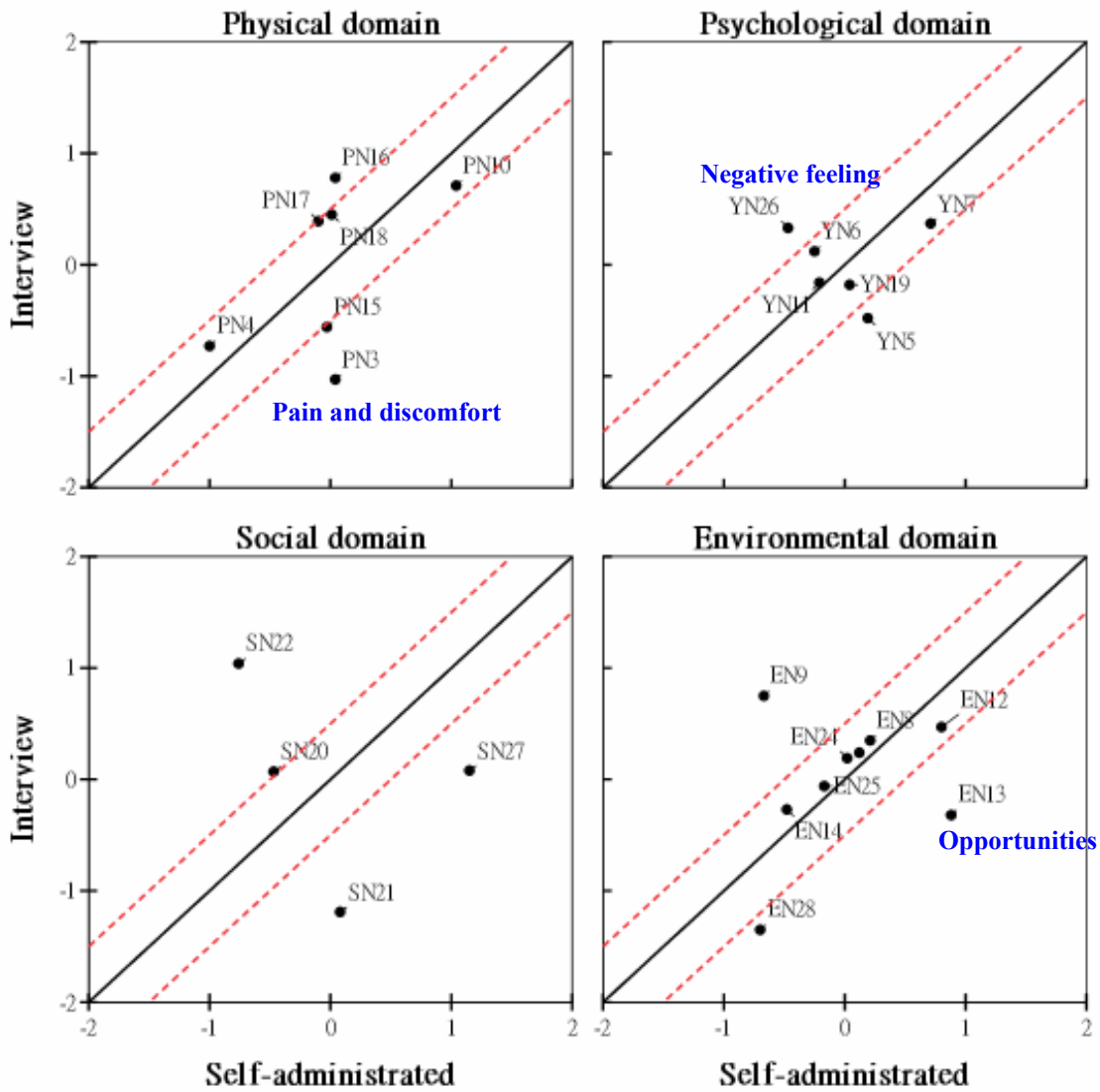


Figure 12. DIF analysis among the mode of self-administration first and the mode of face-to-face interview first in each domain of WHOQOL-BREF for female

項目反應理論分析年齡 40-64 組在自填與面訪心理特質

表十五為年齡 40-64 組在自填與面訪的試題難度與相關指標，其中適合度指標(infit statistic)在 WHOQOL-BREF 的四個範疇中，在面訪組中生理範疇的”睡眠情況” infit 值為 1.43，不符合單一向度的假設。其餘題目均符合單一向度的假設。在年齡 40-64 組信度分析當中，最差的仍是在社會範疇，自填與面訪信度只有 0.45 與 0.25。

在年齡 40-64 組試題難度上，在生理範疇中，自填組範圍為 (-1.45~1.01)，最難的 PN18”滿意自己的工作能力”(1.01)、最簡單 PN4”依賴藥物” (-1)；面訪組的範圍(-1.01~0.24)，最難的 PN16”睡眠狀況”(0.78)，最簡單 PN4”依賴藥物” (-1.01)；在心理範疇中，自填難度範圍為(-0.74~0.73)，最難 YN7”集中精神的能力”(0.71)，最簡單 YN5”享受生活” (-0.74)；面訪難度範圍為(-1.83~1.57)，最難 YN5”享受生活”(1.57)，最簡單 YN6”生命有意義嗎” (-1.83)；社會範疇中，自填難度範圍為(-1.4~1.59)，最難 SN20”滿意人際關係”(1.59)、最簡單 SN22”滿意朋友支持” (-1.4)，面訪難度範圍為 (-0.35~0.27)，最難 SN22”滿意朋友支持”(0.27)、最簡單 SN27”覺得有面子或被尊重嗎”(-0.35)；在環境範疇之中，自填組難度範圍為(-1.01~1.37)，最難的 EN12”足夠的金錢應付所需” (-1.37)，最簡單 EN9”所處環境健康”，面訪組難度範圍為(-1.13~0.99)，最難的 EN12”足夠的金錢應付所需” (-1.35)，最簡單 EN28”想吃的食物能吃到” (-1.13)。

難度順序上，生理範疇有 PN16”滿意睡眠狀況”，心理範疇有 YN5”享受生活”，環境範疇、EN13”方便得到資訊”，難度順序明顯不一致。圖十三為年齡 40-64 組自填與面訪的 DIF 情形，一共有 17/26 題 DIF。

Table 15. Item difficulty, infit statistics, and personal separation reliability of each item among self-administration first group and face-to-face interview first group for age < 65 years (arranged by item difficulty within each domain)

Domains / Item	Self-administration		Face-to-face interview	
	Item difficulty	Infit Mean Square	Item difficulty	Infit Mean Square
Physical domain	Reliability:0.7		Reliability:0.7	
PN4. Dependence on medication or treatment	-1.45	1.06	-1.01	0.93
PN17. Activities of daily living	-0.28	0.87	0.6	0.57
PN15. Mobility	-0.24	1.17	0.18	1.21
PN16. Sleep and rest	-0.01	0.73	1.04	1.43
PN3. Pain and discomfort	0.09	1.28	-0.86	1.15
PN10. Energy and fatigue	0.88	0.94	-0.19	1.02
PN18. Working capacity	1.01	1.03	0.24	0.81
Psychological domain	Reliability:0.65		Reliability:0.67	
YN5. Enjoyment of life	-0.74	1.01	1.57	1.07
YN6. Spirituality, religion and personal beliefs	-0.54	1.02	-1.83	1.03
YN11. Body image and appearance	0	1.08	-1.7	1.14
YN19. Self-satisfaction	0.12	0.94	0.2	0.58
YN26. Negative feeling	0.42	0.82	0.27	1.05
YN7. Thinking, learning, memory & concentration	0.73	1.06	1.48	1.04
Social domain	Reliability:0.45		Reliability:0.25	
SN22. Friends' support	-1.4	0.64	0.27	0.97
SN27. Esteemed and respected	-0.6	1.02	-0.35	0.95
SN21. Sexual activity	0.4	1.23	0.01	1.01
SN20. Personal relationship	1.59	0.99	0.07	1.06
Environmental domain	Reliability:0.82		Reliability:0.79	
EN9. Physical environments	-1.01	1.08	-0.45	1.12
EN28. Eating food	-0.84	1.1	-1.13	0.94
EN23. Home environment	-0.62	1.09	0.53	1.07
EN25. Transportation	-0.2	1.16	0.58	0.87
EN24. Health & social care: availability & quality	-0.06	1.11	0.7	0.96
EN8. Physical safety and security	0.18	0.72	-0.92	1.09
EN13. Opportunities	0.37	0.78	-0.18	0.94
EN14. Participation & support of leisure activities	0.8	1.13	-0.13	0.99
EN12. Financial environment	1.37	0.76	0.99	0.82

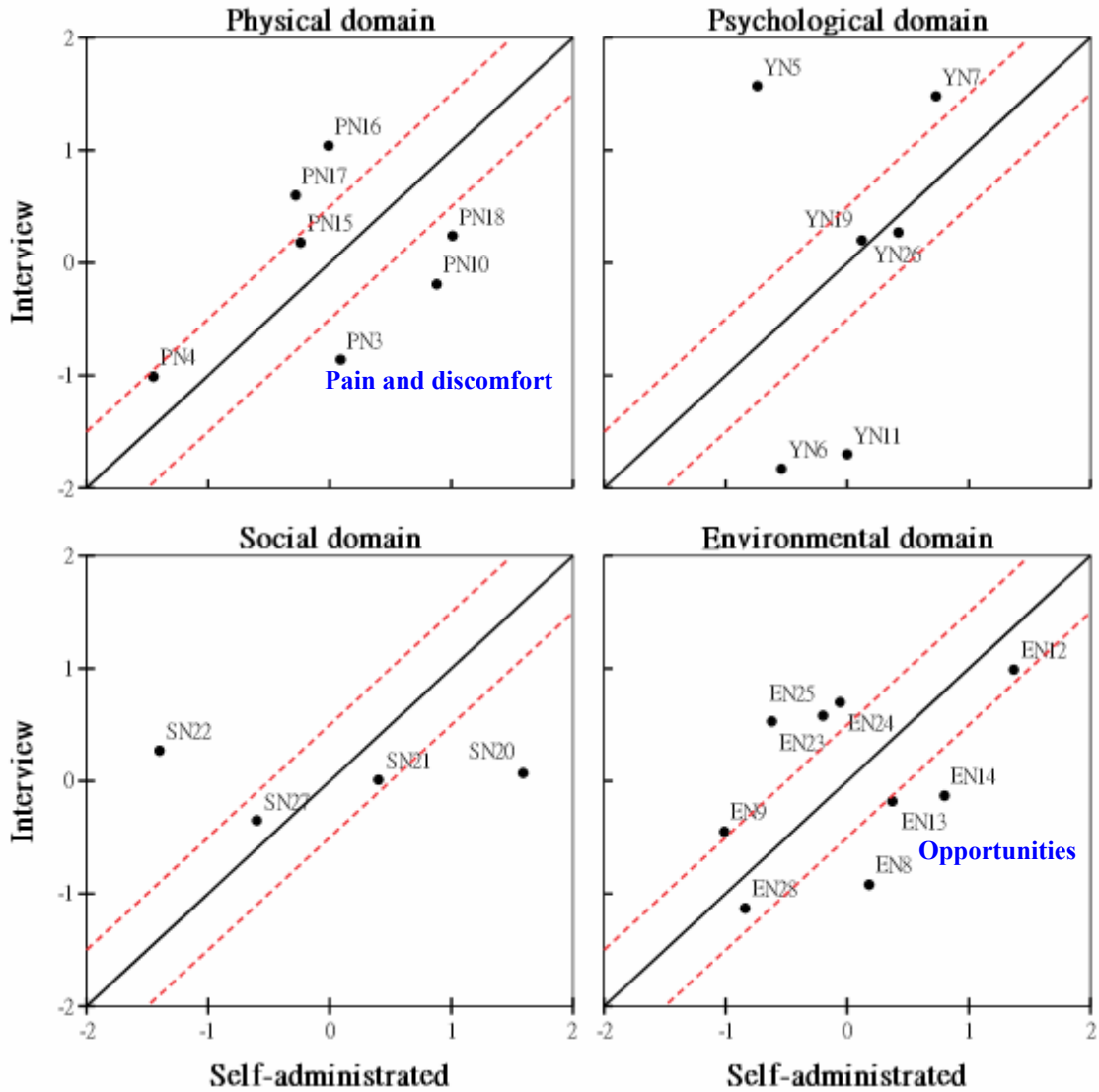


Figure13. DIF analysis among the mode of self-administration first and the mode of face-to-face interview first in each domain of WHOQOL-BREF for age < 65 years

項目反應理論分析年齡大於 65 歲在自填與面訪心理特質

表十六為年齡大於 65 歲組在自填與面訪的試題難度與相關指標，其中適合度指標(infit statistic)在 WHOQOL-BREF 的四個範疇中，生理範疇中的 PN4 ” 依賴醫療的幫助應付日常生活 ” infit 值為 1.6，心理範疇中的 YN26 ” 負面感覺 ” infit 值為 1.59，社會範疇的 SN21 ” 性生活 ” infit 值為 1.42，這三題未符合單一向度的假設。在信度分析當中，最差的仍是在社會範疇，自填與面訪信度只有 0.64 與 0.35。

試題難度上，生理範疇中，自填組範圍為(-1.13~0.86)，最難的 PN18 ” 滿意自己的工作能力 ” (-1.13~0.86)、最簡單 PN10 ” 足夠的精力 ” (1.13)；面訪組的範圍(-1.08~0.94)，最難的 PN17 ” 滿意日常活動的能力 ” (0.94)，最簡單 PN3 ” 疼痛是否會妨礙需要處理的事 ” (-1.08)；心理範疇中，自填難度範圍為(-1.13~0.79)，最難 YN11 ” 接受自己的外表 ” (0.79)，最簡單 YN7 ” 集中精神的能力 ” (-1.13)；面訪難度範圍為(-0.15~0.15)，最難 YN7 ” 集中精神的能力 ” (0.15)，最簡單 YN11 ” 接受自己的外表 ”；社會範疇中，自填難度範圍為(-0.9~1)，最難 N22 ” 滿意朋友支持 ” (1)、最簡單 SN27 ” 覺得自己有面子或被尊重嗎 ” (-0.9)，面訪難度範圍為(-0.65~0.68)，最難 N22 ” 滿意朋友支持 ” (0.68)、最簡單 SN27 ” 覺得自己有面子或被尊重嗎 ” (-0.65)；在環境範疇之中，自填組難度範圍為(-1.35~1.22)，最難的 EN13 ” 日常生活所需資訊 ” (1.22)，最簡單 EN8 ” 感到安全 ” (-1.35)，面訪組難度範圍為(-0.96~0.41)，最難的 EN9 ” 所處環境健康 ” (0.41)，最簡單 EN28 ” 想吃的食物能吃到 ” (-0.96)。

難度順序差別較顯著的題目，生理範疇有 PN3 ” 疼痛會妨礙需要做的事 ”，心理範疇有 YN7 ” 集中精神的能力 ” 和 YN11 ” 能接受自己的外表 ”，環境範疇 EN13 ” 方便得到資訊 ”。圖十四為年齡大於 65 歲自填與面訪的 DIF 情形，一共有 9/26 題 DIF。

Table 16. Item difficulty, infit statistics, and personal separation reliability of each item among self-administration first group and face-to-face interview first group for age ≥ 65 years (arranged by item difficulty within each domain).

Domains / Item	Self-administration		Face-to-face interview	
	Item difficulty	Infit Mean Square	Item difficulty	Infit Mean Square
Physical domain	Reliability:0.61		Reliability:0.75	
PN10. Energy and fatigue	-1.13	0.93	0.46	0.7
PN15. Mobility	-0.52	0.7	-0.59	0.95
PN4. Dependence on medication or treatment	-0.04	1.6	-0.78	1.33
PN16. Sleep and rest	0.07	0.86	0.23	1.24
PN3. Pain and discomfort	0.13	1.25	-1.08	1.04
PN17. Activities of daily living	0.63	0.78	0.94	0.85
PN18. Working capacity	0.86	0.8	0.82	0.86
Psychological domain	Reliability:0.54		Reliability:0.71	
YN7. Thinking, learning, memory & concentration	-1.13	1.04	0.15	0.82
YN26. Negative feeling	-0.13	1.59	0.01	1.37
YN5. Enjoyment of life	-0.07	0.92	0.03	0.83
YN19. Self-satisfaction	0.21	0.78	-0.02	0.84
YN6. Spirituality, religion and personal beliefs	0.33	0.73	-0.02	1.05
YN11. Body image and appearance	0.79	0.79	-0.15	0.95
Social domain	Reliability:0.64		Reliability:0.35	
SN27. Esteemed and respected	-0.9	1.14	-0.65	1.32
SN21. Sexual activity	-0.13	0.94	-0.56	1.42
SN20. Personal relationship	0.03	0.84	0.53	0.7
SN22. Friends' support	1	0.95	0.68	0.7
Environmental domain	Reliability:0.78		Reliability:0.62	
EN8. Physical safety and security	-1.35	0.81	0.04	0.85
EN25. Transportation	-0.91	1.25	0.22	0.92
EN28. Eating food	-0.83	1.06	-0.96	0.98
EN14. Participation & support of leisure activities	-0.12	1.17	0.19	0.92
EN12. Financial environment	0.1	0.94	-0.07	1.15
EN23. Home environment	0.39	0.76	0.29	0.82
EN24. Health & social care: availability & quality	0.42	1.15	0.16	0.95
EN9. Physical environments	1.08	1.02	0.41	1.29
EN13. Opportunities	1.22	0.88	-0.28	1.05

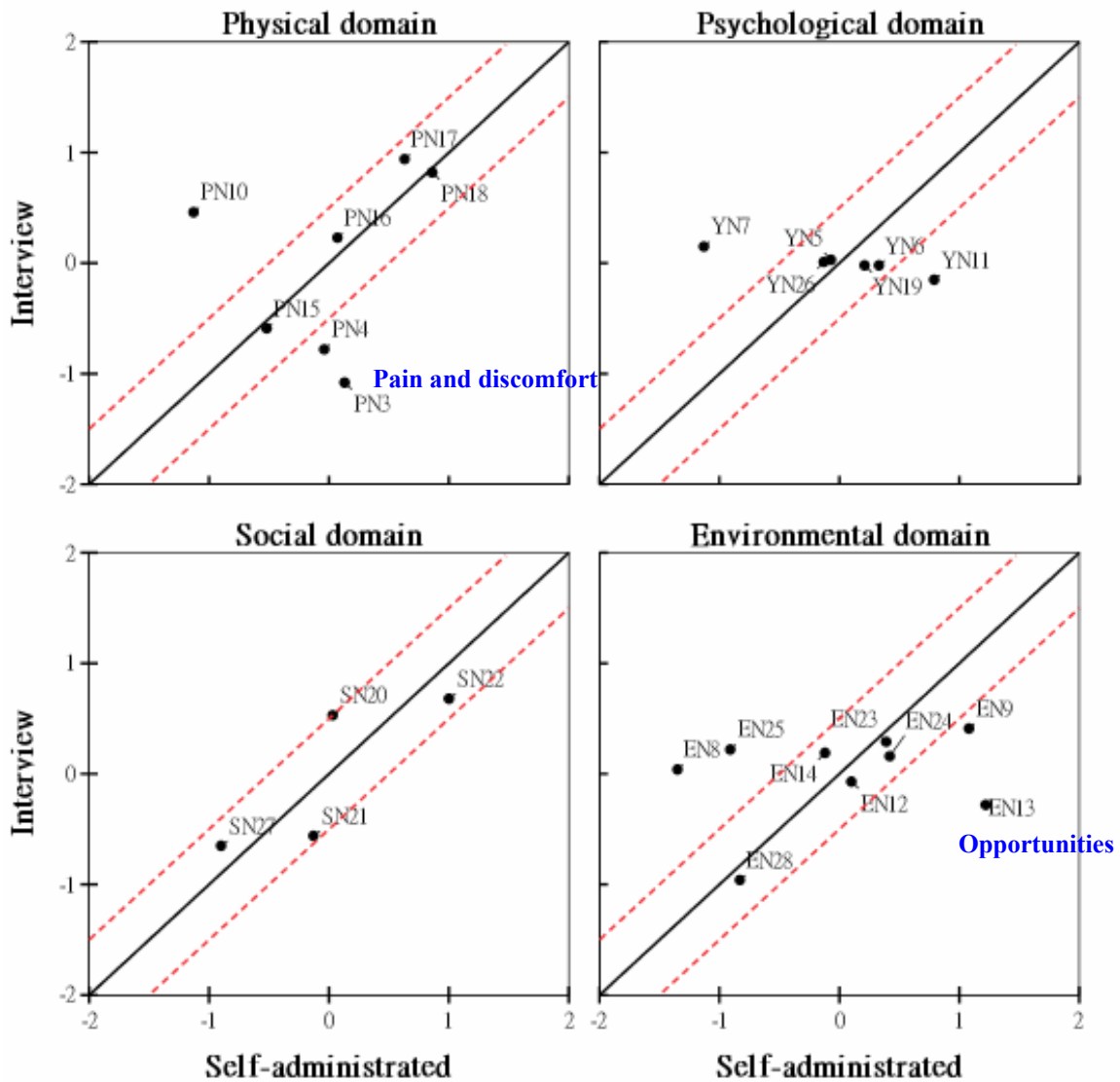


Figure14. DIF analysis among the mode of self-administration first and the mode of face-to-face interview first in each domain of WHOQOL-BREF for age ≥ 65 years.

第五章 討論

不論先自填組或先面訪組，亦即不論收樣順序的前後，整體而言，面訪分數均比自填分數高。過去，Lyons⁴⁸使用隨機交叉研究設計的報告指出，針對醫院門診或出院病人，不論面訪或自填的次序，亦發現使用 SF-36 時發現面訪的分數均高於自填。

在表3中，先自填內在等級相關係數指標(ICC):社會0.56、環境0.66、心理0.68和生理0.78，先面訪組所有範疇都大於0.7，先面訪組的再現性較先自填組好。Leeuw的後設分析結果發現，因為進行面訪時，回答者面對訪員的時候，相對可以給予回答者較強的刺激思考問題，所以填答問題時印象較為深刻，所以，在本研究的先面訪組中，因採取先面訪的形式，給予受測者的刺激較為強烈，故接下來自填問卷時，答題的再現性就容易較高³⁵。

雖然先自填組ICC數值表現較先面訪組差，更顯示出自填與面訪的差異。因WHOQOL-BREF原先即建議使用自填方式，希望受訪者能「主觀」自覺生活品質狀況，而未受其他第三者或觀察者的認定(包括訪員或醫師)；同時，在FDA的PRO 準則中，也強調PRO「直接」由病人方面所測得到的健康狀況方面的測量，而非經過醫生或其他人解釋所得的結果，所以自填法相較於第三者介入的面訪法，更貼切病人「主觀感覺」的重要性。本研究的ICC數值表現，更顯示出以面訪取代自填，對原先以自填方式為主要的研究，要注意其可能產生的偏差。

從不同人口特質探討先自填組組內 ICC 值的表現，可以看到在男性的表現比女性差，尤以社會範疇(ICC=0.5)最差，年齡大於 65 歲表現較年齡小於 65 歲差，其中以社會範疇(ICC=0.4)最差；教育程度在國中以下的 ICC 表現較差，其中也以社會範疇(ICC=0.5)最差；看年齡與性別的交互作用，65

歲以上男性老人 ICC 表現最差，尤其是社會範疇(ICC=0.38)。社會範疇再現性差的原因可能主要受到社會期望效果(social desirability effects)的影響。Kiesler & Sproull 指出，研究者採取的收樣方式會影響受訪者的反應型態和社會期望效果；Sproull 指出，人們在回答問題時，面對人和自己單獨時的答案會不一樣，研究結果顯示面對人回答問題時，其答案容易傾向人們「喜歡」、「期待」的答案¹⁰³。Tourangeau 和 Presser 過去評估不同收樣模式之間的差異性指出：比起自填，回答者通常在面訪或電訪時，在於回答中表現出更多的正面與社會期待性的回答^{38,46,47,44}。Kiesler 更指出，當其問題是性質屬於較敏感的，尤其在性行為時，面訪比起自填更容易有期待性偏差¹⁰³。Sanitaria 在西班牙的 SF-36 隨機分派交叉研究中，比較自我填答以及電訪的組內差異，發現在情感角色上(emotional role)與社會功能(social function)的 ICC 值表現最不一致¹⁰⁴。Hebert 在評估關於性別差異的肥胖相關自我評估時，發現男性較容易有社會期待性偏差，並且在自我評估報告中比起女性明顯高估¹⁰⁵。

另外，看 WHOQOL-BREF 的內在一致性(internal consistency coefficients)使用古典測驗理論(Cronbach's α)與項目反應理論(Person separation reliability)分析，兩者趨勢類似，發現在社會範疇表現最差，比較 Yao⁹¹ 臺灣簡明版生活品質問卷所研究的相關信度的驗證，雖然四個範疇信度值均大於 0.75，但是其中仍以社會範疇(0.75)的信度為最低，而在生理(0.8)與環境範疇(0.8)信度較高，信度的高低趨勢與本研究結果類似。

在 IRT 模式適合度分析的結果顯示：在反向題 PN4”需要醫療應付日常生活”和 YN26”負面感覺” PN3”疼痛會妨礙需要做的事”，其 infit 指標大於 1.4，不符合單一向度的假設。過去有相當多的研究，證實了反向題(reversed item)會降低其單一向度指標。Yao 測量 WHOQOL-BREF 其試題與

概念的相關，其在心理範疇中 YN26” 負面感覺”與心理範疇的相關最低⁹¹。Joel 曾經對測量相關議題探討，反向題會造成假設單一向度指標的下降，並且這個現象，也有相當多的多試題、多向度的研究證實，尤其在測量單一向度，其試題是正反向題並存時，更容易在測量時造成混淆^{108,109}。反向題雖然容易造成降低試題的同一概念假設，但並不代表在問卷的設計上就必須避免反向題，實際上，反向題的存在，因打亂了一部份回答的順序，而能得到更準確結果¹⁰⁸。

另外，在 65 歲以下的族群，其在 PN16”睡眠狀況”也不符合在生理範疇之中單一向度的假設；65 歲以上族群其在 SN21”滿意性生活”亦不符合在社會範疇之中單一向度的假設。本研究當中 65 歲以下的族群其年齡介於 40~64 歲，年齡為中壯年族群，其造成不符合單一向度的假設的原因可能在於，在台灣此族群為工作、家庭責任的壓力較大，而影響睡眠問題，故在此族群中，其睡眠問題多屬於心理層面所造成，而較不是生理層面影響睡眠。SN21”滿意性生活”在 65 歲以上族群，因喪偶、官能退化等問題，故性生活對此族群而言則不符合其單一向度假設。

本研究結果顯示在不同的管理模式，題目難度順序較顯著不同的題目如：在反向題”負面感覺”男性在自填較面訪容易傾向有負面得分，女性則相反；老年人在”接受自己的外表”，自填較面訪容易傾向不滿意的感覺；而在字面上較抽象的題目，如”性行為”、“日常生活資訊”及”集中精神的能力”等，需要訪員進一步解釋題意，也會造成受訪者對不同管理模式之間的感受不同。Lyons 在研究 SF-36 的報告，採用郵寄自填與面訪隨機交叉方式觀察病人在得分上的差異，研究指出，雖然交叉隨機分派的方式可以大幅降低訪問順序差異 (order effect)，但是收樣方式的不同卻造成得分的不同 (mode of administration effect)，並且在關及敏感的範疇差異更顯著⁴⁸。此外，年齡、

性別的不同，對於不同收樣方式的差異量上也不同，其研究結果顯示：男性差異較女性來的大，並且年齡越大，差異越大。

我們看到困難度的順序，在 YN26”負面感受”這一題，男性與女性的表現正好相反，男性在自填時較容易有負面感受，但在面訪時卻較沒有；女性在面訪時較容易有負面感受，但在自填卻較沒有。在台灣社會中，雖然現代提倡男女平等風氣盛行，但在本研究受訪者的成長年代，多仍是在過去父系社會環境下成長，潛在的男性主義(Masculinity)的教導著男性應該越具有男子氣概越好、果斷、工作、晉升、成功、和金錢是重要的，女性應該是順從、溫柔的，所以訪員在詢問其負面感覺時，男性有可能較不會將真實的感覺透露出來，導致在性別上兩個結果相反。在全部和 65 歲以上兩組，對於接受自己外表在面訪中容易接受自己的外表，在自填中卻不容易接受。這可能是社會期待性偏差所致，老年人身體老化和官能、外表慢慢地退化中，其實是較不容易接受自己的外貌，但在面訪時，因面臨第三者，難為情將容貌的不滿說出，所以在訪員面前回答較能接受自己的外表所致。

Steinbuechel 研究使用 WHOQOL-BREF 的因素分析結果指出，年齡的不同會影響某些題目，譬如說，在老年人當中，其在性行為的滿意度這一題的表現較不符合。另外，在試題平均得分來看，”接受自己的外表”和”生命有意義”這兩題對於老年人與年輕人的意義是不同的。他也發現，男性與女性在”社會支持”和”負面感覺”，其結果並不相同，女性較滿意社會支持，但也較容易有負面感覺。但並沒有比較實際客觀的狀況，例如經濟來源以及健康照護系統和個人的期望的試題得分比較¹¹⁰。

PN14“日常生活資訊”在面訪顯得較簡單，在自填卻較難。這題完整的題目呈現為「您能方便得到每日生活所需的資訊嗎？」受測者在作答時，普遍因題目較抽象，難以意會，相形在面訪時許多受訪者不懂題目意思，需要

訪員進一步解釋題意，而受訪者反應比較自填方式簡單所致。

同樣受到管理模式解釋力不同造成回答者的感受不同的還有在老年人組的 YN7”集中精神的能力好嗎”造成影響，本題最主要是在詢問關於思考、學習或者是在記憶方面時的專注能力，但在自填時，容易將其題目解讀為單純集中精神的能力，而訪問時，訪員除了唸出題目外，並對於受訪者對題目的疑問加以解釋，所以在實際使用面訪和自填時的反應有落差，尤其是對於老年人而言，因對於「集中精神的能力」隨著官能的退化而變差，在這兩種模式間的反應差異性就較為顯著。

試題差別表現(differential item functioning, DIF) 結果顯示在”身體疼痛會妨礙您所需要做的事”、“性生活”和”方便得到資訊”在不同性別上兩種模式均出現明顯偏差，尤其”性生活”這題在不同年齡層表現亦不同。過去 Naumann 使用 WHOQOL-BREF 測試憂鬱症老人其結果的信效度，發現除了關於滿意自己的人際關係和性生活這兩題表現之外，其他都算是令人滿意的¹⁰⁶。Skevington 測試 WHOQOL 的跨文化差異，其中報告指出，在這過程中他發現對於某一些族群，像是認知障礙、溝通困難、或是中風等，都會影響到使用 WHOQOL-BREF 的適用性，另外，尤其在年齡大於 65 歲以上的族群，因需求不同，建議是另外再發展一份 WHOQOL-Old，較符合 65 歲以上老年人的特殊需求¹⁰⁷。

研究中的有病族群，囊括了 13 種疾病（高血壓、腎病、牙周病、心血管疾病、糖尿病、肺病、B 型肝炎、消化性潰瘍、高血脂病、中風、攝護腺癌、攝護腺肥大與其他疾病），這些疾病的表現特徵不同，本身影響生活品質的原因也不同，所以如要進一步瞭解其在收樣模式的差異，必須要區分其疾病影響特性分組分析；同理可證，在教育程度方面，也因涉及多層面多變異的關係，會使結果的呈現不穩定。其他進一步的交互作用分析，例如：性

別與年齡（65 歲以下男性、65 歲以下女性、65 歲以上男性和 65 歲以上女性）因樣本數不足而受限。欲探究這些人口特性的在不同管理模式上的心理特性差異，最好能有較多的樣本支持分析，本研究樣本數限制，以較穩定且足夠的樣本數來看性別、年齡分層，探究其心理特性差異。

綜合上面的古典測驗理論和項目反應理論的結果，由 CTT 的結果觀察到面訪範疇平均得分高於自填平均得分，細切至看單題平均值的結果也是如此。古典測驗理論看受測者在兩種模式間單題試題差異，用平均值、標準差、平均值差與檢定值，我們僅僅只能看出得分高低是否有顯著差異。基於這樣的結果，我們再用項目反應理論分析試題，在兩種管理模式之中依據其受測者的反應性而產生試題難度，在進而討論自填與面訪個別的試題難度順序，討論兩種模式在試題之中的感受性差異。而後再用人口特性去細分成男性、女性、65 歲以下和 65 歲以上四組看自填與面訪間的感受性。利用 IRT 針對試題及各不同族群有系統的一系列分析，對面訪與自填兩種模式的比較提供了 CTT 所無法表現的深層剖析。

第六章 結論與建議

第一節 結論

整體而言，不論何種施測模式的順序先後，WHOQOL-BREF全部範疇的面訪得分均高於自填得分。由人口特質來看：施測模式的不同在年齡大於65歲和男性兩組的結果差異較大。

內在等級相關係數指標(ICC)比較兩組的再現性，先面訪組(0.71~0.8)較先自填組(0.56~0.77)穩定，因為面訪模式使受測者印象較深刻所致，故再現性較高，雖自填法再現性較差，但自填法相較於第三者介入的面訪法，更貼切 PRO Guideline 中病人「主觀感覺」的重要性。進一步看年齡、性別分組的 ICC 值，發現男性、年齡大於 65 歲其 ICC 值較不穩定，社會期待性偏差顯見在這兩族群。

古典測驗理論(CTT)與項目反應理論(IRT)看問卷各範疇內部一致性，其生理、心理與環境範疇信度均不錯，但在社會範疇之中，IRT 與 CTT 信度均較差，結果與 WHOQOL-BREF 過去其他相關研究結果是吻合的。

項目反應理論看單一向度假設指標的結果顯示在兩種管理模式，反向題”需要醫療應付日常生活”、”負面感覺”和”疼痛會妨礙需要做的事”，不符合單一向度的假設，此 3 題均為反向題，反向問題會降低其試題在概念之中的單一向度假設。此外，老年人在社會範疇中的”性生活”以及中壯年在生理範疇中的”睡眠狀況”，因角色實際情形相較於試題不符合其範疇概念，所以結果顯示不符合單一向度的假設。

不同的管理模式，題目相對的難度順序亦不同，產生變化的主要原因有(1)較敏感的題目造成社會期待性偏差(2)不同的管理模式對試題的解釋力不同。在敏感的題目中變化較顯著的題目如：在反向題”負面感覺”男性在

面訪比起自填，較不容易有負面感覺，女性則相反，在面訪較容易有負面感覺；老年人在“接受自己的外表”，自填較面訪容易傾向有不滿意的感覺，因為問題較敏感，社會期待性偏差導致在兩種管理模式難度順序不同。因管理模式中對於受測者的解釋力不同造成難度順序不同在全部分組的“方便得到每日生活所需的資訊”和老年人“集中精神的能力”之中，因面訪模式可以對題目提供較多訊息和解釋，造成差異。

用 IRT 看兩種管理模式與題目特質之間的關係，其不分組以及分男女兩組的結果均顯示在生理範疇 PN3“身體疼痛會妨礙您所需要做的事”、SN21“性生活”和 EN13“方便得到資訊”則出現嚴重 DIF，顯示這些題目在這兩種模式間相當不穩定。

第二節 研究限制

1. 本研究並沒有做訪員間信度(Inter-rater reliability)，因在健診中心的參與健檢者有季節性的不同，而導致三位訪員所收案的族群分佈不均勻，所以無法切確測得其信度，但我們收樣的所在場地是在醫院的健診中心，受訪者填寫環境寧靜不被打擾，且配合度亦較高，並且訪員訓練過程都有嚴格的標準化，並且不定期抽檢，所以訪員間的可信度仍可以維持相當高的可信度。

2. 本研究收樣標準需兩種模式均可以完成才納入，故其中不識字者、視力因素或其他原因無法完成兩份問卷者則不納入研究。

3. 本研究採隨機收樣的方式，故其兩組的人口特質分層很均勻，可比性也很高，仍然可以看出整體大致趨勢。

4. 本研究也因樣本數問題，無法做更細部的分層與交互作用，進一步探究各不同人口族群的表現。

第三節 應用與建議

1. 研究兩種管理模式施測世界衛生組織生活品質簡明版問卷 (WHOQOL-BREF) 得分差異，結果面訪高於自填，建議未來在使用其他問卷施測時，若不瞭解其問卷的心理特性可能所造成的偏差相關訊息時，則要避免合併管理模式使用，以免產生過多的偏差削弱了真實分數。
2. WHOQOL-BREF 在兩種管理模式之中，受測者為 65 歲以上及男性得分差異較大，並且在四個範疇中，以社會範疇相差最大。
3. 在敏感的題目，例如性生活、負面感覺、接受自己的外表，在自填與面訪的反應性不同，若未來採取面訪模式收案時，建議在研究方法上改良，以減低偏差。
4. 面訪模式因有訪員的引導，所以對於用詞較艱深或題意含糊不清的題目可以立即釐清受訪者對於題目的困惑，建議問卷設計的題目用詞盡可能明確、簡單明瞭。
5. 透過瞭解 WHOQOL-BREF 在管理模式差別所在後，對於其可能所造成的偏差，加以調節，提高在臨床實際應用上合併管理模式使用時的測量準確性。

參考文獻

- 1 Goyder J. Face-to-face interviews and mailed questionnaires: The net difference in response rate. *Public Opin Q* 1985; 49:234-252
- 2 Wallace RB, Woolson RF. *The Epidemiologic Study of the Elderly*. New York: Oxford University Press, 1992
- 3 Fink P, Ornbol E, Huyse FJ, et al. A brief diagnostic screening instrument for mental disturbances in general medical wards. *J Psychosom Res* 2004; 57:17-24
- 4 Siemiatycki J. A comparison of mail, telephone, and home interview strategies for household health surveys, *Am J Public Health* 1979; 69: 238-245
- 5 Bassett SS, Magaziner J, Hebel JR. Reliability of proxy response on mental health indices for aged, community-dwelling women. *Psychol Aging* 1990; 5:127-132
- 6 Rothman ML, Hedrick SC, Bulcroft KA, et al. The validity of proxy-generated scores as measures of patient health status. *Med Care* 1991; 29:115-124
- 7 Magaziner J, Bassett SS, Hebel JR, et al. Use of proxies to measure health and functional status in epidemiologic studies of community-dwelling women aged 65 years and older. *Am J Epidemiol* 1996; 143:283-292
- 8 Magaziner J, Zimmerman SI, Gruber-Baldini AL, et al. Proxy reporting in five areas of functional status. Comparison with self-reports and observations of performance. *Am J Epidemiol* 1997; 146:418-428
- 9 Sneeuw KCA, Aaronson NK, Haan RJ, et al. Assessing Quality of life after stroke the value and limitations of proxy ratings. *Stroke* 1997; 28:1541-1549
- 10 WHO. *Field trial WHOQOL-100: Scoring the WHOQOL*. Geneva: World Health Organization, 1995
- 11 WHO. *Field trial WHOQOL-100: The 100 questions with response scales*. Geneva: World Health Organization, 1995
- 12 林茂榮, 姚開屏, 黃景祥, et al. 台灣版世界衛生組織生活品質問卷量尺語詞的選擇. *中華公共衛生雜誌* 1999; 18:262-270
- 13 Leung KF, Wong WW, Tay MS, et al. Development and validation of the interview version of the Hong Kong Chinese WHOQOL-BREF. *Qual Life Res* 2005; 14:1413-1419
- 14 Liang WM, Kuo HW, Lin CF, et al. Factor construct of health-related quality of life in Taiwanese workers by WHOQOL-BREF questionnaire. *Mid Taiwan J Med* 2005; 10:113-122
- 15 台灣版世界衛生組織生活品質問卷發展小組. *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(第一版)*. 台北: 國立台灣大學公衛學院生活品質研究室, 2000
- 16 Bjorner JB, Kreiner S, Ware JE, et al. Differential item functioning in the Danish translation of the SF-36. *J Clin Epidemiol* 1998; 51:1189-1202
- 17 Food US, Administration D. *Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims*: February, 2006
- 18 Hawthorne G. Measuring Social isolation in older adults: Development and initial validation of the friendship scale. *Soc Indic Res* 2006; 77:521-548
- 19 Weinberger M, Oddone EZ, Samsa GP, et al. Are health-related quality-of-life measures affected by the mode of administration? *J Clin Epidemiol* 1996; 49:135-140
- 20 Brambilla DJ, McKinlay SM. A comparison of responses to mailed questionnaires and telephone interviews in a mixed mode health survey. *Am J Epidemiol* 1987;

- 126:962-971
- 21 O'Toole BI, Battistutta D, Long A, et al. A comparison of costs and data quality of three health survey methods: mail, telephone and personal home interview. *Am J Epidemiol* 1986; 124:317-328
- 22 Bowling A. Mode of questionnaire administration can have serious effects on data quality. *J Public Health* 2005; 27:281-291
- 23 Bowling A. *Research Methods in Health: Investigating Health and Health Services*: Open University Press, 2001
- 24 Simon D, Schorr G, Wirtz M, et al. Development and first validation of the shared decision-making questionnaire. *Patient Educ Couns*, 2006; 63(3):319-27
- 25 McColl E, Jacoby A, Thomas L. Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients: Core Research on behalf of the NCCHTA, 2001
- 26 Cartwright A. *Health Surveys in Practice & Potential*. London: King Edward's Hospital Fund for London, 1989
- 27 Scott C. Research on mail surveys. *J R Stat Soc Ser A Stat Soc* 1961; 124:143-205
- 28 Austin P, Lewis D, Scammell B. A review of postal surveys: OPCS methodology series paper. London: Office for Population Censuses and Surveys, 1977
- 29 Walker A, Government statistical S, Office for national S. *Living in Britain: Results from the 2000 general household survey*: Stationery Office, 2001
- 30 Dillman DA. *Mail and internet surveys: the tailored design method*. New York: Wiley, 2000
- 31 Sykes W CM. *Effect of mode of interview: experiments in the UK*. New York: John Wiley and Sons, 1988
- 32 Gasquet I, Falissard B, Ravaud P. Impact of reminders and method of questionnaire distribution on patient response to mail-back satisfaction survey. *J Clin Epidemiol* 2001; 54:1174-1180
- 33 Measures O. Assessing inner-city patients' hospital experiences: A controlled trial of telephone interviews versus mailed surveys. *Med Care* 1997; 35:70-76
- 34 Brazier JE, Harper R, Jones NM, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ* 1992; 305:160-164
- 35 Leeuw ED. *Data quality in mail, telephone and face to face surveys*: ERIC, 1992
- 36 Frankfort-Nachmias C, Nachmias D. *Research methods in the social sciences*: Edward Arnold, 1976
- 37 Johnson WD, Mercante DE. Analyzing multivariate data in crossover designs using permutation tests. *J Biopharm Stat* 1996; 6:327-342
- 38 Tourangeau RK, Jobe JB, et al. Sources of error in a survey of sexual behaviour. *J Official Stats* 1997; 13:341-365
- 39 Tourangeau RK. *The psychology of survey response*; Chapter 10: Mode of data collection: Cambridge University Press, 2000
- 40 Bowling A, Bond M, Jenkinson C, et al. Short form 36 (SF-36) health survey questionnaire: which normative data should be used? Comparisons between the norms provided by the omnibus survey in Britain, the health survey for England and the oxford healthy life survey. *J Public Health Med* 1999; 21:255-270
- 41 Schuman PS. *Questions and answers in attitude surveys: Experiments on question form, wording, and context*: Social Science, 1981
- 42 Sudman BN. *Asking questions*. San Francisco: Jossey Bass, 1983

- 43 Schwarz N, Hippler HJ. Subsequent questions may influence answers to preceding questions in mail surveys. *Public Opin Q* 1995; 59:93-97
- 44 Mikkelsen RL, Middelboe T, Pisinger C, et al. Anxiety and depression in patients with chronic obstructive pulmonary disease (COPD). *Nord J Psychiatry* 2004; 58:65-70
- 45 Warner SL. Randomized response: a survey technique for eliminating evasive answer bias. *J Am Stat Assoc* 1965; 60:63-66
- 46 Tourangeau R, Smith TW. Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opin Q* 1996; 60:275-304
- 47 Presser S, Stinson L. Data collection mode and social desirability bias in self-reported religious attendance. *Am Sociol Rev* 1998; 63:137-145
- 48 Lyons RA, Wareham K, Lucas M, et al. SF-36 scores vary by method of administration: implications for study design. *J Public Health Med* 1999; 21:41-45
- 49 Vuillemin A, Oppert JM, Guillemin F, et al. Self-administered questionnaire compared with interview to assess past-year physical activity. *Med Sci Sports Exerc* 2000; 32:1119-1124
- 50 Urbach DR, Tomlinson GA, Harnish JL, et al. A measure of disease-specific health-related quality of life for achalasia. *Am J Gastroenterol* 2005; 100:1668-1676
- 51 McHorney CA, Kosinski M, Ware JE, Jr. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. *Med Care* 1994; 32:551-567
- 52 Perkins JJ, Sanson-Fisher RW. An examination of self- and telephone-administered modes of administration for the Australian SF-36. *J Clin Epidemiol* 1998; 51:969-973
- 53 Garcia-Losa M, Unda M, Badia X, et al. Effect of mode of administration on I-PSS scores in a large BPH patient population. *Eur Urol* 2001; 40:451-457
- 54 Rhodes T, Girman CJ, Jacobsen SJ, et al. Does the mode of questionnaire administration affect the reporting of urinary symptoms? *Urology* 1995; 46:341-345
- 55 Klepac RK, Dowling J, Rokke P, et al. Interview vs. paper-and-pencil administration of the McGill pain questionnaire. *Pain* 1981; 11:241-246
- 56 Fowler FJ, Gallagher PM, Nederend S. Comparing telephone and mail responses to the CAHPS survey instrument. *Med Care* 1999; 37:MS41-49
- 57 Adams R, Rosier M, Campbell D, et al. Assessment of an asthma quality of life scale using item-response theory. *Respirology* 2005; 10:587-593
- 58 Bozlu M, Doruk E, Akbay E, et al. Effect of administration mode (patient vs physician) and patient's educational level on the Turkish version of the international prostate symptom score. *Int J Urol* 2002; 9:417-421
- 59 Durant LE, Carey MP. Self-administered questionnaires versus face-to-face interviews in assessing sexual behavior in young women. *Arch Sex Behav* 2000; 29:309-322
- 60 Kaplan CP, Hilton JF, Park-Tanjasiri S, et al. The effect of data collection mode on smoking attitudes and behavior in young African American and Latina women: Face-to-face interview versus self-administered questionnaires. *Eval Rev* 2001; 25:454-473
- 61 Nebot M, Celentano DD, Burwell L, et al. AIDS and behavioural risk factors in women in inner city Baltimore: a comparison of telephone and face to face surveys. *J Epidemiol Community Health* 1994; 48:412-418
- 62 Nicolaas G, Thomson K, Lynn P, et al. The feasibility of conducting electoral surveys in the UK by telephone: National Centre for Social Research, 2000
- 63 Midanik LT, Greenfield TK. Telephone versus in-person interviews for alcohol use:

- results of the 2000 national alcohol survey. *Drug Alcohol Depend* 2003; 72:209-214
- 64 Evans M, Kessler D, Lewis G, et al. Assessing mental health in primary care research using standardized scales: can it be carried out over the telephone? *Psychol Med* 2004; 34:157-162
- 65 Fouladi RT, McCarthy CJ, Moller NP. Paper-and-pencil or online? Evaluating mode effects on measures of emotional functioning and attachment. *Assessment* 2002; 9:204-215
- 66 Harewood GC, Yacavone RF, Locke GR, 3rd, et al. Prospective comparison of endoscopy patient satisfaction surveys: e-mail versus standard mail versus telephone. *Am J Gastroenterol* 2001; 96:3312-3317
- 67 Bower WF, Wong EM, Yeung CK. Development of a validated quality of life tool specific to children with Bladder dysfunction. *Neurourol Urodyn* 2006; 25:221-227
- 68 Bowling A. *Research methods in health: Investigating health and health services*: Open University Press, 1997
- 69 Sudman S BN. *Asking questions*. San Francisco: Jossey Bass, 1983
- 70 Knauper B, Schwarz N. Why your research may be out of order. *Psychologist* 2004; 17:28-31
- 71 Krosnick JA, Alwin DF. An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opin Q* 1987; 51:201-219
- 72 Schwarz N, Knauper B, Hippler HJ, et al. Rating scales: numeric values may change the meaning of scale labels. *Public Opin Q* 1991; 55:570-582
- 73 Sudman S, Bradburn NM, Schwarz N. *Thinking about answers: the application of cognitive processes to survey methodology*: Jossey-Bass, 1996
- 74 Knauper B. The impact of age and education on response order effects in attitude measurement. *Public Opin Q* 1999; 63:347-370
- 75 Park DC. *The basic mechanisms accounting for age-related decline in cognitive function. Cognitive aging: A primer*, 2000
- 76 Schwarz N. *Cognition, aging, and self-reports*: psychology press philadelphia: PA, 1998
- 77 Dorant E, van den Brandt PA, Goldbohm RA, et al. Agreement between interview data and a self-administered questionnaire on dietary supplement use. *Eur J Clin Nutr* 1994; 48:180-188
- 78 Pruchno RA, Hayden JM. Interview modality: effects on costs and data quality in a sample of older women. *J Aging Health* 2000; 12:3-24
- 79 Parker C, Dewey M. Assessing research outcomes by postal questionnaire with telephone follow-up. *Int J Epidemiol* 2000; 29:1065-1069
- 80 Bradburn NM. Response effects. *Handbook of Survey Research*: Academic Press, 1983
- 81 Turner CF, Ku L, Rogers SM, et al. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science* 1998; 280:867-873
- 82 Dooley D. *Social research methods*: Prentice Hall, 1990
- 83 Groves RM, Kahn RL. *Surveys by telephone: a national comparison with personal interviews*: Academic Press, 1979
- 84 Ryan JM, Corry JR, Attewell R, et al. A comparison of an electronic version of the SF-36 general health questionnaire to the standard paper version. *Qual Life Res* 2002; 11:19-26
- 85 Bowling A, Bannister D, Sutton S, et al. A multidimensional model of quality of life in older age. *Aging Ment Health* 2002; 6:355-371

- 86 Hambleton RK. Good practices for identifying differential item functioning. *Med Care* 2006; 44:S182-S188
- 87 Staquet MJ, Hays RD, Fayers PM. Quality of life assessment in clinical trials: methods and practice. New York: Oxford Univeristy Press, 1999
- 88 Embretson SE, Reise SP. Item response theory for psychologists: Lawrence Erlbaum Assoc Inc, 2000
- 89 Uttaro T, Lehman A. Graded response modeling of the quality of life interview. *Eval Program Plann* 1999; 22:41-52
- 90 Garratt AM. Rasch analysis of the Roland disability questionnaire. *Spine* 2003; 28:79-84
- 91 Wang WC, Yao G, Tsai YJ, et al. Validating, improving reliability, and estimating correlation of the four subscales in the WHOQOL-BREF using multidimensional Rasch analysis. *Qual Life Res* 2006; 15:607-620
- 92 Roorda LD, Jones CA, Waltz M, et al. Satisfactory cross cultural equivalence of the Dutch WOMAC in patients with hip osteoarthritis waiting for arthroplasty. *Ann Rheum Dis* 2004 ;63:36-42
- 93 Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. *J Am Geriatr Soc* 1975; 23:433-441
- 94 Martinez J, Duenas R, Onis MC, et al. Adaptacion validacion al castellano del cuestionario de Pfeiffer (SPMSQ) para detectar la existencia de deterioro cognitivo en personas mayores de 65 anos. *Med Clin (Barc)* 2001; 117:129-134
- 95 張家銘, 蔡智能. 老年人之周全性評估. *台灣醫學* 2003; 7:364-374
- 96 Hwang HF, Liang WM, Chiu YN, et al. Suitability of the WHOQOL-BREF for community-dwelling older people in Taiwan. *Age Ageing* 2003; 32:593-600
- 97 Fenton KA, Johnson AM, McManus S, et al. Measuring sexual behaviour: methodological challenges in survey research. *Sex Transm Infect* 2001; 77:84-92
- 98 Rejean H ea. Refusal and information bias association with postal questionnaire and face-to-face interviews in very elderly subjects. *J Clin Epidemiol* 1996; 49:373-381
- 99 Carmines EG, Zeller RA. Reliability and validity assessment: Sage Publications, 1979
- 100 Shrout PE FJ. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86:420-428
- 101 Chu M. Arthritis impact measurement scales 2: translation, validation, utilization in clinical studies and the development of a short form: University of Hong Kong , 2005
- 102 Lawton G, Bhakta BB, Chamberlain MA, et al. The Behcet's disease activity index. *Rheumatology*. *Rheumatology* 2004; 43:73-78
- 103 Kiesler S, Sproull LS. Response effects in the electronic survey. *Public Opin Q* 1986; 50:402-413
- 104 Sanitaria G. Comparison between telephone and self-administration of short form health survey questionnaire (SF-36). *Gac Sanit* 2005; 19:6-15
- 105 Hebert JR, Ma Y, Clemow L, et al. Gender differences in social desirability and social approval bias in dietary self-report. *Am J Epidemiol* 1997; 146:1046-1055
- 106 Naumann VJ, Byrne GJ. WHOQOL-BREF as a measure of quality of life in older patients with depression. *Int Psychogeriatr* 2004; 16:159-173
- 107 Skevington SM. Advancing cross-cultural research on quality of life: observations drawn from the WHOQOL development. *Qual Life Res* 2002; 11:135-144
- 108 Herche J, Engelland B. Reversed-polarity items and scale unidimensionality. *Int*

- Psychogeriatr 1996; 24:366-374
- 109 Reckase MD, Ackerman TA, Carlson JE. Building a unidimensional test using multidimensional items. J Educ Res 1988; 25:193-203
- 110 Steinbuchel N, Lischetzke T, Gurny M, et al. Assessing quality of life in older people: psychometric properties of the WHOQOL-BREF. Eur J Ageing 2006; 3:116-122
- 111 Brown Jr BW. The crossover experiment for clinical trials. Biometrics 1980; 36:69-79



世界衛生組織生活品質問卷(WHOQOL-BREF)

請您用自己的標準、希望、愉快、以及關注點，針對您最近兩星期內的生活情形，來回答問題。

1. 整體來說，您如何評價您的生活品質？
極不好 不好 中等程度好 好 極好
2. 整體來說，您滿意自己的健康嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
3. 您覺得身體疼痛會妨礙您處理需要做的事情嗎？
完全沒有妨礙 有一點妨礙 中等程度妨礙 很妨礙 極妨礙
4. 您需要靠醫療的幫助應付日常生活嗎？
完全沒有需要 有一點需要 中等程度需要 很需要 極需要
5. 您享受生活嗎？
完全沒有享受 有一點享受 中等程度享受 很享受 極享受
6. 您覺得自己的生命有意義嗎？
完全沒有 有一點有 中等程度有 很有 極有
7. 您集中精神的能力有多好？
完全不好 有一點好 中等程度好 很好 極好
8. 在日常生活中，您感到安全嗎？
完全不安全 有一點安全 中等程度安全 很安全 極安全
9. 您所處的環境健康嗎？(如污染、噪音、氣候、景觀)
完全不健康 有一點健康 中等程度健康 很健康 極健康
10. 您每天的生活有足夠的精力嗎？
完全不足夠 少許足夠 中等程度足夠 很足夠 完全足夠
11. 您能接受自己的外表嗎？
完全不能夠 少許能夠 中等程度能夠 很能夠 完全能夠
12. 您有足夠的金錢應付所需嗎？
完全不足夠 少許足夠 中等程度足夠 很足夠 完全足夠
13. 您能方便得到每日生活所需的資訊嗎？
完全不方便 少許方便 中等程度方便 很方便 完全方便
14. 您有機會從事休閒活動嗎？

- 完全沒有機會 少許機會 中等程度機會 很有機會 完全有機會
15. 您四處行動的能力好嗎？
完全不好 有一點好 中等程度好 很好 極好
16. 您滿意自己的睡眠狀況嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
17. 您對自己從事日常活動的能力滿意嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
18. 您滿意自己的工作能力嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
19. 您對自己滿意嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
20. 您滿意自己的人際關係嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
21. 您滿意自己的性生活嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
22. 您滿意朋友給您的支持嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
23. 您滿意自己住所的狀況嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
24. 您對醫療保健服務的方便程度滿意嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
25. 您滿意所使用的交通運輸方式嗎？
極不滿意 不滿意 中等程度滿意 滿意 極滿意
26. 您常有負面的感受嗎？（如傷心、緊張、焦慮、憂鬱等）
從來沒有 不常有 一半有一半沒有 很常有 一直都有
27. 您覺得自己有面子或被尊重嗎？
完全沒有 有一點有 中等程度有 很有 極有
28. 您想吃的食物通常都能吃到嗎？
從來沒有 不常有 一半有一半沒有 很常有 一直都有

簡易心智狀態問卷調查表(SPMSQ)

姓 名：_____ 日 期：_____

錯誤請打 X	問 題	注 意 事 項
	1.今天是幾號？	年、月、日都對才算正確。
	2.今天是星期幾？	星期對才算正確。
	3.這是什麼地方？	對所在地有任何的描述都算正確； 說「我的家」或正確說出城鎮、醫院、機構的名稱都可接受。
	4-1.您的電話號碼是幾號？	經確認號碼後証實無誤即算正確； 或在會談時，能在二次間隔較長時間內重覆相同的號碼即算正確。
	4-2.您住在什麼地方？	如長輩沒有電話才問此問題。
	5.您幾歲了？	年齡與出生年月日符合才算正確。
	6.您的出生年月日？	年、月、日都對才算正確。
	7.現任的總統是誰？	姓氏正確即可。
	8.前任的總統是誰？	姓氏正確即可。
	9.您媽媽叫什麼名字？	不需要特別証實，只需長輩說出一個與他不同的女性姓名即可。
	10.從 20 減 3 開始算，一直減 3 減下去。	期間如有出現任何錯誤或無法繼續進行即算錯誤。

進行方式：依下表所列的問題，詢問長輩並將結果紀錄下來，(如果長輩家中沒有電話，可將 4-1 題改為 4-2 題)，答錯的問題請記錄下來。