

中國醫藥大學

碩士論文

編號：IEH-1611

以 Rasch 模式分析世界衛生組織生活品質
問卷簡明版在社區老人的心理計量特質

**Rasch model analysis for studying
psychometric characteristics of the
WHOQOL-BREF in community-dwelling
older people**

所別：環境醫學研究所

指導教授：梁文敏、林茂榮

學生：葉懿諄 Yeh Yi-Chun

學號：9365011

中華民國九十五年六月

誌謝

時間一下子就過去了，感覺真的過得很快。感覺好像才剛考試研究所而已，沒想到已經要畢業了。在這段過程中，有各位同學們，曜慶、媛婷、懷芝、舒婷、玟陵、佳蓉、鈞萍、錦蓉、心縵及秀貞等等，還有許多沒有提名的同學們，在課業上一同的學習使我獲益良多。

記得剛到梁老師辦公室時，就去香港參加 ISOQOL 研討會，感覺實在是相當的陌生一點也不知道到底生活品質、項目反應理論是甚麼東西。感謝梁老師在這段過程中，一直耐心的教導，使我在對於問卷的分析及評估上有更多、更深的了解。以及林茂榮老師針對邏輯能力所給我的幫助上使我受益良多。另外還有宏偉學長總是非常孜孜不倦地學習並且還要教我，常覺得他好像是忙碌的小蜜蜂，嗡嗡嗡，飛到西又飛到東，每天總有許多事要忙，好像都忙不完一樣。還有總是很親切的金燕學姊及很熱鬧的弼哥，謝謝你們在這段時間的照顧。

還有泰進、雅琳及玉君，你們要加油喔！雖然項目反應理論剛開始接觸會覺得很難懂，但是以你們的努力及天份一定可以有一個很好的成績。另外，還有中台姊妹之家的姊妹們，桂蘭、詩芳、思妍及曉語，謝謝你們在這段時間中總是擔待我的軟弱，體恤我。以及我最親愛的家人，謝謝你們總是尊重我的決定。

箴言九章 10 節「信靠耶和華是智慧的開端」，一直都覺得自己不是一個聰明的人，但是我總是會跟主禱告說：「是你帶我來的，我所需的智慧，你就一定不會讓我缺乏」。能夠完成論文實在不是我能，乃是神作我一切的智慧。

摘要

目的：

面對傳統心理計量理論不敷所需，愈來愈多人應用項目反應理論（item response theory, IRT）來分析問卷。本研究使用 IRT 中的 Rasch 系列模式分析世界衛生組織生活品質問卷簡明版（WHOQOL-BREF）測量台灣社區老人的心理計量特質。

材料與方法：

研究對象為 2001 年在台灣台中縣新社鄉的 1200 個社區老人。平均年齡為 73.4 歲，介於 65 歲至 103 歲，男性佔 59%。由訪員至受訪者家中以面談的方式取得基本資料及 WHOQOL-BREF 生活品質等資料。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：以適合度指標（infit statistic）來檢驗各範疇單一向度的假設。以題目難度的測定值來檢驗題目難度的範圍，重疊性及順序。並且，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。

結果：

WHOQOL-BREF 各範疇的題目均符合單一向度的假設。老人能力在生理、心理、社會及環境範疇分別為-6.7~7.76、-5.78~8.79、-9.34~10.98 及-5.25~5.41，而題目難度的分佈分別為-1.1~0.48、-1.19~1.29、-0.77~1.17 及-1.08~0.96，然而，在環境範疇的題目難度有較多重覆性的題目。在題目順序方面，與國健局的結果做題目順序的比較，發現在生理範疇及心理範疇有少數題目的順序不同，然而，在社會及環境範疇的題目順序則是相同的。在 DIF 的檢驗上，除了心理範疇的”接受自己的外表”在年齡有明顯的 DIF，以及社會範疇的”性生活”及”朋友支持”在性別及年齡皆有

顯著 DIF 存在外，其他的範疇都沒有顯著的 DIF 存在。WHOQOL-BREF 在生理、心理、社會及環境範疇中的信度指標分別為 0.76、0.77、0.68 及 0.78。除了社會範疇外，其他的範疇都有良好的信度。

結論：

傳統測量方法從加總的分數來看人的能力，該結果會受題目難度的影響，在題目難度估計上也沒有考慮到誤差因人而異的特質。然而，項目反應理論建構在一完整的理論架構下，同時考慮題目的特質及人的能力，能補足許多傳統測量方法之不足，並有系統的發掘問題。從本分析中更進一步地證實 WHOQOL-BREF 在老人的適用性及未來可以改進的部分。

關鍵字：項目反應理論、Rasch 模式、世界衛生組織生活品質問卷簡明版、社區老人、台灣



Abstract

Objectives:

Item response theory (IRT) offers one of the best alternatives for optimizing scales and performing item analysis. The Rasch family of model which was one of item response theory was used to analyze the psychometric characteristics of brief version of World Health Organization Quality of life (WHOQOL-BREF) in community-dwelling older people in Taiwan.

Methods:

One thousand and two hundreds community-dwelling older people living in Shin-Sher Township of Taichung County in 2001, completed the WHOQOL-BREF at their residences either by themselves or with the assistance of an interviewer. The age of these 1200 subjects ranged from 65 to 103 years with an average of 73.4 year, and 59% of the subjects were males. Partial credit model, one of the Rasch family of model, was used to analyze the psychometric characteristics of domains and items of the WHOQOL-BREF. Infit statistics were applied to examine the unidimensionality of each domain; Item calibration of logit unit was to examine the range of item difficulty, redundancy, and hierarchy; and differential item functioning analysis was to examine whether item properties were the same between males and females older people.

Results:

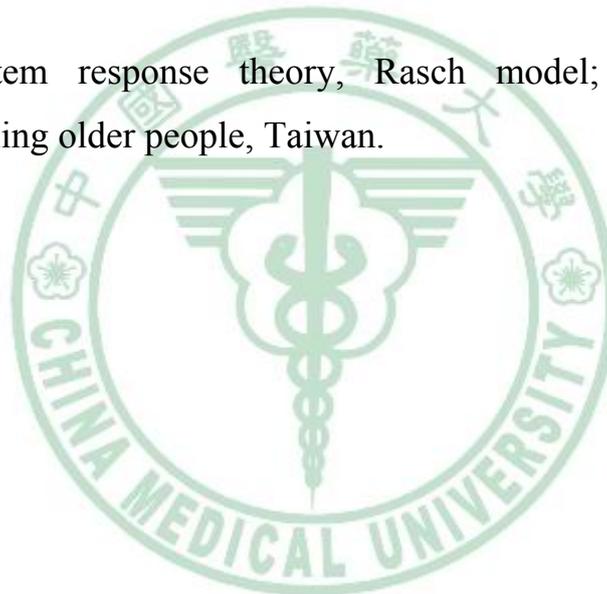
All of the items within each domain satisfied the assumption of unidimensionality. The range of mean of item difficulty (range) in physical, psychological, social, environmental domain were -1.1~0.48 (1.58), -1.19~1.29 (2.48), -0.77~1.17 (1.94), and -1.08~0.96 (2.04), respectively; the corresponding range (range) of threshold difficulty were -3.86~5.3 (9.16), -6.74~6.78 (13.52), -8.39~9.66 (18.05), and -3.72~4.75 (8.47), respectively; the corresponding range of ability were -6.7~7.76 (14.46), -5.78~8.79 (14.57), -9.34~10.98 (20.32), -5.25~5.41 (10.66), respectively. More item redundancies occurred in environmental domain than those in other domains. Comparison of item hierarchy with the results from the NHRI health survey, found that the two studies had different hierarchy in physical and psychological domains, and a similar hierarchies in social and environmental domains. While “body image” of psychological domain had a significant DIF in age group, “sexual activity” and “friends’ support” of social domain had a

significant DIF among gender and age group, there was no apparent DIF in other domains. The reliability indices in physical, psychological, social and environmental domains were 0.76, 0.77, 0.68, and 0.78, respectively.

Conclusion:

The summated score method is used to measuring personal ability in classical test theory (CTT) analysis and it does not account for the variability of item properties. However, IRT is based on a comprehensive structure of psychometrical theory and provides a foundation for scaling personal ability and item difficulty by using the responses of assessment items. IRT can overcome a lot of limitations of CTT. While our study confirms the suitability of WHOQOL-BREF in older people, it also highlights that there is still a room of improvement for the WHOQOL-BREF in the older people.

Key Words: Item response theory, Rasch model; WHOQOL-BREF, community-dwelling older people, Taiwan.



目 錄

中文摘要.....	I
英文摘要.....	III
目 錄.....	V
表 目 錄 (Table content).....	VII
圖 目 錄 (Figure content)	VIII
附 錄.....	74
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 老人的特性.....	7
第二節 老人生活品質研究.....	7
第三節 心理計量特質的測量 — 由古典到項目反應理論.....	9
第四節 項目反應理論.....	11
第五節 Rasch 模式的家族.....	14
第六節 Rasch 模式分析.....	20
第七節 心理計量特質.....	24
第八節 現代測量理論在生活品質的應用.....	25
第三章 研究方法.....	28
第一節 研究設計.....	28

第二節	研究對象.....	29
第三節	研究工具的擬定.....	29
第四節	資料收集過程.....	30
第五節	資料統計與分析.....	31
第四章	研究結果.....	35
第一節	基本人口學變項之敘述統計.....	35
第二節	問卷心理計量特質分析.....	37
第五章	討論.....	63
第六章	結論與建議.....	68
第一節	結論.....	68
第二節	研究限制.....	69
第三節	應用與建議.....	70
參考文獻	71



Table contents

Table 1. Demographic characteristic of community-dwelling older people.....	36
Table 2. Item calibrations, misfit statistics and slope estimates of the WHOQOL-BREF domains in community elderly and control group.....	39
Table 3. The range of item threshold and ability across four domains of WHOQOL-BREF in community older people.....	40
Table 4. Items with redundancy in each domain of WHOQOL-BREF...	42
Table 5. Comparison of item hierarchy between two studies.....	42
Table 6. Targeting and Separation indices of WHOQOL-BREF in community older people.....	43
Table 7. Item calibrations of the WHOQOL-BREF domains in male and female.....	45
Table 8. Item calibrations of the WHOQOL-BREF domains in young and old older people.....	49
Table 9. Item information for ranges of the latent trait.....	55
Table 10. Item thresholds for the WHOQOL-BREF domains.....	60
Table 11. Discriminant validity analysis: means of domain or facet score by characteristic in classical test theory and item response theory.....	65

Figure contents

Figure 1. The response curve.....	15
Figure 2. Jumper stronger than fence clears. Jumper weaker than fence tumbles.	15
Figure 3. Category probability curve for a rating scale item with three thresholds	19
Figure 4. A mean of item difficulty for the four domains of WHOQOL-BREF separated in community elderly.....	41
Figure 5. DIF comparison on male and female in physical domain of WHOQOL-BREF.....	46
Figure 6. DIF comparison on male and female in psychological domain of WHOQOL-BREF.....	46
Figure 7. DIF comparison on male and female in social domain of WHOQOL-BREF.....	47
Figure 8. DIF comparison on male and female in environmental domain of WHOQOL-BREF.....	47
Figure 9. DIF comparison on young and old older people in physical domain of WHOQOL-BREF.....	50
Figure 10. DIF comparison on young and old older people in psychological domain of WHOQOL-BREF.....	50
Figure 11. DIF comparison on young and old older people in social domain of WHOQOL-BREF.....	51
Figure 12. DIF comparison on young and old older people in environmental domain of WHOQOL-BREF.....	51
Figure 13. DIF comparison on male and female in social domains of WHOQOL-BREF between young and old older people.....	53
Figure 14. DIF comparison on young and old older people in social domains of WHOQOL-BREF between male and female.....	53

Figure 15. Person-threshold map for physical domain of WHOQOL-BREF.....	61
Figure 16. Person-threshold map for psychological domain of WHOQOL-BREF.....	61
Figure 17. Person-threshold map for social domain of WHOQOL-BREF.....	62
Figure 18. Person-threshold map map for environmental domain of WHOQOL-BREF.....	62



第一章 緒論

第一節 研究背景與研究動機

生活品質的議題在近年來被廣泛地利用在健康結果測量上，各類型的生活品質問卷也不斷的被發展、修正及驗證，以期能提高測量的品質^{1,2}。隨著問卷的發展及需求，在測量的理論與分析的技術上也一直在尋求突破，傳統的測量理論中許多的限制，隨著現代測量理論的發展，一直在改進與突破，而現代測量理論中最具代表性的應推項目反應理論 (item response theory, IRT) 的發展^{3,4}，此理論最早在教育及心理測量領域中造成了很大的衝擊並蓬勃的發展，目前已居領導地位，許多著名的測驗如托福測驗、加拿大的護理執照考試、IQ 測驗等其測驗方式均根據 IRT 來發展，近年來，此理論又跨足到醫學結果測量的領域^{5,6}，IRT 不但能幫助研究者發展更好的工具來測量所感興趣之特質⁷、並可以幫助問卷使用者了解所欲使用之問卷特質，增加使用者對已有資料之了解，並能有效地幫助使用者選擇適合的問卷。它較傳統測量理論更能結構性地提供許多有用的訊息，例如每份問卷的心理計量特質，甚至了解到如何區辨每個題目的好壞⁸。

老年人口的增加在開發中及已開發國家中為一個共同的現象。在衰老的過程中有一些症狀是很難避免的，如身體及心理的老化，而且每一個人老化的程度是不相同的，甚至是相當兩極化的。每一個人都有不同的特質，但是如何在老化的過程中歸納出老人面臨老化時共同的特點及在老化過程中的需求是非常重要的。老人在健康上常會面臨的問題很多，其中常見有多種的慢性病、憂鬱症、失智等。目前有許多學者利用生活品質來測量老人的健康，這些研究指出生活品質可以用來預測老人憂鬱症⁹老人在照護中心的死亡率等。

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。

Hwang 等利用傳統心理計量方法來評估 WHOQOL-BREF 在新社鄉社區老人的適用性¹³，顯示 WHOQOL-BREF 具有良好的信度、效度且適用於評估老人的生活品質。WHOQOL-BREF 問卷中包含很多問卷中所沒有考量到的環境範疇，對於老人在環境範疇上較敏感的需求是很重要的。同時，也顯示不論是請老人自填或是由訪員訪問所花費的時間差不多（花費平均時間分別為 11 分及 15 分，兩者間並沒有統計上顯著的不同），並不會造成老人在答題上的負擔，表示 WHOQOL-BREF 是適用於測量老人生活品質的。然而以上的方法大都以傳統分析為主，在傳統分析中常利用範疇的得分來看人的特質，但項目反應理論(IRT)乃是更細緻地探討每一個題目的特質且從題目的特質來看人的能力。另外，傳統分

析中題目特質與人的能力之間是無法相比較的，但 IRT 利用一個數學模式將人的能力及題目特質放在同一個尺度上使其可以相互比較，並能更深入有系統地發掘問題。

項目反應理論 (item response theory, IRT) 一般可作為現代測量理論的代表，其基礎架構主要是根據心理計量理論利用數學模式將人的能力及題目的難度放在同一個尺度上，使得人的能力與題目的難度間可以加以比較評估。此理論發展大致分為兩個系列的模式，一為 Rasch 系列模式，另一為參數系列模式。項目反應理論目前已廣泛使用在教育界、心理界、商業界，且在近十幾年來在健康結果測量上的應用亦是大量增加。在 2001 年著名的雜誌 *Medical Care* 對 IRT 專闢一系列的討論^{1, 2, 14}。實際應用的文獻亦增加中，因應在 2001 年在美國伊利諾州芝加哥所舉辦的客觀測量的國際研討會 (International Conference on Objective Measurement, ICOM) 中，117 篇被接受的論文中，有 104 篇使用 Rasch 測量模式，有 8 篇利用 2-或 3-參數的項目反應理論模式，只有 5 篇是利用傳統測驗理論 (classic test theory)。Rasch 模式亦廣泛地被利用在健康照護的研究中，在 2004 年 1 月 *Medical Care* 針對 IRT 中的 Rasch 模式在健康照護的研究中有一系列的探討^{5, 15-20}。

使用傳統的方法來驗證問卷是簡單且易懂，但是其並沒有一個完整的理論架構作支持，利用 IRT 雖不像傳統測量理論來得易懂，但是其背後有完善的理論架構，針對每一題提供的訊息更多、且更符合理論的假設。目前已有許多研究利用 IRT 來驗證生活品質的問卷並較傳統方法提供更多題目的訊息。在 Ducan 等人利用 IRT 中的 Rasch 模式來驗證一份新的中風衝擊量表 (stroke-impact scale) 的研究中也提到利用 IRT 分析所提供的訊息不僅可以用來驗證問卷的信、效度，也利用所提供的訊息來評估受測者的特質⁸。Nijsten 等進一步比較傳統測量理論與項目反應理論，

在評估如何縮減牛皮癬衝擊問卷(impact of Psoriasis Questionnaire)²¹，結果顯示在縮減問卷方面利用 Rasch 分析所提供的訊息較傳統測量理論不僅題目更少並且在信度，單一向度的假設上有更好的心理計量特質。

因此，本研究擬針對 Hwang(2003)所使用 WHOQOL-BREF 來測量新社鄉老人的生活品質，以 IRT 理論中之 Rasch 模式重新探討 WHOQOL-BREF 應用在社區老人之心理計量特質及 WHOQOL-BREF 應用在老人的適用性。

第二節 研究的重要性

- 1.可使研究者更瞭解 WHOQOL-BREF 問卷在範疇及題目的心理計量特質，方便未來研究使用時之選擇參考。
- 2.可利用問卷的特質來更瞭解 65 歲以上社區老人的生活品質，檢驗老人對 WHOQOL-BREF 問卷的適合性作為發展更適宜問卷的參考。

第三節 研究目的

本研究目的為以 Rasch 模式分析 WHOQOL-BREF 問卷在測量老人生活品質的心理計量特質。應用 Rasch 分析來評估下列各項心理計量特質：

- (1) 單一向度 (unidimensionality)
- (2) 題目的難度 (item difficulty)
- (3) 題目的適中性 (targeting)
- (4) 區別性指標 (separation index)
- (5) 試題差別功能(differential item functioning)
- (6) 訊息功能 (information function)
- (7) 鑑別力 (discrimination)
- (8) 難度與能力整合 (integrated analysis)。

第四節 研究問題與研究假設

研究問題

- (1) 從現代測量理論觀點，世界衛生組織生活品質問卷簡明版 (WHOQOL-BREF) 是否有良好的心理計量特質？
- (2) 世界衛生組織生活品質問卷簡明版 (WHOQOL-BREF) 是否適用於社區老人生活品質的測量？

研究假設

- (1) WHOQOL-BREF 各範疇具單一向度的概念。
- (2) WHOQOL-BREF 各範疇題目難度均勻分佈。
- (3) WHOQOL-BREF 各範疇題目難度具適中性。
- (4) WHOQOL-BREF 各範疇在不同族群沒有 DIF 存在。
- (5) 訊息函數 (information function) 能準確提供各題目訊息量。
- (6) 心理計量特質能反映出測量品質的好壞。
- (7) IRT 分析可以更準確且全面性來檢驗問卷的心理計量特質。
- (8) WHOQOL-BREF 適用於社區老人生活品質的測量。

第五節 名詞界定

古典測量理論：又稱為傳統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（贊同）機率為人的能力與一個或以上的題目參數的聯合函數。反應機率的表現是以題目特徵曲線作為潛在特質的函數。

Rasch 分析：為項目反應理論中的一種。

心理計量特質：統計上評估問卷適合度的方法，如信度、效度及內部一致性。

效度：研究工具能測到研究者想要測量概念之程度。

建構效度：評估測量工具能否符合研究者所感興趣理論的假設。

信度：測量工具中顯出變項誤差的一致性。

區別性指標 (Separation index)：評估測量工具能否區別出特質間的差異。

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。

第二章 文獻查證

第一節 老人的特性

由於公共衛生及醫療介入，平均餘命的延長使老年人口增加，台灣在 1997 年時 65 歲以上老年人口已超過總人口的 7%，根據台灣行政院衛生署公佈民國 93 年 65 歲以上的老年人口約占總人口的 9.48%²²。老年人口的增加使老人的議題愈來愈受到關注¹³。因老化所造成身體上功能的缺陷及慢性疾病常常影響老人的健康，這些退化症狀及慢性疾病成了老年人口中常見的問題。疾病或治療深深地會影響老人的健康，在現今重視病患自我感受（patient reported outcome, PRO）的醫療模式中，常利用問卷來評估病患全面性的健康。

第二節 老人生活品質研究

當應用一般性問卷在老人時必須考慮測量方面的幾個問題¹³。首先，老人填寫問卷相較於年輕人負擔增加許多，因為老人身體的虛弱、文盲或認知缺陷的比例較高。第二，生活品質測量中的題目及範疇需要考慮到老人認為重要的部分，例如：健康服務的可近性及環境的安全²³。第三，假如直接使用而沒有加以修改問卷，生活品質測量的地板值（floor value）在老人可能是較常見的²⁴，例如：許多老人比較沒有機會從事休閒活動或四處行動的能力比較差。地板值會降低區辨有病及健康人之間不同的能力及降低偵測不同時間或健康介入後生活品質改變的能力²⁵。

生活品質是一個廣泛且多範疇的概念，根據世界衛生組織所定義生活品質是「生活品質是指個人在所生活的文化價值體系中，對於自己的目標、期望、標準、關心等方面的感受程度，其中包括一個人在生理健康、心理狀態、獨立程度、社會關係、個人信念以及環境六大方面」²⁶，但世界衛生組織在實地研究之後提出生活品質至少應包含生理，心理及

社會等範疇²⁷。隨著平均餘命的增加，人們晚年的健康愈來愈受到重視。疾病或治療會影響老人的健康，生活品質的測量結果可以協助建構老人經歷疾病或治療後全面性的健康剖析，藉此可以幫助醫療政策的制定以改善老人的生活品質²⁸。生活品質問卷常見的如由世界衛生組織所發展的 WHOQOL-BREF 問卷及由 QualityMetric 公司所發展的 Short Form-36。

在各地老年人口的精神疾病中，以憂鬱症的盛行率最高，其次為認知缺陷。因此，有研究利用 WHOQOL-BREF 來測量憂鬱症老人的生活品質並驗證 WHOQOL-BREF 問卷，且探討生活品質與臨床及社會人口學因子的相關性。結果發現憂鬱症愈嚴重生活品質愈差，並且自覺生理症狀（self-report physical symptom）愈多者生活品質亦愈差。但是在共病症（diagnostic co-morbidity）與生活品質間並沒有顯著的關係存在⁹。

在台灣亦有學者利用 SF-36 建立在都市，鄉村及離島地區老人生活品質的參考基準且針對三個區域加以比較，並且探討人口學變項與生活品質之間的關係。研究結果發現在大多數的範疇中，都市老人的生活品質較鄉村及離島地區老人顯著的比較好。離島地區老人在活力及心理健康上是最好，反之，鄉村老人的生活品質是最差的，尤其是鄉村老人中的女性。在一些其他的特徵上則顯示慢性疾病愈多者生活品質愈差，且需要看護者及最近三個月內有住院的老人生活品質都較差²⁹。

第三節 心理計量特質的測量 — 由古典到項目反應理論³⁰

一般的科學研究中常常需要測量一些變項，但並不是每一個特質的測量都像身高、體重等等可以利用工具明確的量出來。常常研究者感興趣的特質是無法直接測量得到，例如：英文能力、憂鬱的程度及生活品質等等。當面臨這些無法直接測量得到的特質時，傳統上，研究者首先發展一組問題來評估一個概念。當由一群受測者完成問卷的填寫後，再將這些結果加總成一個總分的數值來代表該受測者在某特質上的表現，這種測量方式較接近古典測量理論。舉例來說，假設一個研究者發展評估一份飲食疾病的量表，量表得分愈高表示有飲食疾病的機會愈高，量表得分愈低表示愈沒有機會有飲食疾病。題目的選項通常為5分法：1（非常不同意），2（不同意），3（普通），4（同意），5（非常同意）。假想問卷中的幾個題目如下：

	非常不同意	不同意	普通	同意	非常同意
	SD	D	N	A	SA
1. 我規律地嘔吐來控制我的體重	1	2	3	4	5
2. 我會計算食物的熱量	1	2	3	4	5
3. 我用運動來燃燒脂肪	1	2	3	4	5

假如有一個人在等級評價量表(rating scale response)的反應分別為2, 4, 5。利用傳統的計分方式這一個人在飲食疾病的量表中可以得到11分。之後，在隨後的統計分析中這個11分會被視為一個“測量值”。另一個人的反應為5, 5, 1，他的分數也是11分。直接將分數等級加總有兩個假設如下：(1) 每一題目的貢獻都是相同的(constant information)，並且(2) 每一個題目選項的測量都具有相同的間距(equal interval scale)上，就是假設選項間具有等距的特質。以下針對這兩個假設所面臨的問題進行說明。

根據第一個假設表示所有的題目在評估這個概念的重要性上是相同的。以上述假想中三個不同性質的題目來討論，可發現假設此三題對於

飲食疾病量表有相同的貢獻似乎是不明智的。例如，在第一題回答”非常同意”是較第三題回答”非常同意”更可能有飲食疾病上的問題。另外，有爭議的地方則是第二題所涉及飲食議題部份是較少的。因此，當在已知的概念下題目代表不同的程度時，資料不應該用直接加總後的總分來分析，因為總分表示所有題目所貢獻的價值，但每題所貢獻價值儘管分數相同，但意義並不同。

當然在第一題回答”非常同意”對於整體飲食疾病分數的貢獻是大於第二題回答”非常同意”的。

第二個假設是題目的選項間間隔都是相同的 (equal interval scale)。將等級評價量表(rating scale response)相加表示每一個選項的距離是相同的。我們以這虛構飲食疾病問卷的第一題為例。在心理學上，對於受測者這一題”不同意”及”非常不同意”選項的間距是較其他選項的間距接近的。同樣地，認為”同意”及”非常同意”可能是較接近的。因此，對於這一題在心理學上的每一個選項的距離可能如下。

1. 我規律地嘔吐來控制我的體重 SD D N A SA

這個間隔說明如何從回答”不同意(D)”越過”同意(A)”的反應是相當大的，在同意尺度(A 或 SA)或不同意尺度(D 或 SD)上難以捉摸的些微的變化並不容易區分。例如，對於一個並未規律地以嘔吐來控制體重的受測者而言(第一題)，他頗有可能選擇”不同意”或”非常不同意”中任一個，因為這兩個選項間心理學上的距離是相當近的，依此類推，此受測者是較難以回答”同意”或”非常同意”中的任一個。在這個例子中，回答”同意”與”不同意”之心理學上的距離是相當大的。

此外，對於每一個題目在每個序位選項間距離的重要性可能不同。例如，對於一些題目在”同意”及”非常同意”可能有短的心理學上躍過的距離，但其他的就不一定。將題目各選項相加而得到的總分會忽略這個現

象，這所引起的另一個問題為當原始的總分被視為這個概念的測量值時會影響測量的精確度。因此，我們這 3 題飲食疾病量表中較精確的概念如以下的推論：

1. 我規律地嘔吐來控制我的體重 SD D N A SA
2. 我會計算食物的熱量 SD DN A SA
3. 我用運動來燃燒脂肪 SD DN A SA

在傳統的方法中將這兩個假設強加在問卷的計分上，著實會影響測量的精確性。因應以上的問題心理計量學者提出現代測量理論以解決傳統方法上的限制。其中最具有代表性的應推項目反應理論 (item responses theory)，以下將針對項目反應理論進行介紹。

第四節 項目反應理論³¹

項目反應理論 (item response theory, IRT) 因應古典測量理論的一些限制而發展。在目前的應用上，IRT 分為兩個發展的路線，一為 Rasch 系列，一為參數系列。IRT 模式是利用非線性函數整合受測者在一個特質中潛在能力 (latent trait) 與題目特質之間的關係³²。在測量的過程中所觀察到題目的反應與潛在特質(或能力)之間的關係，可以題目特徵曲線 (item characteristic curve)³³ 來表現。在 IRT 模式中，假設題目反應具有單一向度 (unidimensionality) 與局部獨立 (local independence) 的特性。單一向度表示題目間評估單一潛在的概念。局部獨立性表示在固定的能力下，題目間彼此不相關。

IRT 模式有不同的模式，其主要的區別是在潛在能力與題目反應機率間關係的函數類型不同。最簡單的 IRT 模式為 Rasch 模式³⁴，此模式原則上只考慮困難度參數，故又稱為一個參數的邏輯斯函數 (logit function)。Rasch 模式允許題目間有不同的難度，但是卻假設所有的題目

都有相同的鑑別力。假如 Rasch 模式符合 IRT 基本假設（單一向度與局部獨立性），其次觀察到的題目得分，可利用受測者的能力（theta）與題目難度的函數式來描述之。舉例來說，對於二項式的題目（dichotomous item），答對的機率利用數學模式形成邏輯斯的頻度曲線（logistic ogive）：

$$P_{ni} = \frac{1}{1 + \exp(B_n - D_i)}$$

P_{ni} 是隨機選第 n 位受測者回答第 i 題答對的機會。 B_n 表示受測者的潛在特質， D_i 為題目的難度。難度參數表示在特定的能力下，受測者答對（或贊同）這一題的機率為 0.5 時所對應在量尺上的值，此定義是根據題目為只有兩個可能的反應之二項式情況，在最初建構 IRT 模式時，就是以二項式題目作為一個基礎。二參數模式（two parameter IRT model）是在單參數模式之後發展的，不僅估計題目的難度也估計題目的鑑別力，二參數模式的反應機率為：

$$P_{ni} = \frac{1}{1 + \exp[-a_i(B_n - D_i)]}$$

a_i 為第 i 題的鑑別力參數。鑑別力參數相似於在傳統測量理論中題目與總分的相關性（item-total correlation）；這個參數的值愈高表示題目有較好的能力來區分鄰近能力的程度，並且這也可以由潛在能力（X 軸）或 theta 與特定回答的機率（Y 軸）間較陡的斜率來證明。而三參數模式是加入偽猜測（pseudo-guessing, c ）參數來調整事實上個體可能因為運氣而使得分數高於期望值的情形。

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³⁰。

訊息函數 (information function)

在 IRT 中訊息函數(information matrix)是一個重要的特點來評估測量的精確度，IRT 所提供的訊息函數 (information function) 在建構一份問卷或題目庫及比較問卷上是很有用的指標⁴。IRT 的測量模式的訊息 (information) 是一個基本的特徵，不論是二分類選項 (dichotomous response) 或多個題目回答選項 (polytomous item-response)，它可被表現在二分類選項中以題目反應曲線 (item response curve) 或是在多個題目回答選項以類別反應曲線 (category response curve)，或是轉換成題目訊息曲線 (item information curve) 來表現受測者能力與其對題目反應之間的關係。以二分類選項的題目為例，可以寫出題

$$I(\theta) = \frac{P_i^*(\theta)^2}{P_i(\theta) * (1 - P_i(\theta))}$$

目訊息曲線如下：

$P_i(\theta)$ 為答對或贊同一個題目的條件機率，在一些二分類的 IRT 模式中可由參數估計值得到，且 $P_i^*(\theta)$ 這個術語代表在特定能力下題目反應曲線第一次微分的值。為了方便計算可以將公式改寫，在三參數模式中公式可以幫助研究者計算條件式題目訊息曲線 (conditional item information curve)^{3, 35}。

$$I(\theta) = \left[a_i^2 \frac{1 - P_i(\theta)}{P_i(\theta)} \right] * \left[\frac{(P_i(\theta) - c_i)^2}{(1 - c_i)^2} \right]$$

若偽猜測參數 ($c=0$) 如雙參數模式，則公式如下為 $I(\theta) = \alpha_i^2 P_i(\theta)(1 - P_i(\theta))$ ，並且在單參數模式中題目訊息函數為 $I(\theta) = P_i(\theta) * (1 - P_i(\theta))$ 。

訊息函數可以讓研究者瞭解每一個題目在連續的能力特質尺度的每一點所能提供的訊息，當訊息函數的數值愈高時表示題目在對於此能力下的人所能提供的訊息愈高，即能較準確的估計在此能力下的人。在本

分析中藉由將每一個題目的訊息函數相加，來評估哪些題目所能提估的訊息最高，即評估哪些題目在評估老人的生活品質時，能夠估計的較準確。另外，在每個範疇中將在每一個能力間隔下的訊息函數相加，來探討每個範疇在哪一個能力區間下能夠提供較高的訊息，亦即提供較準確的估計^{3,35}。

第五節 Rasch 模式的家族

Rasch 模式的發展，Georg Rasch 佔了重要的地位。Georg Rasch (1901-1980) 是一個丹麥的數學家，他同時也是個統計學家及心理計量學家，他最著名的成就是發展 Rasch 測量模式。在 1919 年，Rasch 開始在哥本哈根大學讀數學系，並於 1925 年完成碩士學歷，在 1930 年獲得博士學位。畢業後因為無法以數學家的身份找到工作，故在同年轉向去做統計學顧問³⁶。

Rasch 最常見的貢獻是在心理計量領域上，一開始他以 Poisson 分佈來建立學生考試時分數偏差的模式。他將模式視為一個 multiplicative Poisson model，之後他發展用於二項式題目 (dichotomous items) 的 Rasch 模式，並將此模式用在丹麥軍人的智力及成就測驗的回答資料探討。目前這個模式已經廣泛地用於評估教育及教育心理學的測量，尤其是在成就及認知的評估³⁶。

Rasch 模式公式如下：

$$P_{ni} = \frac{e^{(B_n - D_i)}}{1 + e^{(B_n - D_i)}}, i = 1, 2, \dots, n$$

從以上的數學式可推導出圖一的反應曲線。利用數學模式將人的能力與題目的難度放在同一個尺度上，使其間的關係可以相比較。從圖一中我們可以知道當人的能力與題目難度相同 ($B_n = D_i$) 時，他答對這一題的機率為 0.5。當人的能力遠比題目難度低時，答對的機率就小。當能力

遠比題目的難度高時，答對的機會就高。

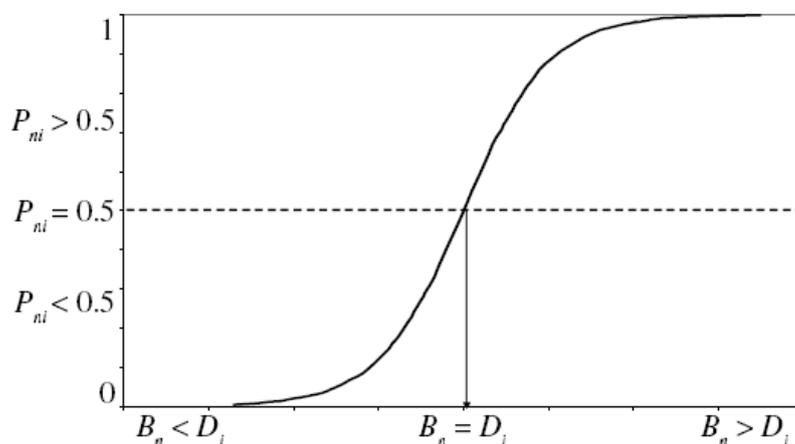


Figure 1. The response curve³⁷

Wright 等³⁷ 利用圖二簡單的描繪出 Rasch 模式背後的心理計量理論。圖二的跨欄表示題目難度，如果人的能力與題目的難度相同 ($B_n - D_i = 0$)，那麼該生答對這一題的機會為 0.5，即該生有一半的機會答對此題目。圖中的跨欄高手其能力遠比此題目的難度高 ($B_n - D_i > 0$)，那麼越過此跨欄的機會就增加，即跳過這個跨欄是容易的。但是若是人的能力遠比題目的難度低 ($B_n - D_i < 0$)，就難以越過此跨欄。

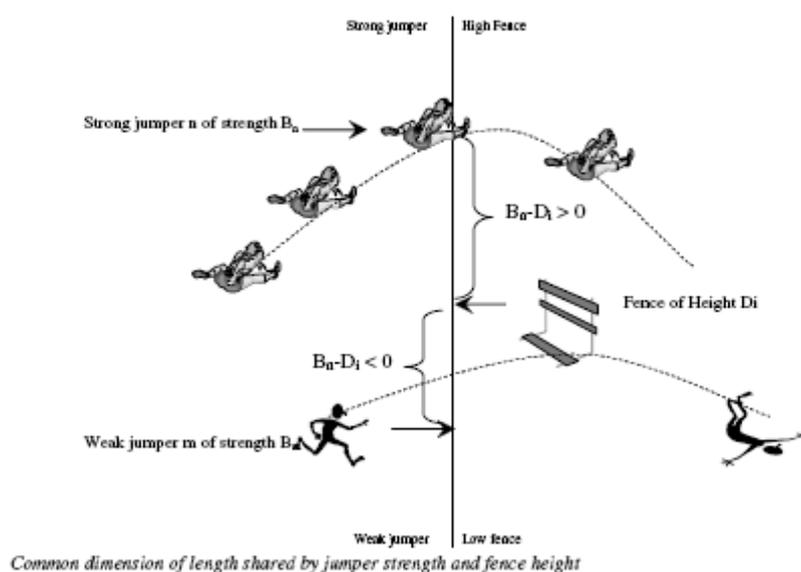


Figure 2 Jumper stronger than fence clears. Jumper weaker than fence tumbles.³⁷

二項式模式 (dichotomous model)

二項式模式 (dichotomous model) 為 Rasch 模式家族中最簡單的成員。據此我們可在已知人的能力及題目的難度之下，預測二項式結果 (是/否) 的條件機率。假如回答”是”編碼為”1”，回答”否”編碼為”0”，該模式是以答”是”的機率作為在第 n 個人的能力與第 i 個題目難度差異大小的函數。此基礎為假設題目難度及人的能力這兩個測量值間的差異是會影響人在回答特定題目時答”是”的機率，這個基本的邏輯是容易理解的，在較簡單的題目中所有的人有較高的機會答”是”，而在較難的題目中則有較低的機會答”是”。

當一群合適的樣本填寫一份問卷時，測量的啟動點是開始於計算每一個人 (答”是”的題目個數除以總題目個數) 及每一個題目 (回答”是”的人的個數除以所有人的個數) 正確的百分比。這些原始分數相加的序位得分，在估計人的能力及題目難度測量值時是必需且充分的^{38,39}。

在估計 B_n (第 n 個人的能力估計值) 的第一個步驟是轉換原始得分的百分比成答”是”的勝算，勝算的計算為以答”是”的百分比除上答”否”的百分比。舉例來說，原始分數有 40% 回答”是” (p) 除以回答”否” ($1-p$) 的百分比，獲得 40/60 的比值，此比值我們稱之為勝算。之後，取這個勝算的自然對數 ($\ln 40/60 = -0.4$) 就是人能力的估計值。估計題目的難度 D_i (第 i 個題目的難度) 也是相同的計算過程，人們在回答某一題目時答”是”百分比除以答”否”百分比並且將這比值取自然對數就是題目難度的估計值。

這些題目難度 (D_i) 與人的能力 (B_n) 的估計值被放在一個 logit 的量尺上。平均難度的 logit 設為在 0，正的 logit 表示高於平均機率，而負的 logit 表示低於平均機率。Rasch 模式的計算經常開始於忽略人的估計值，先計算題目的估計值，之後使用第一輪得到的題目估計值來計算人

的能力，以獲得第一輪人的估計值。這些估計值彼此間反覆的比對，以獲得一組穩定且內部一致性高的題目難度與受測者能力的參數估計值，所以 $B_n - D_i$ 的值能表現出受測者答”是”的 Rasch 機率值。這個重覆比對的過程稱為收斂(converge)。當我們已經估計出受測者能力及題目難度，受測者回答”是”的機會可利用以下數學式表示：

$$P_{ni}(x=1) = f(B_n - D_i) \quad (1)$$

這裡的 P_{ni} 是機率值， x 是一個已知的回答，並且 1 表示回答”是”。因此這個方程式所指的情況為，在一個已知的題目(i)下，第 n 個人回答”是”(x=1)的機率(P_{ni})是人能力(B_n)及題目難度(D_i)的差異的函數。

希臘符號在統計上常用來表示估計值的參數，公式如下：

$$\pi_{ni}(x_{ni}=1) = f(\beta_n - \delta_i) \quad (2)$$

這裡的 π_{ni} (p_i) 代表機率， β (beta) 表示人的能力，並且 δ 代表題目的難度。

已知 B_n 及 D_i 之後，我們可以方程式(1)得到下列的關係式：

$$P_{ni}(x_{ni}=1|B_n, D_i) = \frac{e^{(B_n - D_i)}}{1 + e^{(B_n - D_i)}} \quad (3)$$

$P_{ni}(x_{ni}=1|B_n, D_i)$ 是在已知人的能力及題目難度下，第 n 個人回答第 i 題答”是”(x=1)的機率。

Rating Scale Model

Rating scale model 是 dichotomous model 的延伸，當題目有超過 2 個以上的反應選項（如 Likert scale），例如，一個題目有五個反應的選項（0=相當不同意，1=不同意，2=普通，3=同意，4=相當同意）。這模式中有 4 個閾值（threshold），閾值的定義為一個人有 50/50 的機會選擇一個的選項越過另一個選項，每一個題目之閾值(k)對應一難度估計值(F)，例如，第一個閾值設定為選擇選項”1(不同意)”來代替選項”0(非常不同意)”的機率，可用以下公式來估計：

$$P_{ni1}(x=1|B_n, D_i, F_1) = \frac{e^{(B_n - [D_i + F_1])}}{1 + e^{(B_n - [D_i + F_1])}} \quad (5)$$

P_{ni1} 是第 n 個人在第 i 題越過非常不同意(category 0) 而選擇不同意 (category 1)的機率。在這個公式中， F_1 是第一閾值的難度，在 rating scale model 中，這個閾值難度(F_1)的估計值在整個範疇的題目中只估計一次。閾值難度(F_1) 與題目難度(D_i)相加 (如 $D_i + F_1$) 來表示第 i 題於第一個閾值的難度。假設 $B_n - (D_i + F_1)$ 與 $B_n - D_i - F_1$ 有相同的數值，這樣的方式更能幫助我們對於 Rasch model 的了解。

例如選擇同意(category 2)越過不同意(category 1)設定如下：

$$P_{ni2}(x=2|B_n, D_i, F_2) = \frac{e^{(B_n - D_i - F_2)}}{1 + e^{(B_n - D_i - F_2)}} \quad (6)$$

這裡的 B_n 是人的能力， D_i 是整個題目的難度，以及 F_2 是第 2 個閾值的難度，且這個閾值難度是由所有題目所估計。Rating scale model 的通式是任一個人在任一個題目選擇任一個已知選項的機率，為第 n 人的能力(B_n) 第 i 題的難度(D_i)及第 k 個閾值難度(F_k)的函數，亦即

$$P_{nik} = \frac{e^{(B_n - D_i - F_k)}}{1 + e^{(B_n - D_i - F_k)}} \quad (7)$$

然而，假如將這些機率轉換成勝算，參數分離也可以利用 rating scale model 來證實：

$$\ln\left(\frac{P_{nik}}{1 - P_{nik}}\right) = B_n - D_i - F_k \quad (8)$$

一般在統計上常常利用希臘字母來表示母群體參數的估計，公式表示如下：

$$\ln\left(\frac{\pi_{nik}}{1 - \pi_{nik}}\right) = \beta_n - \delta_i - \tau_k \quad (9)$$

這裡的 β_n 是人的能力， δ_i 是整個題目的難度，以及 τ_k 是第 k 個閾值的難度，且這個閾值難度是由所有題目所估計。

對於一組 rating scale item 的閾值可以用題目機率曲線的交叉點來描述每個反應選項。圖三中的 F_1 表示選擇從 category 0 越過 category 1 的閾值， F_2 表示選擇從 category 1 越過 category 2 的閾值， F_3 表示選擇從 category 2 越過 category 3 的閾值， F_4 表示選擇從 category 3 越過 category 4 的閾值。x 軸表示人的能力 (B_n) 及題目難度 (D_i) 之間的差異。Y 軸表示反應的機率值。

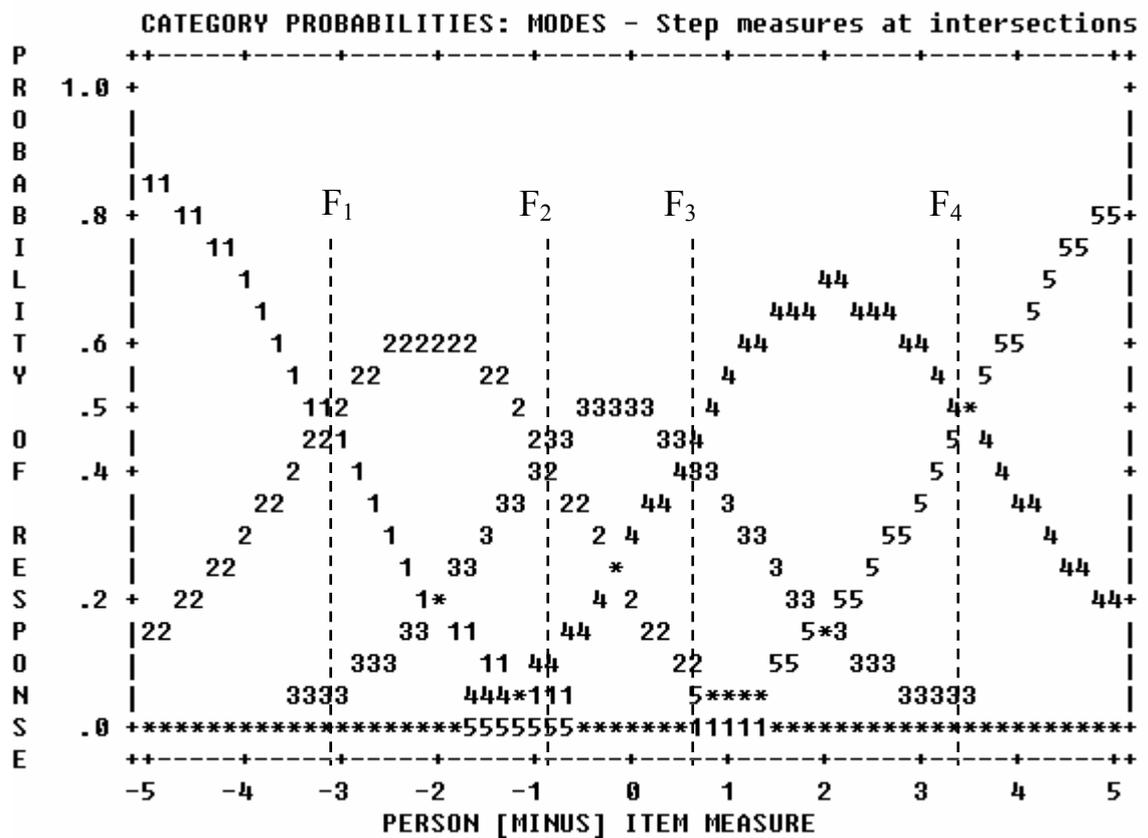


Figure 3. Category probability curve for a rating scale item with three thresholds³⁰

Partial Credit Model

Partial credit model 可以視為 rating scale model 的一種，但其不會強迫在一組題目間閾值估計值都是固定的，閾值估計值在每一個題目都是不相同的。當用 rating scale model 來處理 rating scale 的資料，它允許每一個題目有它本身的閾值估計值。反之，rating scale model 會將一組預值估計值套用於整組題目，partial credit model 對每一個題目(i)會有其自己

的(k)個閾值估計值。

$$\ln\left(\frac{P_{nik}}{1-P_{nik}}\right) = B_n - D_{ik} \quad (10)$$

因此，從方程式 8 以 D_{ik} 來取代 $(D_i + F_k)$ ，表示在 partial credit model 中，每一組閾值估計值對於個別的題目都是唯一的。在本文中是利用 Rasch 系列中的 Partial credit model 來進行分析。

第六節 Rasch 模式分析

(1) 單一向度假設檢定^{8, 30, 40}

單一向度是在評估一組題目所測量的概念是否相同。Rasch model 是利用嚴謹的數學方程式將所有題目難度與人能力之間的關係經由一連續的等距尺度來結合起來，所以不可能所有的題目及所有人都會與模式完美的符合。適合度指標的估計是計算每一個人(B_n)在每一題(D_i)的反應殘差，即真實的反應(x_{ni})與 Rasch 模式所估計的期望值(E_{ni})差多少？

$$y_{ni} = x_{ni} - E_{ni}$$

殘差的分佈常為一個 z 或 t 分佈，當要檢查適合度統計量時常以平均值均方 (mean square fit statistic) 或標準化的適合度指標 (standard fit statistic) 表示。另外，這兩種適合度指標可進一步被分為(a)較強調於非期望的反應與受測者及題目測量間的差異 (outfit statistic)，(b)較強調在非期望的反應在受測者或題目測量附近 (infit statistic)。

1. 未加權的均方適合統計量---Outfit MNSQ

Outfit statistic 是考慮受測者的能力與題目難度關係後，標準化殘差的平方取平均值。這個平均值是沒有加權的，不會被其他的訊息所影響。這個指標可以提供非期望的反應對受測者及題目測量的相對的影響。

$$Z_{ni}^2 = \frac{(x_{ni} - E_{ni})^2}{\sigma_{ni}^2}$$

$$\text{outfit} = \text{average} \left[\frac{(x_{ni} - E_{ni})^2}{\sigma_{ni}^2} \right] = \frac{\sum Z_{ni}^2}{N}$$

X 為觀察值， E 為期望值， σ^2 為期望值的變異數， N 為觀察值的數目。

期望值為 1，Outfit MNSQ 範圍介於 0 至無限大。Outfit MNSQ 值若等於 1，則表示題目具有局部獨立的特性，題目與資料適合。Outfit 是建立在標準化殘差平方和（sum of squared standardized residuals）的基礎所發展的指標，假設標準化殘差為常態分佈（a unit normal distribution），故均方和逼近於卡方分佈（ χ^2 distribution）。

Outfit MNSQ 有高的敏感度用來測量受測者在填達此題時，題目是否會太簡單或太困難。若 Outfit MNSQ 逼近 0，則表示此題目在測量上有過度預期的效果。若大於 1.3，則表示此題目與資料不合適⁴⁰⁻⁴²，亦有許多文章建議以大於 1.4 作為判別標準^{43, 44}。

2. 加權的均方適合統計量---Infit MNSQ

Infit 是 Outfit 經訊息加權後之形式（information-weighted form），此加權過程將減低較偏離主要研究對象能力所在範圍部份所造成的影響， W_{ni} 為經由殘差本身變異數加權的權重，則 Infit 數學公式為：

$$\text{infit} = \frac{\sum (Z_{ni}^2 * W_{ni})}{\sum W_{ni}}$$

若 Infit MNSQ 值逼近於 1，則表示題目具有局部獨立的特性，亦是題目與資料適合。Infit MNSQ 若接近 0，則表示此題目在測量上有過度預期的效果。因為 Rasch 模式為一機率理論，所以容許與期望值之間差異的存在，在大部分的研究中大都容許 30% 的差異性存在，所以若是 infit 或是 outfit MNSQ 大於 1.3，則表示此題目與範

疇不合適⁴⁰⁻⁴²，亦有許多文章中建議以 1.4 作為判別依據^{43,44}。

一般而言，使用 Rasch 分析可以得到四種適合度指標，常用的為平均值均方（mean square infit statistic），因 infit statistic 較 outfit statistic 敏感，且不易受極端值的影響。較少是用 misfit t statistic 可能是因 t 分佈是除以標準誤，當樣本數愈大時，標準誤就愈小，t statistic 很有可能會大於 2⁴⁵。

(2) 題目鑑別力(item discrimination)

Rasch 模式中題目的鑑別力在每一個題目間都是固定的。這樣的特性支持相加性及概念的穩定性。在 Rasch 模式的斜率是所有題目的平均鑑別力。因為區辨參數是非線性的，所以並不是個別斜率的平均值。當 Rasch 模式是以邏輯斯（logits）為公式，在數學上將平均的斜率設為 1。在實際估計題目的應用上，首先計算及固定（anchoring）Rasch 估計值之後再計算鑑別力。這是一個事後分析，針對每一個題目估計鑑別參數(a_i)。此估計模式如下：

$$\log\left(\frac{P_{nik}}{1-P_{nik}}\right) = a_i(B_n - D_i - F_k)$$

a_i 可能的數值範圍為 $-\infty$ 到 $+\infty$ ，數值愈趨近 $+\infty$ 表示鑑別力愈佳。Rasch 估計值會強迫平均的題目鑑別力接近 1。結果鑑別力的估計值為 1 與 Rasch 模式的期望一致。數值若大於 1 表示鑑別力佳(over-discrimination)，並且數值小於 1 表示鑑別力差(under-discrimination)。鑑別力佳被認為在原始分數及 IRT 分析的分數是有益的。高的鑑別力經常與低的 MNSQ 數值相一致，低的鑑別力則與高的 MNSQ 一致⁴⁶。

(3) 受測者區別性分析

Rasch 測量模式提供指標來幫助研究者了解在一個連續的尺度上題

目是否夠分散並且人的能力是否區分得開。Rasch 能提供用以評估信度的指標 person reliability index。

Person reliability index 是估計測量相同概念的其他題目間，人能力相對位置的重覆性。Person reliability 會受能力的估計值誤差的影響及 targeted item 的個數所影響，即當題目估計值的誤差愈小並且題目愈多時，person reliability 就愈高，此概念與傳統方法中用以評估信度的 Cronbach's alpha 相似。這是觀察的反應變異重複的百分比。

$$R_p = \frac{SA_p^2}{SD_p^2} = \frac{SD_p^2 - SE_p^2}{SD_p^2}$$

分母 (SD_p^2) 代表人在感興趣的測量上的所有變動。分子 (SA_p^2) 代表所有的變動中重複的部份，用 Rasch 模式可重複的量稱為調整後人的變異數 (adjusted person variability, SA_p^2)，這個 SA_p^2 可以用總變異減去誤差變異數 ($SD_p^2 - SE_p^2 = SA_p^2$) 而得。這個重複的部份除以人的總變異可以獲得人的信度估計值，這個值 (R_p) 的範圍為 0~1。

在測量的變項中，估計人的分散程度的另一個指標為 person separation index (G_p)，這個估計值為調整後人的標準差 (SA_p) 除以平均的測量誤差 (SE_p)，這個測量誤差被定義為由 Rasch 模式所沒有辦法解釋的部份³⁰。

$$G_p = \frac{SA_p}{SE_p}$$

不同於人分散的信度 (person separation reliability)，人分散的指標 (person separation index) 並不會固定在 0 跟 1 之間，因此對於比較幾個不同的能力分層中可能是比較有用的。例如，為了要在統計上將人區分成不同的能力分層，使用 3 個標準差來定義每一個分層，利用這個公式可以指出在樣本中可區分出不同的能力分層。

(4) 試題差別功能分析 (differential item functioning, DIF)

比較感興趣的不同族群（例如：男/女，有工作/沒有工作，已婚/離婚/未婚）間題目的估計值，來檢查在不同的族群間題目是否有顯著不同的重要性，這稱為試題差別功能 (differential item functioning, DIF)。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³⁰。在本分析中我們要比較性別及年齡是否有 DIF 存在。在年齡方面，以 75 歲作為切點，大於等於 75 歲的老人分為老老人組，小於 75 歲為年輕老人組。

以性別為例，要評估 WHOQOL-BREF 的題目在新社鄉社區老人中，題目的難度會不會因為性別的不同而不同。所以需要以性別分群分別估計男性與女性題目的難度。以男性及女性題目難度畫散佈圖，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⁴⁵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。

第七節 心理計量特質

- (1) 單一向度 (unidimensionality)：單一向度是在評估一組題目所測量的概念是否相同。
- (2) 題目難度與人的能力 (item difficulty and person ability)：Rasch 模式是一個很容易被接受工具來幫助我們來建構一個客觀，可相加的量尺。這個模式可以將原始分數轉換成抽象的，等間距的量尺 (equal-interval scale)。在 Rasch 模式中，利用 logit scale 來測量受測者的能力及題目難度。將機率以 logit 轉換，大部分的資料會落到

-5~5之間。受測者能力是以logit scale來測量，題目的難度也是以logit scale來測量。固定題目的難度後，在能力量尺相差一個單位，則相當於在 θ 量尺上的成功（贊同）的勝算相差約2.72倍。在固定人的能力，在題目難度上相差一個單位，則相當於在成功（贊同）的勝算相差約2.72倍³⁰。

Probability p	Odds p/(1-p)	Logit ln(p/(1-p))
0.00	0.00	-5.00
0.1	0.11	-0.95
0.2	0.25	-0.60
0.3	0.43	-0.37
0.4	0.67	-0.18
0.5	1.00	0.00
0.6	1.50	0.18
0.7	2.33	0.37
0.8	4.00	0.60
0.9	9.00	0.95
1.00	99999.00	5.00

- (3) 難度的重疊性 (item redundancy)：在測量尺度上題目難度差距小於平均題目難度的兩個測量標準誤。
- (4) 難度的順序 (item hierarchy)：Hierarchy表示用題目難度層級來排序的概念，如從最簡單到最難的題目。例如：爬一層樓梯是比爬幾層樓梯簡單的或是穿衣服是比做粗重的家事簡單。在Rasch分析中，人的能力及題目難度的單位都是以logit表示，一個人做一項特定工作的勝算。logit愈大表示題目愈難。一個人能夠做一個特定工作的勝算是這個人能夠做這一項工作的機率比上這個人無法做這個工作的機率。當一份問卷在發展時，若有使用概念化等級的題目，就可以用Rasch分析所產生的實證的題目順序與理論的順序相比較，並且結果可以被視為建構效度 (construct validity) 的證實^{8, 30}。

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

第八節 現代測量理論在生活品質的應用

項目反應理論是在心理計量的方法下提供一個較好的選擇可用於設計合適的問卷、考卷及調查，並且也可用來作題目分析。項目反應理論較傳統測量理論有的優點如下：(a) 題目參數與潛在能力間是獨立的，(b) 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，(c) 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，(d) 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，(e) 可用來產生簡短，複本及特定形式的問卷，並且(f) 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)³⁵。

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。Grratt 等利用 Rasch 分析 Roland Disability Questionnaire (RDQ) 題目是否都是測量相同的概念（單一向度）及題目的分散性如何。結果發現大部分的題目都符合單一向度的假設，並且發現 RDQ 題目的難度在 0 附近，有許多題目都是重複的，故作者建議日後若要修改問卷時可以刪除掉不符合單一向度的題目及重複性的題目，使問卷更簡短且精確的測量人的能力⁴⁰。

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⁴⁷。DIF 分析被用來評估模式-資料的適合度，在不同性別、年齡分層及教育程度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 DIF。根據這

個標準刪除了 7 題，其中包括：生理範疇 1 題、社會範疇 1 題以及環境範疇 5 題，剩下 19 題。發現 19-題的 WHOQOL-BREF 較原本的設計能較簡潔地估計受測者潛在能力。多向度的方法不僅能夠較精確的評估範疇間的相關性，並且也較單一向度的方法有較高的信度⁴⁷。



第三章 研究方法

第一節 研究設計

本研究為一橫斷式研究，主要是以社區老人為研究對象，以訪視的方式收集老人基本資料及 WHOQOL-BREF 生活品質問卷進行兩種測量方法的比較，基本架構如下：



第二節 研究對象

本研究對象為 65 歲以上之新社鄉老人。新社鄉位於台灣中部的台中縣，是一個有高比例老人的地區，所以是一個適合進行老人社區介入的地方。在 2001 年，新社鄉居民中 65 歲以上老人的比例為 11.9%，相較於全台灣的 8.6% 是較高的。在新社鄉 13 個村里中篩選 6 個老人人口比例最高的村里。在台灣所有的人都需要登記在當地的戶政事務所，戶政登記程序可用來核對並提供人口學資訊，並提供官方用來識別個人的狀態及親戚關係。因此，研究開始時在新社鄉戶政事務所登記的基礎上得知 6 個村里中 2072 個 65 歲以上老人的姓名、生日、性別和教育程度。

在兩週的訪視期間，1200 個合適的對象同意參加研究。沒有參加的 872 位研究對象，24 位已死亡，59 位住院或臥病在床，252 位搬到其他地方，323 位在訪視期間不在家及 214 位拒絕訪視。此外，相較於沒有回應的對象，有回應的對象在性別及教育程度有相似的分佈，但有回應的較沒有回應的年輕且有統計的差異¹³。

第三節 研究工具的擬定

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⁴⁸ 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¹³。

第四節 資料收集過程

本研究從當地的戶政資料篩選 65 歲的老人，訪視前寄發明信片給老人告知本研究目的及訪視的時間。受過訓練的訪員進行以結構化的問卷在老人家中進行訪視並收集相關的資料。藉由參與 4 個小時的訓練而標準化訪視的過程及訪視的態度。在開始訪視每一個老人時首先問他/她是否願意自填，假如不願意則有訪員進行個人訪視。結果只有 86 位老人自填 WHOQOL-BREF 問卷。不管是自填或是個人訪視都會計時每一次訪視的長度¹³。

另外，也收集年齡，性別，教育程度，慢性共病症，憂鬱及認知狀態等資訊。利用 24 個可能影響老人較大的慢性病表格來評供老人共病症的情形。利用 15 題老人憂鬱量表(15-item Geriatric Depression)簡短型來評估憂鬱症的情形，GDS 分數愈高表示有憂鬱的情形。利用簡易智能評估量表(Mini-Mental State Examination, MMSE)來評估老人的認知情形。這份問卷常見在流行病學研究中著重於評估研究對象在方向感、記錄事項、訊息回想、注意力、算數、語言及視覺空間架構的認知情形。在流行病學委託的研究(Epidemiologic Catchment Area Studies)中，MMSE 計分分類的慣例為 0-17 分表示認知缺陷⁴⁹。

第五節 資料統計與分析

本研究利用 Rasch 分析評估 WHOQOL-BREF 是否具有下列心理計量特質：

- (1) 單一向度(unidimensionality)
- (2) 題目難度(item difficulty)
- (3) 題目適中性(targeting)
- (4) 訊息函數功能(information function)
- (5) 鑑別力分析(discriminat analysis)
- (6) 區別性指標(person separation index)
- (7) 試題差別功能 (differential item functioning, DIF)
- (8) 整合性分析(integrated analysis)

分別詳述於後：

(1) 單一向度 (unidimensionality)

在同一個範疇中的題目應該要評估一個單一的方向或概念。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)⁵⁰，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⁵¹，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。適合度的評估與卡方分析有相同的概念都是比較觀察值與預測值之間的關係。

(2) 題目的難度 (item difficulty) 分佈

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)^{40,52}。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)⁸。先依題目的難度先排序，再求毗鄰題目間難度的差異，若是難度差異值小於 2 個測量標準誤，表示題目有重疊性存在。

(3) 題目適中性 (targeting)

Targeting 指標用來評估題目難度與研究對象能力的適中性。在題目的校準上，在不同題目間平均難度及 rating scale categories 都固定在 0，因此，範疇的平均得分若為 0，表示這些題目的難度對研究對象是適中的。當平均範疇得分離 0 越遠，表示這一組題目是 mis-targeted。當範疇平均分數為 0.5 時，表示 slight mis-targeted。範疇平均分數為 1.0 時，表示 substantial mis-targeted。範疇平均分數若為 '正' 表示對於這個樣本這一組題目是簡單的，範疇平均分數若為 '負' 表示對於這個樣本這一組題目是困難的⁸。

(4) 訊息函數功能 (information function)

訊息函數可以讓研究者瞭解每一個題目在連續的能力特質尺度的每一點所能提供的訊息，當訊息函數的數值愈高時表示題目在對於此能力下的人所能提供的訊息愈高，即能較準確的估計在此能力下的人。在本分析中藉由將每一個題目的訊息函數相加，來評估哪些題目所能提估的訊息最高，即評估哪些題目在評估老人的生活品質時，能夠估計的最準確。另外，在每個範疇中將在每一個能力間隔下的訊息函數相加，來探

討這些題目在哪一個能力區間下能夠提供較高的訊息，即能較準確估計
35。

(5)鑑別力分析 (discriminant analysis)

鑑別力是用於測定該題目是否能有效地將不同能力的人區別出來。一般用 Rasch 系列的結果比較少在看鑑別力，因為在 Rasch 模式中會將此參數預測為 1，估計完所有的參數之後再來估計鑑別力，是屬於事後估計所得的。如果小於 1 表示題目鑑別能力的程度差；如果大於 1 表示題目鑑別能力的程度佳⁴⁶。

(6)區別性指標(person separation index)

Rasch 分析產生一個 person separation 指標在每一個範疇中來區別出人的分層(能力明顯不同的人)。人的區別性指標 (person separation index, G)，公式如下：

$$\text{Separation index (G)} = \frac{\text{adjusted person variability}}{\text{error variability}}$$

G 指標愈大，愈能在測量中區別出人的能力。Person separation index 為 1.5 代表為可接受的區辨能力，2.0 代表為良好的區辨能力，3.0 代表非常好的區辨能力。

在樣本中區別明顯的分層可用以下公式計算⁵¹：

$$\text{分層個數} = \frac{(4G + 1)}{3}$$

假如得到的數值為 1.5 表示樣本中可以分為兩層(高、低)，數值為 2.00 表示可以分為三層(高、中、低)。題目的區別性指標(item separation index)與人的區別性指標概念同。

Person separation reliability

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。

$$\text{Separation reliability} = \frac{\text{adjusted variability}}{\text{total variability}}$$

數值的範圍為 0~1 並且結果解釋同 Cronbach's alpha。Separation reliability 為 1.5 時相似於係數為 0.7（可接受），2.0 時相似於係數為 0.8（良好），3.0 時相似於係數為 0.9（極好）⁸。

(7) 試題差別功能 (differential item functioning, DIF)

DIF 方法是用於測定在不同的族群間題目測量值是否精確地測量相同的概念。在本研究中分性別（及年齡）次族群針對題目難度畫散佈圖，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³⁵，及以 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。

(8) 整合性分析 (integrated analysis) : Person-threshold map

將人的能力及題目的難度放在同一個尺度上，使人的能力及題目的難度可以相比較。在本分析中以 Person-threshold map 來進行兩者間的比較。

統計分析工具

1. 以 SAS 作統計分析及資料處理。
2. 以 WINSTEP 套裝軟體進行 partial credit model 分析。

第四章 研究結果

第一節 基本人口學變項之敘述統計

表一為新社鄉老人基本人口學特質。新社鄉老人總樣本數為1200人，其平均年齡為73.4歲，標準差為6.04，範圍介於65~103歲，其中男性有709人(佔59%)，與配偶同住者有766人(佔66%)，沒有受過正式的教育者有387人(佔37%)，這些老人中有821位曾經醫師診斷為患有慢性疾病，其中353人罹患一種慢性疾病，225人罹患兩種慢性疾病，並且有243人至少罹患三種慢性疾病以上。以老人憂鬱症量表評估有憂鬱症者有63人(佔5%)；以簡短心智評估量表來評估，有認知缺陷者有220人(佔18%)。利用WHOQOL-BREF評估老人的生活品質，以傳統的計分方式在生理、心理、社會及環境範疇的得分分別為13.3、12.8、13.2及13.1，這些範疇的內部一致性都在0.7以上顯示各範疇皆有良好的信度。



Table 1 Demographic characteristic of community-dwelling older people

Variable	number	Percent (%)
Age	73.4 (6.04) ^a	65 ~ 103 ^b
Gender		
Male	709	59.08
Female	491	40.92
Live with spouse		
Yes	766	56.84
No	517	43.16
Physician-diagnosed chronic condition	821	
1	353	43
2	225	27
>3	243	30
Depression		
No	1135	94.74
Yes (GDS score >10)	63	5.26
Cognitive impairment		
No	980	81.67
Yes (MMSE score >18)	220	18.33
Score distribution of WHOQOL-BREF		
Overall	12.77 (2.54) ^a	
Physical domain	13.31 (2.25) ^a	0.80 ^c
Psychological domain	12.76 (2.14) ^a	0.81 ^c
Social domain	13.16 (1.92) ^a	0.73 ^c
Environmental domain	13.08 (1.99) ^a	0.80 ^c

a .means (SD); b: range c. Cronbach's alpha

第二節 問卷心理計量特質分析

本節將利用Rasch模式針對世界衛生組織生活品質問卷簡明版在社區老人的各項心理計量特質進行分析：

(1) 單一向度假設(unidimensionality)

從表二的適合度指標(infit statistic)，WHOQOL-BREF的四個範疇中，適合度指標(infit statistic)的範圍在生理、心理、社會及環境範疇分別為0.83~1.35、0.82~1.38、0.81~1.24及0.8~1.17，WHOQOL-BREF各範疇中的題目都符合單一向度的假設。

(2) 題目難度的分析

題目難度的估計及順序

表二顯示WHOQOL-BREF各範疇題目難度的估計值結果，並依難度順序排列。在生理範疇中難度的範圍介於-1.1~0.48，最難以回答滿意的題目為PN10”足夠的精力”(0.48)，其餘依次為PN15”四處行走的能力”(0.42)、PN18”工作能力”(0.42)、PN17”日常活動的能力”(0.21)、PN16”睡眠狀況”(0.16)、PN4”依賴藥物”(-0.6)及PN3”身體疼痛”(-1.1)。在心理範疇中難度範圍為-1.19~1.29，最難以回答滿意的題目為YN5”享受生活”(1.29)，其餘依次為YN7”集中精神”(0.67)、YN6”生命有意義”(0.37)、YN11”接受自己的外表”(-0.52)、YN19”對自己滿意”(-0.61)及YN26”常有負面感受”(-1.19)。在社會範疇中難度範圍為-0.77~1.17，最難以回答滿意的題目為SN21”性生活”(1.17)，其餘依次為SN27”受尊重或是有面子”(0.31)、SN20”人際關係”(-0.71)及SN22”朋友支持”(-0.77)。在環境範疇中難度範圍為-1.08~0.96，最難以回答滿意的題目為EN14”有機會從事休閒活動”(0.96)，其餘依次為EN12”足夠的金錢應付所需”(0.94)、EN13”每日生活的資訊”(0.29)、EN8”感到安全”(0.1)、EN24”醫療保健服務的方便

性”(0.01)、EN23”住所的狀況”(-0.56)、EN9”環境健康”(-0.63)及EN28”想吃的食物的都吃的到”(-1.08)。



Table 2 Item calibrations, misfit statistics and slope estimates of the WHOQOL-BREF domains in community elderly and control group.

Domains/Item	Slope parameter $\hat{\alpha}$ (s.e.#)	Item difficulty (S.E.)	Infit Mean Square
Physical domain			
PN3. Pain and discomfort	0.94	-1.1 (0.04)	1.05
PN4. Dependence on medication or treatment	0.59	-0.6 (0.04)	1.35
PN10. Energy and fatigue	1.07	0.48 (0.05)	0.93
PN15. Mobility	1.1	0.42 (0.04)	0.9
PN16. Sleep and rest	0.81	0.16 (0.05)	1.2
PN17. Activities of daily living	1.26	0.21 (0.05)	0.72
PN18. Working capacity	1.15	0.42 (0.05)	0.83
Psychological domain			
YN5. Enjoyment of life	1.12	1.29 (0.05)	0.87
YN6. Spirituality, religion and personal beliefs	1.16	0.37 (0.06)	0.82
YN7. Thinking, learning, memory & concentration	1.05	0.67 (0.05)	0.95
YN11. Body image and appearance	1.04	-0.52 (0.06)	0.95
YN19. Self-satisfaction	1.02	-0.61 (0.06)	0.97
YN26. Negative feeling	0.55	-1.19 (0.05)	1.38
Social domain			
SN20. Personal relationship	1.12	-0.71 (0.07)	0.84
SN21. Sexual activity	0.92	1.17 (0.08)	1.04
SN22. Friends' support	1.17	-0.77 (0.07)	0.81
SN27. Esteemed and respected	0.79	0.31 (0.07)	1.24
Environmental domain			
EN8. Physical safety and security	1.03	0.1 (0.04)	0.99
EN9. Physical environments	0.88	-0.63 (0.05)	1.14
EN12. Financial environment	1	0.94 (0.04)	1
EN13. Opportunities	1.14	0.29 (0.05)	0.86
EN14. Participation & support of leisure activities	0.82	0.96 (0.04)	1.15
EN23. Home environment	1.2	-0.56 (0.05)	0.8
EN24. Health & social care: availability & quality	1.03	0.01 (0.05)	0.96
EN25. Transportation	1.04	-0.03 (0.05)	0.96
EN28. Eating food	0.82	-1.08 (0.05)	1.17

Estimates were obtained from partial credit model by WINSTEP.

* Item misfit (infit statistic >1.4)

Abbreviations: PN = physical domain; YN = psychological domain; SN = social domain; EN = environmental domain; the numbers were the item number.

題目難度的範圍

表三將WHOQOL-BREF四個範疇中題目難度及門檻估計值的範圍與受測老人之能力估計值作一對照。就題目平均難度而言，生理範疇介於-1.1~0.48、心理範疇-1.19~1.29、社會範疇-0.77~1.17、及環境範疇-1.19~1.29；就題目的門檻值而言，生理-3.86~5.3、心理-6.74~6.78、社會-8.39~9.66、及環境-3.72~4.75；老人能力範圍則生理範疇介於-6.7~7.76、心理-5.78~8.79、社會-9.34~10.98、及環境-5.25~5.41。從門檻值與95%老人能力的分佈中可看出題目門檻值的分佈是較大，故WHOQOL-BREF是適用於評估大部分老人的生活品質，但是在能力最好及最差則較不準。

Table 3. The range of item threshold and ability across four domains of WHOQOL-BREF in community older people.

	Physical domain	Psychological domain	Social domain	Environmental domain
Item threshold	-3.86~5.3	-6.74~6.78	-8.39~9.66	-3.72~4.75
Ability	-6.7~7.76	-5.78~8.79	-9.34~10.98	-5.25~5.41
95% ability *	-1.52~3.28	-2.37~3.27	-3.29~5.2	-1.62~2.38

*: the range of 95% was from the 2.5th percentile to 97.5th percentile.

難度間距及重疊性

圖四可方便看出題目是否均勻分佈在量尺上，完美的均勻分佈表示題目之間沒有間距且沒有重疊。若定義相差一個 logit 表示有間距(gap)存在，則在各範疇題目難度的範圍內，僅社會範疇在 SN20(難度=-0.71)及 SN27(0.31)之間有一間距(gap)存在。進一步針對重疊性，我們定義若兩個題目難度的差距是小於兩倍的各範疇平均難度標準誤，則表示兩題具有重疊性，表四列出各範疇具有重疊性的各組題目，在生理範疇中 PN15”四處行走的能力(0.42)”與 PN18”工作能力”(0.42)有重疊性。在心理範疇中 YN19”對自己滿意”(-0.61) 與 YN11”接受自己的外表”(-0.52)有重疊性。在社會範疇題目分佈均勻，但在 SN20”人際關係”(-0.71) SN22”朋友支

持”(-0.77)有題目重疊性存在。在環境範疇中題目分佈均勻但是重疊性的題目較多，有 EN9”環境健康”(-0.63)與 EN23”住所的狀況”(-0.56)；EN24”醫療保健服務的方便性”(0.01)與 EN25”交通運輸方式”； EN24”醫療保健服務的方便性”(0.01)與 EN8”感到安全”(0.1)；EN12”足夠的金錢應付所需”(0.94)與 EN14 ”有機會從事休閒活動”(0.96)。

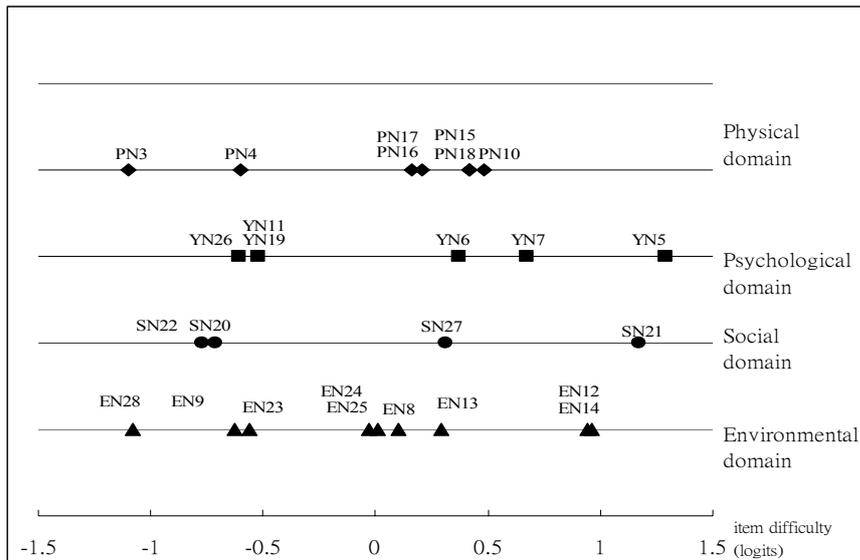


Figure 4. A mean of item difficulty for the four domains of WHOQOL-BREF separated in community elderly.

難度順序 (item hierarchy)

表五為題目難度順序的評估，將本研究社區老人難度順序與 Wang(2006)⁴⁷ 研究中20~65歲之題目順序相比較。在生理範疇中，在20~65歲的人中PN4”依賴藥物”是較PN3”身體疼痛”容易回答滿意的，但是在老人PN4”依賴藥物”是較PN3”身體疼痛”容易回答滿意。以及，PN16”睡眠狀況”在老人這一題是較容易回答滿意；但在較年輕的族群這是較難回答滿意。在心理範疇中，YN19”對自己滿意”及YN11”接受自己的外表”在這兩族群間題目順序的不同可能是題目的重覆性所造成的。另外，在YN26”常有負面感受”相對於20~65歲的人，老人是較容易回答滿意的。

Table 4 Items with redundancy in each domain of WHOQOL-BREF

Redundancy Group	Item	difficulty	difficulty difference
Physical domain (2SE=0.1)			
1	PN15. Mobility	0.42	0
1	PN18. Working capacity	0.42	
Psychological domain (2SE=0.12)			
1	YN19. Self-satisfaction	-0.61	0.09
1	YN11. Body image and appearance	-0.52	
Social domain (2SE=0.14)			
1	SN22. Friends' support	-0.77	0.06
1	SN20. Personal relationship	-0.71	
Environmental domain (2SE=0.1)			
1	EN9. Physical environments	-0.63	0.07
1	EN23. Home environment	-0.56	
2	EN25. Transportation	-0.03	0.04
2	EN24. Health & social care: availability & quality	0.01	
3	EN24. Health & social care: availability & quality	0.01	0.09
3	EN8. Physical safety and security	0.1	
4	EN12. Financial environment	0.94	0.02
4	EN14. Participation & support of leisure activities	0.96	

Table 5. Comparison of item hierarchy between two studies.

	Physical domain ^a Item no.(calibration)	Psychological domain Item no.(calibration)	Social domain ^b Item no.(calibration)	Environmental domain ^c Item no.(calibration)
Wang W.C. ⁴⁷ *	PN4(-0.77) , PN3(-0.45), PN17(-0.1), PN18(0.09), PN10(0.61), PN16(0.61)	YN11(-0.57) , YN19(-0.54), YN6(-0.21), YN26(-0.03) , YN7(0.03), YN5(1.32)	SN22(-0.39), SN20(-0.22), SN27(0.61)	EN9(-0.27), EN25(-0.25), EN24(0.08),EN8(0.43)
This study**	PN3(-1.1), PN4(-0.6) , PN16(0.16) , PN17(0.21), PN18(0.42), PN10(0.48)	YN26(-1.19) , YN19(-0.61), YN11(-0.52) , YN6(0.37), YN7(0.67), YN5(1.29)	SN22(-0.77), SN20(-0.71), SN27(0.31)	EN9(-0.63), EN25(-0.03), EN24(0.01), EN8(0.1)

*Item parameters were estimated by multidimensional Rasch model.

**Item parameters were estimated by unidimensional Rasch model.

a. Deleted item 15 in physical domain; b Deleted item 21 in social domain; c. Deleted item 9, item12, item13, item14 and item 28 in environmental domain.

Abbreviations: PN = physical domain; YN = psychological domain; SN = social domain; EN = environmental domain; the numbers were the item number.

(3) 題目適中性 (targeting)

表六中受測者能力估計值的平均在WHOQOL-BREF中生理、心理、社會及環境範疇分別為0.57、0.44、1.02及0.36。除了社會範疇的題目對於老人是相對簡單的題目外，其他範疇的題目對於這些老人是適中的。

(4)區別性指標 (separation index)

在表六中以題目作為基準，在題目難度中各指標所具有的意義僅作參考不作詳細討論。針對受試者潛在特質(person latent trait)，於WHOQOL-BREF的生理、心理、社會及環境範疇之區別性指標依序為1.79、1.83、1.45及1.86。根據此結果可計算出理想的分層個數，結果分別為2.7、2.8、2.3及2.8，因都小於3顯示各範疇皆能將老人的能力區分為三組。反應出使用WHOQOL-BREF可將老人區分為高、中及低能力三組。

在信度的評估中，信度的指標是經區別性指標轉換而得的，故其與區別性指標的結果是相互呼應的。在IRT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，在生理、心理、社會及環境範疇的信度指標分別為0.76、0.77、0.68及0.78。除了社會範疇的信度指標輕微的小於0.7外，其他的範疇皆顯示良好的內部一致性。

Table 6 Targeting and Separation indices of WHOQOL-BREF in community older people

Domains	Person latent trait (N=1200)			Item difficulty		
	Targeting index mean $\hat{\theta}$ (s.e.#)	Separation index	Reliability	Targeting index mean $\hat{\beta}$ (s.e.#)	Separation index	Reliability
Physical	0.57 (1.45)	1.79	0.76	0 (0.05)	11.5	0.99
Psychological	0.44 (1.86)	1.83	0.77	0 (0.06)	14.98	1.00
Social	1.02 (2.66)	1.45	0.68	0 (0.07)	10.73	0.99
Environmental	0.36 (1.30)	1.86	0.78	0 (0.05)	13.45	0.99

Estimates were obtained from partial credit model by WINSTEP. # Standard error

(5) 試題差別功能 (differential item functioning)

以性別及年齡將老人分為兩群來比較各範疇中題目間，是否會因為人特質的不同導致題目的特質也不同。表七為分性別的題目的難度，在生理範疇中男性難度的範圍為-1.03 ~ 0.46，女性為-1.22 ~ 0.49。在心理範疇中男性難度的範圍為-1.33 ~ 1.3，女性為-1.12 ~ 1.26。在社會範疇中男性難度的範圍為-0.7 ~ 1，女性為-1.79 ~ 2.07。在環境範疇中男性難度的範圍為 0.88 ~ 1，女性為 -0.79 ~ 0.94。大致上，題目難度的分佈是相同的，然而，女性在生理及環境範疇的分佈較男性廣。

圖五為生理範疇分性別的來看題目難度的散佈圖，發現生理範疇的題目並沒有 DIF。從圖六中我們可以發現心理範疇中 YN6 及 YN19 這兩題有輕微的 DIF。從圖七可以發現社會範疇中 SN21 及 SN22 嚴重的 DIF，SN21”性生活”在男性是較難的題目，但在女性卻是較簡單的題目；SN22”朋友支持”在女性是較難的題目，但在男性是較簡單的。圖八為環境範疇，沒有題目的點落在此範圍外，沒有 DIF 存在。

Table 7 Item calibrations of the WHOQOL-BREF domains in male and female.

Items	Male (N=709)		Female (N=491)	
	difficulty	error	difficulty	error
Physical domain				
PN3. Pain and discomfort	-1.14	0.05	-1.01	0.07
PN4. Dependence on medication or treatment	-0.68	0.05	-0.54	0.07
PN10. Energy and fatigue	0.52	0.06	0.4	0.08
PN15. Mobility	0.24	0.05	0.68	0.07
PN16. Sleep and rest	0.37	0.06	-0.17	0.08
PN17. Activities of daily living	0.28	0.07	0.13	0.09
PN18. Working capacity	0.41	0.07	0.51	0.09
Psychological domain				
YN5. Enjoyment of life	0.93	0.06	1.33	0.09
YN6. Spirituality, religion and personal beliefs	0.01	0.07	0.38	0.1
YN7. Thinking, learning, memory & concentration	0.15	0.07	0.84	0.09
YN11. Body image and appearance	1.51	0.08	-0.67	0.1
YN19. Self-satisfaction	-1.12	0.08	-0.59	0.1
YN26. Negative feeling	-1.48	0.06	-1.29	0.08
Social domain				
SN20. Personal relationship	-0.89	0.09	-0.49	0.11
SN21. Sexual activity	1.12	0.1	-1.51	0.14
SN22. Friends' support	-0.62	0.09	1.74	0.12
SN27. Esteemed and respected	0.38	0.08	0.26	0.11
Environmental domain				
EN8. Physical safety and security	0.08	0.06	0.14	0.08
EN9. Physical environments	-0.57	0.06	-0.73	0.09
EN12. Financial environment	1.06	0.05	0.77	0.08
EN13. Opportunities	0.1	0.06	0.62	0.09
EN14. Participation & support of leisure activities	0.87	0.05	1.13	0.07
EN23. Home environment	-0.57	0.06	-0.53	0.09
EN24. Health & social care: availability & quality	0.09	0.06	-0.17	0.09
EN25. Transportation	0.01	0.06	-0.11	0.08
EN28. Eating food	-1.06	0.06	-1.12	0.08

Estimates were obtained separately from partial credit model by WINSTEP.

Abbreviations: PN = physical domain; YN = psychological domain; SN = social domain; EN = environmental domain; the numbers were the item number.

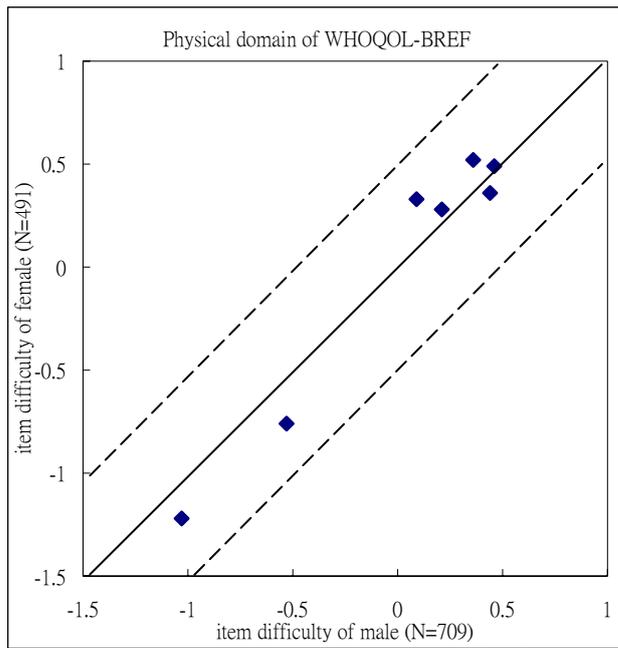


Figure 5. DIF comparison on male and female in physical domain of WHOQOL-BREF

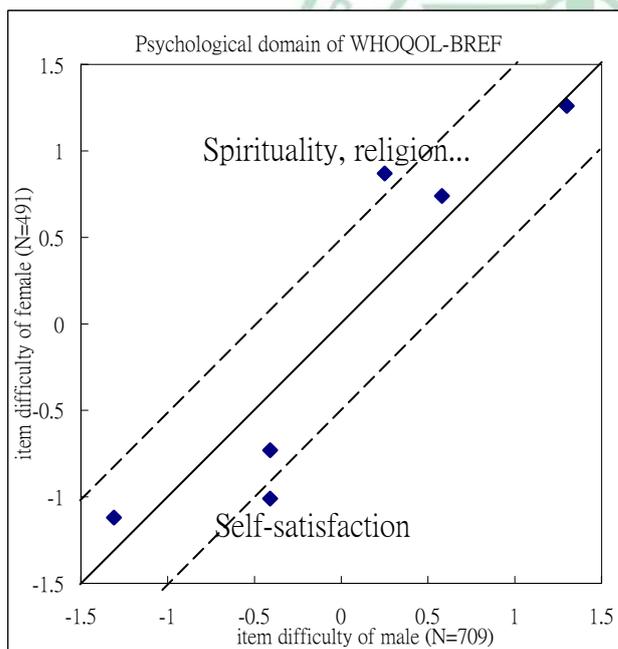


Figure 6. DIF comparison on male and female in psychological domain of WHOQOL-BREF

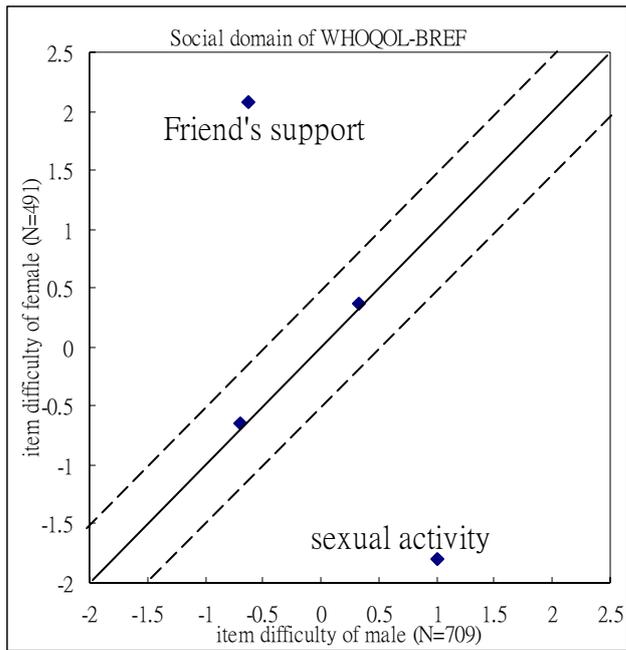


Figure 7. DIF comparison on male and female in social domain of WHOQOL-BREF

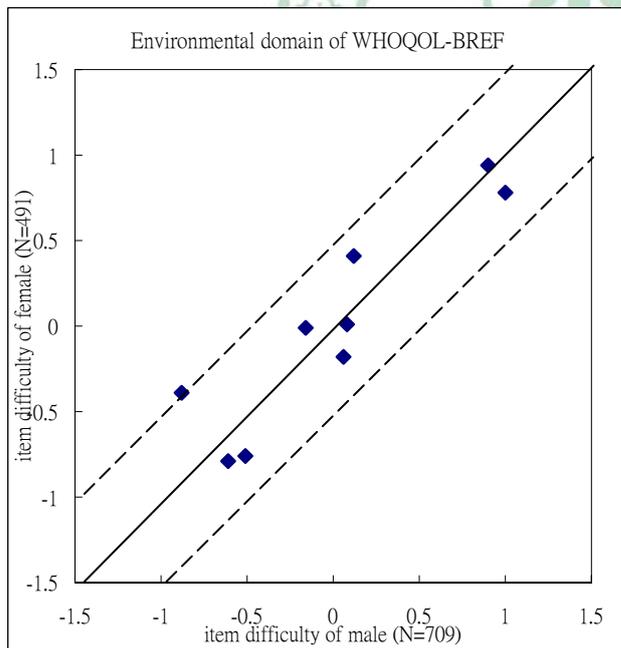


Figure 8. DIF comparison on male and female in environmental domain of WHOQOL-BREF

表八將老人分為年輕老人(65~75 歲)及老老人(≥ 75 歲)，分年齡層的題目的難度，在生理範疇中年輕老人難度的範圍為-1.14~0.52，老老人為-1.01~0.68。在心理範疇中年輕老人難度的範圍為-1.48~1.51，老老人為-1.29~1.33。在社會範疇中年輕老人難度的範圍為-0.89~1.12，老老人為-1.51~1.74。在環境範疇中年輕老人難度的範圍為-1.06~1.06，老老人為-1.12~1.13。大致上，題目難度的分佈是相似的。

圖九為生理範疇分年齡層來看題目難度的散佈圖，發現生理範疇的題目 PN16”睡眠狀況”有輕微的 DIF。從圖十中可以發現心理範疇中”YN7”集中精神”有輕微的 DIF，以及 YN11”接受自己的外表”有明顯的 DIF，YN11”接受自己的外表”在年輕老人是最難的題目，但在老老人則是相對簡單的題目。從圖十一可以發現社會範疇中 SN21”性生活”及 SN22”朋友支持”明顯的 DIF，SN21”性生活”在年輕老人是較難的題目，但在老老人卻是較簡單的題目；SN22”朋友支持”在老老人是較難的題目，但在年輕老人是較簡單的。而在圖十二環境範疇中 EN13”每日生活的資訊”有輕微的 DIF 存在。

Table 8 Item calibrations of the WHOQOL-BREF domains in young and old older people

Items	Young older people (N=800)		Old older people (N=400)	
	difficulty	error	difficulty	error
Physical domain				
PN3. Pain and discomfort	-1.14	0.05	-1.01	0.07
PN4. Dependence on medication or treatment	-0.68	0.05	-0.54	0.07
PN10. Energy and fatigue	0.52	0.06	0.4	0.08
PN15. Mobility	0.24	0.05	0.68	0.07
PN16. Sleep and rest	0.37	0.06	-0.17	0.08
PN17. Activities of daily living	0.28	0.07	0.13	0.09
PN18. Working capacity	0.41	0.07	0.51	0.09
Psychological domain				
YN5. Enjoyment of life	0.93	0.06	1.33	0.09
YN6. Spirituality, religion and personal beliefs	0.01	0.07	0.38	0.1
YN7. Thinking, learning, memory & concentration	0.15	0.07	0.84	0.09
YN11. Body image and appearance	1.51	0.08	-0.67	0.1
YN19. Self-satisfaction	-1.12	0.08	-0.59	0.1
YN26. Negative feeling	-1.48	0.06	-1.29	0.08
Social domain				
SN20. Personal relationship	-0.89	0.09	-0.49	0.11
SN21. Sexual activity	1.12	0.1	-1.51	0.14
SN22. Friends' support	-0.62	0.09	1.74	0.12
SN27. Esteemed and respected	0.38	0.08	0.26	0.11
Environmental domain				
EN8. Physical safety and security	0.08	0.06	0.14	0.08
EN9. Physical environments	-0.57	0.06	-0.73	0.09
EN12. Financial environment	1.06	0.05	0.77	0.08
EN13. Opportunities	0.1	0.06	0.62	0.09
EN14. Participation & support of leisure activities	0.87	0.05	1.13	0.07
EN23. Home environment	-0.57	0.06	-0.53	0.09
EN24. Health & social care: availability & quality	0.09	0.06	-0.17	0.09
EN25. Transportation	0.01	0.06	-0.11	0.08
EN28. Eating food	-1.06	0.06	-1.12	0.08

Estimates were obtained separately from partial credit model by WINSTEP.

Abbreviations: PN = physical domain; YN = psychological domain; SN = social domain; EN = environmental domain; the numbers were the item number.

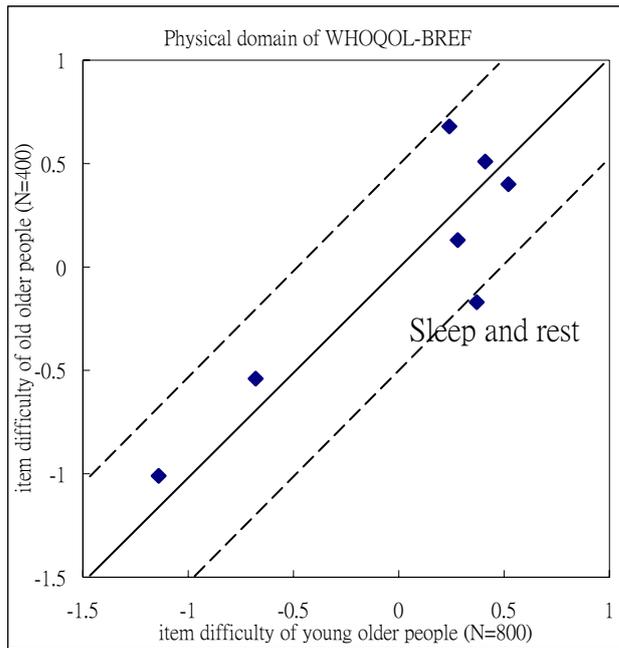


Figure 9. DIF comparison on young and old older people in physical domain of WHOQOL-BREF

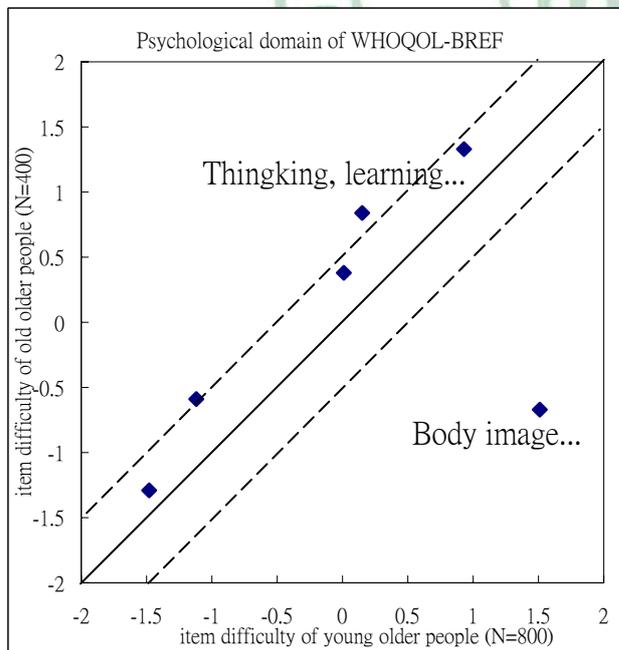


Figure 10. DIF comparison on young and old older people in psychological domain of WHOQOL-BREF

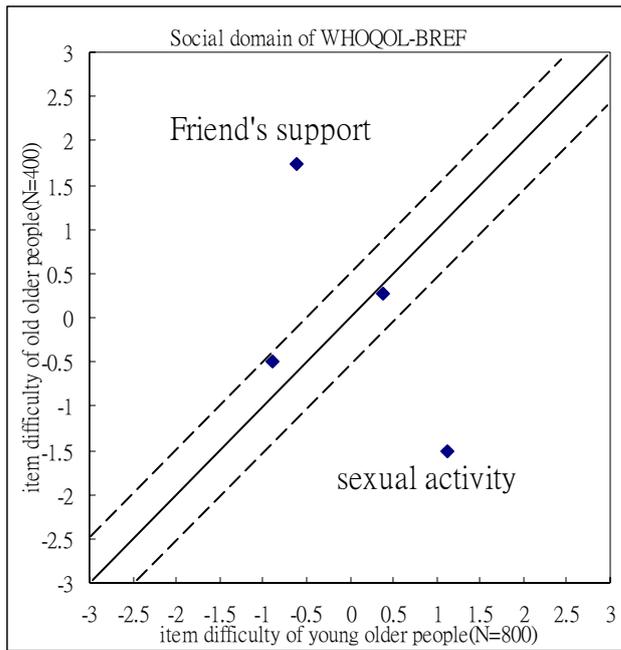


Figure 11. DIF comparison on young and old older people in social domain of WHOQOL-BREF

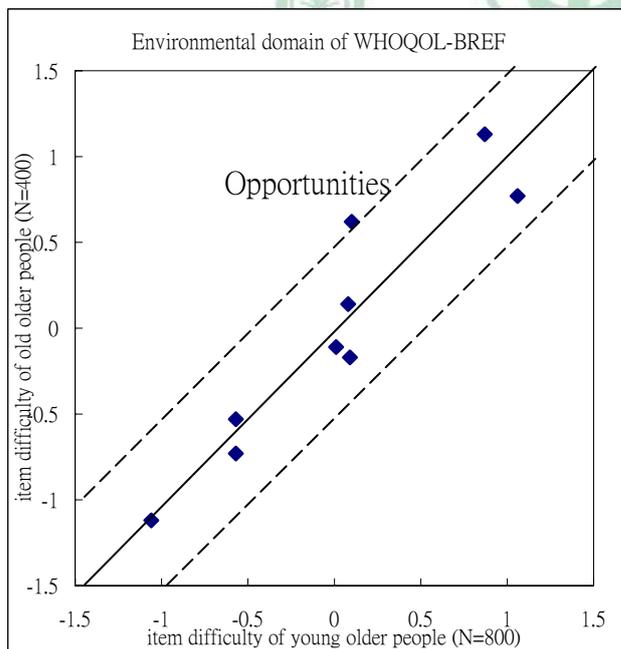


Figure 12. DIF comparison on young and old older people in environmental domain of WHOQOL-BREF

從以上 DIF 的分析，我們可以發現社會範疇的 SN21”性生活”及 SN22”朋友支持在性別及年齡有明顯的 DIF 存在，所以將針對以下兩種方式來評估社會範疇中題目在性別及年齡間 DIF 的情形：(1)年齡分組來看性別間 DIF 的情形，(2)分性別來看年輕老人與老老人間 DIF 的情形。

從圖十三中以分年齡來看性別間 DIF 的情形，左邊為較年輕老人性別間 DIF 的情形，SN21”性生活”在男性是較難的題目，在女性則是較簡單的題目；SN20”人際關係”及 SN22”朋友支持”在男性是相對簡單的題目，在女性是相對難的題目。右邊為老老人性別間的 DIF，SN22”朋友支持”及 SN27”受尊重或是有面子”有輕微的 DIF。

圖十四分性別來看年輕老人與老老人間 DIF 的情形，左邊為男性在年齡上 DIF 的情形，SN21”性生活”在年輕老人是較難的題目，在老老人則是較簡單的題目；SN22”朋友支持”在年輕老人是較簡單的題目，在老老人則是較難的題目。右邊為女性在年齡上 DIF 的情形，SN21”性生活”在年輕老人及老老人都是最簡單的題目，但是老老人中對於性生活是更容易滿足的；SN20”人際關係”及 SN22”朋友支持”在年輕老人在題目的難度上是很相近的，但是在老老人 SN22”朋友支持”是非常難的題目。

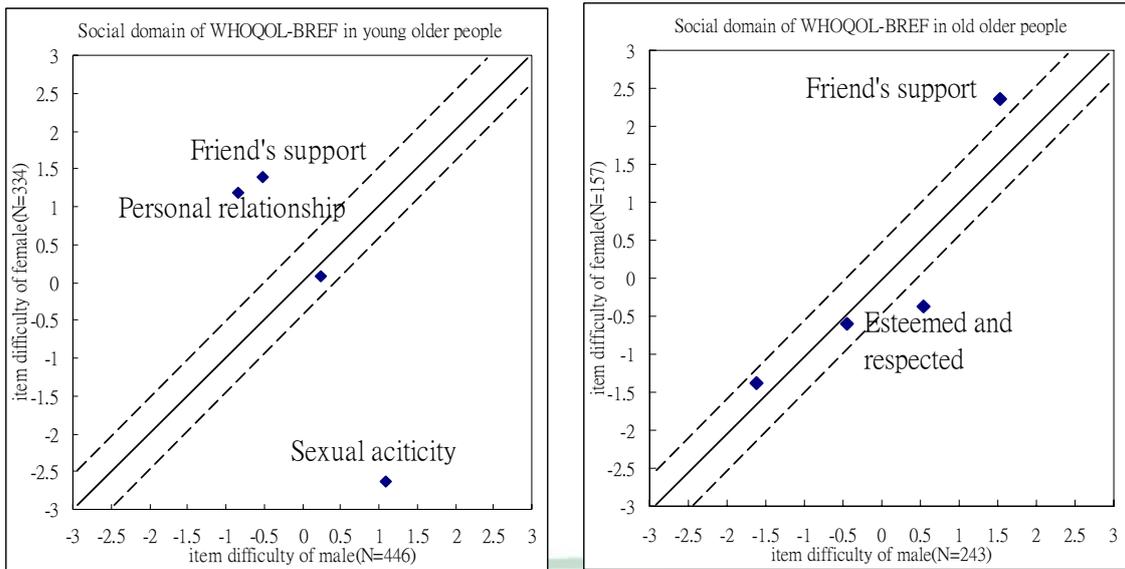


Figure 13. DIF comparison on male and female in social domains of WHOQOL-BREF between young and old older people.

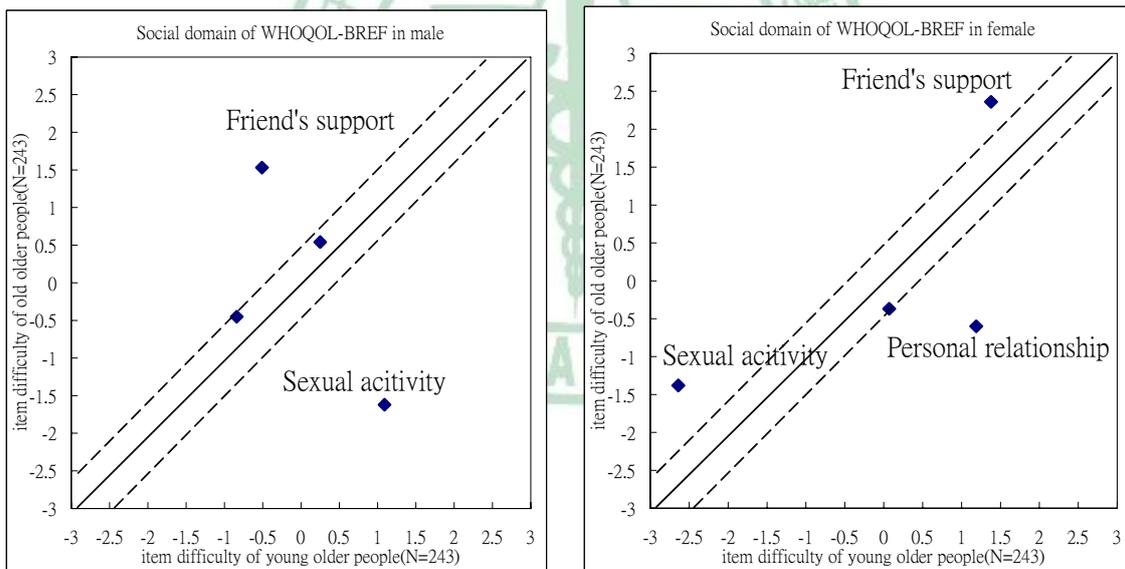


Figure 14. DIF comparison on young and old older people in social domains of WHOQOL-BREF between male and female.

(6) 訊息函數 (information function)

表九為 WHOQOL-BREF 每個範疇中在從-5 到 5 能力尺度下每題目的訊息函數 (item information function)，題目加總每個能力下的訊息函數來評估題目對於老人所能提供的訊息量為何，以及加總範疇中每一個能力下所能提供的訊息量來評估範疇對於哪一個能力區間下的人所能提供的訊息最大，就是測得較準確。

訊息函數受到很多因子的影響，尤其是測量的誤差，而測量的誤差很容易受樣本數的影響。以 PN3 為例，PN3 是較簡單的題目，老人們填答的分佈 1”極妨礙”(1.25)、2”很妨礙”(7.69)、3”中等程度妨礙”(26.09)、4”有一點妨礙”(34.45)、5”完全沒有妨礙”(30.52)。老人們的回答集中在選項 3、4、5，所以在這些選項中的樣本數會較多，表五也顯示 PN3 在能力為-1.5~2 附近能估計的較準確。

加總每個能力下的訊息函數後，來看每一個各題目在估計老人生活品質時所能提供的訊息，如果加總的訊息量大於 7，即表示此題目對於老人生活品質能夠測的較準確。在生理範疇方面，PN3、PN4 及 PN15 所提供的訊息較高。在生理範疇，除了 YN26 外，其他題目加總的訊息量都小於 7。在社會範疇的，每個題目能提供的訊息都較差。而環境範疇，除了 EN25，大部分的題目的訊息量都大於 7。這可能與題目的門檻值及老人的能力分佈有顯著的相關。

Table 9 Item information for ranges of the latent trait

Item	-5	-4.5	-4	-3.5	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	Weighted I(θ)
Physical domain																						
PN3	0.08	0.12	0.18	0.26	0.33	0.41	0.47	0.505	0.54	0.565	0.585	0.61	0.62	0.605	0.54	0.44	0.345	0.24	0.16	0.105	0.06	7.77
PN4	0.12	0.18	0.24	0.3	0.335	0.36	0.38	0.395	0.42	0.455	0.485	0.54	0.61	0.655	0.64	0.53	0.405	0.27	0.17	0.11	0.07	7.67
PN10	0.24	0.295	0.33	0.35	0.36	0.37	0.385	0.39	0.38	0.37	0.365	0.36	0.34	0.315	0.27	0.245	0.23	0.24	0.26	0.27	0.27	6.635
PN15	0.05	0.085	0.14	0.24	0.365	0.54	0.7	0.75	0.7	0.595	0.515	0.45	0.395	0.365	0.34	0.33	0.315	0.29	0.23	0.18	0.13	7.705
PN16	0.22	0.285	0.34	0.375	0.39	0.4	0.41	0.4	0.39	0.37	0.365	0.36	0.335	0.315	0.27	0.245	0.24	0.25	0.27	0.28	0.26	6.77
PN17	0.24	0.32	0.39	0.445	0.46	0.44	0.4	0.35	0.3	0.28	0.28	0.29	0.3	0.3	0.28	0.25	0.235	0.24	0.255	0.27	0.27	6.595
PN18	0.22	0.31	0.39	0.455	0.48	0.46	0.41	0.355	0.3	0.275	0.275	0.29	0.305	0.305	0.29	0.265	0.255	0.26	0.275	0.28	0.27	6.725
Weighted I(θ)	1.17	1.595	2.01	2.425	2.72	2.98	3.155	3.145	3.03	2.91	2.87	2.9	2.905	2.86	2.63	2.305	2.025	1.79	1.62	1.495	1.33	
Psychological domain																						
YN5	0.27	0.35	0.44	0.485	0.495	0.45	0.395	0.32	0.275	0.26	0.27	0.29	0.29	0.275	0.24	0.21	0.18	0.185	0.205	0.23	0.26	6.375
YN6	0.305	0.34	0.36	0.36	0.36	0.35	0.335	0.295	0.255	0.23	0.23	0.25	0.275	0.295	0.3	0.28	0.26	0.25	0.26	0.27	0.28	6.14
YN7	0.275	0.34	0.41	0.445	0.455	0.43	0.39	0.33	0.285	0.27	0.27	0.29	0.3	0.285	0.26	0.23	0.21	0.21	0.225	0.25	0.27	6.43
YN11	0.235	0.22	0.22	0.235	0.265	0.28	0.275	0.245	0.21	0.18	0.17	0.18	0.205	0.245	0.27	0.28	0.27	0.245	0.23	0.23	0.25	4.94
YN19	0.355	0.37	0.37	0.37	0.345	0.31	0.27	0.215	0.175	0.17	0.17	0.21	0.235	0.265	0.27	0.25	0.21	0.19	0.18	0.18	0.21	5.32
YN26	0.18	0.23	0.29	0.325	0.34	0.35	0.37	0.405	0.445	0.47	0.48	0.47	0.455	0.415	0.39	0.36	0.35	0.325	0.28	0.23	0.16	7.32
Weighted I(θ)	1.62	1.85	2.09	2.22	2.26	2.17	2.035	1.81	1.645	1.58	1.59	1.69	1.76	1.78	1.73	1.61	1.48	1.405	1.38	1.39	1.43	

Abbreviations: PN = physical domain; YN = psychological domain; SN = social domain; EN = environmental domain; the numbers were the item number.

Table 9 (continued) Item information for ranges of the latent trait

Item	-5	-4.5	-4	-3.5	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	Weighted I(θ)
Social domain																						
SN20	0.28	0.28	0.29	0.3	0.28	0.25	0.195	0.16	0.135	0.12	0.135	0.16	0.2	0.24	0.26	0.255	0.225	0.175	0.14	0.12	0.12	4.32
SN21	0.25	0.26	0.28	0.26	0.225	0.18	0.13	0.095	0.08	0.07	0.075	0.1	0.13	0.18	0.22	0.245	0.25	0.225	0.19	0.15	0.12	3.715
SN22	0.2	0.22	0.245	0.27	0.255	0.23	0.175	0.14	0.115	0.1	0.105	0.12	0.16	0.21	0.245	0.255	0.245	0.2	0.16	0.13	0.105	3.885
SN27	0.36	0.36	0.355	0.33	0.29	0.25	0.185	0.155	0.135	0.13	0.15	0.19	0.225	0.25	0.255	0.235	0.2	0.15	0.12	0.11	0.11	4.545
Weighted I(θ)	1.09	1.12	1.17	1.16	1.05	0.91	0.685	0.55	0.465	0.42	0.465	0.57	0.715	0.88	0.98	0.99	0.92	0.75	0.61	0.51	0.455	
Environmental domain																						
EN8	0.065	0.12	0.205	0.32	0.49	0.66	0.73	0.68	0.57	0.48	0.42	0.36	0.31	0.27	0.25	0.25	0.26	0.27	0.27	0.25	0.2	7.43
EN9	0.085	0.135	0.22	0.32	0.48	0.62	0.68	0.655	0.575	0.5	0.43	0.37	0.31	0.27	0.24	0.24	0.255	0.27	0.27	0.25	0.2	7.375
EN12	0.105	0.17	0.265	0.37	0.49	0.58	0.585	0.52	0.44	0.38	0.36	0.36	0.37	0.38	0.37	0.36	0.35	0.33	0.29	0.23	0.17	7.475
EN13	0.18	0.26	0.34	0.4	0.44	0.46	0.45	0.415	0.375	0.35	0.33	0.34	0.35	0.35	0.335	0.325	0.32	0.32	0.31	0.27	0.22	7.14
EN14	0.075	0.125	0.19	0.26	0.35	0.45	0.52	0.585	0.625	0.63	0.6	0.54	0.46	0.41	0.375	0.35	0.325	0.28	0.22	0.16	0.11	7.64
EN23	0.1	0.17	0.28	0.42	0.58	0.67	0.64	0.525	0.42	0.36	0.34	0.33	0.32	0.29	0.27	0.26	0.26	0.28	0.28	0.27	0.23	7.295
EN24	0.135	0.215	0.325	0.43	0.54	0.58	0.56	0.49	0.42	0.38	0.36	0.34	0.31	0.27	0.24	0.22	0.225	0.24	0.26	0.27	0.25	7.06
EN25	0.19	0.27	0.35	0.42	0.47	0.49	0.475	0.435	0.395	0.37	0.36	0.34	0.32	0.28	0.245	0.22	0.22	0.23	0.25	0.27	0.26	6.86
EN28	0.09	0.15	0.235	0.34	0.47	0.58	0.6	0.54	0.45	0.39	0.36	0.37	0.41	0.43	0.44	0.43	0.39	0.34	0.26	0.19	0.13	7.595
Weighted I(θ)	1.025	1.615	2.41	3.28	4.31	5.09	5.24	4.845	4.27	3.84	3.56	3.35	3.16	2.95	2.765	2.655	2.605	2.56	2.41	2.16	1.77	

Abbreviations: PN = physical domain; YN = psychological domain; SN = social domain; EN = environmental domain; the numbers were the item number.

(7) 鑑別力 (discriminant) 分析

接著從表二我們來看題目鑑別力參數，鑑別力參數大於1，表示鑑別力佳；鑑別力參數小於1，表示鑑別力不佳。以生理範疇來看，鑑別力較差的題目有PN3 (0.94)，PN16 (0.81)，其中以PN4 (0.59)的鑑別力最差。在生理範疇中，只有YN26 (0.55)的鑑別力最差，其餘的題目皆有良好的鑑別力。在社會範疇中，SN21 (0.92)及SN27 (0.79)鑑別力較差。而在環境範疇中，鑑別力較差的題目有EN9 (0.88)，EN14 (0.82)及EN28 (0.82)。

(8) 整合性分析 (integrated analysis)

受測者能力分佈

除了以上探討的題目特質外，IRT 也提供受測者能力的估計值，幫助我們了解老人的特質，分別以 WHOQOL-BREF 的四個範疇來看老人特質的分佈。在生理範疇中，能力最差的為-6.7，而最好的為 7.76。在能力的分佈中，範疇的平均分數為 0.38，這些題目難度對於社區老人是適中的，約 95%的老人能力集中在-1.52 ~ 3.28 之間。在心理範疇中，老人中能力最差的為-5.78，能力最好的為 8.79。在能力的分佈中，範疇的平均分數為 0.4，這些題目難度對於老人是適中的，約有 95%的老人能力集中在 -2.37 ~ 3.27。在社會範疇中，老人中能力最差的為-9.34，能力最好的為 10.98。在老人能力的分佈，範疇中人能力的平均為 1.02，可知這些題目難度對於老人是相對簡單的(mis-targeted)，約有 95%的老人能力集中在 -3.29 ~ 5.2。在環境會範疇中，老人中能力最差的為-5.25，能力最好的為 5.41，能力的平均為 0.36，這些題目難度對於老人是適中的，約有 95%的人能力集中在-1.61 ~ 2.38。

題目回答選項與能力的關係

表十為 WHOQOL-BREF 的四個範疇中題目的門檻值。所有的題目的

選項皆為 5 分法，有 4 個門檻值。從表十我們可以看到每一個題目中越過每一個門檻值的能力。以 PN10”您每天的生活有足夠的精力嗎?”為例，四個門檻值($\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ 及 $\hat{\beta}_4$)的數值分別為-3.73, -1.12, 1.48 及 5.3。第一個門檻值為-3.73, 表示-3.73 能力的人回答 PN10 時, 選擇”完全不足夠”及”少許足夠”這兩個選項的機會各為 0.5。接下來我們來看, 若人的能力為 3 他回答生理範疇的 7 個題目時, 最有可能的選擇類型 (response pattern) 為何? 在 PN3 這一題選擇 5”完全沒有妨礙”的機會最大。在 PN4 最有可能選擇 5”完全沒有需要”。PN10、PN15、PN16、PN17 及 PN18 此人的能力皆大於第三個門檻值, 但都沒有越過第四個門檻值, 所以此人最有可能的回答為 4。如果這些門檻值愈分散, 表示題目的選項愈能將不同能力的人分開, 但是不能單就著門檻值的分散性來看題目, 也必須考量閾值估計值的標準誤。估計的標準誤與選項中是否有夠多的人選擇有顯著的關係, 從圖十四心理範疇的 person-threshold map 來看, 可以發現老人的能力遠在第一個門檻值以上所以大多數的老人會選擇”2”以上的選項, 從表十我們可以發現第一個門檻值的標準誤也比較大。若是標準誤太大, 此估計值也是較無效的。

相較於傳統測量理論, 以 IRT 模式來分析最大的優勢是以心理計量理論模式為基礎的測量(model-based measurement)。利用數學模式將人的能力及題目特質放在同一個尺度, 使人的能力與題目特質可以相比較。從圖十五到圖十八我們將人能力的分佈與題目門檻值放在同一個尺度上。圖十五為生理範疇每 7 個题目的門檻值與老人生理功能能力的分佈, 除了”PN15 四處行走的能力”外, 由此相對的分佈我們可以發現老人的能力遠在第一個門檻值以上, 表示這些老人的能力已經越過第一個門檻值, 老人比較可能的回答為第二個選項以上的選項。

圖十六為心理範疇 6 個题目的四個門檻值與老人心理特質的分佈,

從圖中我們可以發現 YN10”生活有足夠的精力”及 YN11”接受自己的外表”，所有的老人能力遠在第一個門檻值以上，表示這些老人回答兩題時傾向回答第二個選項以上的選項。並且除了 YN26”常有負面的感受”這一題外，也發現老人的能力是比第四個門檻值來得低的，表示老人不容易回答最滿意的選項，老人回答的分佈集中在 2 到 4 之間。即

WHOQOL-BREF 心理範疇的題目對老人是適中的。

圖十七為社會範疇 4 個題目的四個門檻值與老人社會關係特質的分佈，從圖中老人在社會範疇的能力分佈不像在其他範疇的分佈一樣，那麼的對稱。能力的分佈大都落在能力比較好的區域。大多數老人的能力是比所有題目第一個門檻值來得高的，表示老人在社會範疇的題目較不會回答”非常不滿意”。在”SN21：滿意性生活”這一題，幾乎不會回答”非常滿意”。

圖十八為環境範疇 9 個題目的四個門檻值與老人環境範疇得分的分佈，所有老人的分佈約在-2 到 5 之間且其能力的分佈是相對稱的，題目間門檻值的分佈是較老人的能力分佈廣。除了 EN8”感到安全”、EN12”足夠的金錢應付所需”及 EN14”有機會從事休閒活動”外，老人的能力大都越過了第一個門檻值，即老人有較高的機會回答 2 以上的選項。相對的這些題目第一個門檻值比其他門檻值的誤差大一點（見表六）。

Table 10 Item thresholds for the WHOQOL-BREF domains

Item	$\hat{\beta}_1$ (s.e.)	$\hat{\beta}_2$ (s.e.)	$\hat{\beta}_3$ (s.e.)	$\hat{\beta}_4$ (s.e.)
Physical domain				
PN3. Pain and discomfort	-3.69 (0.31)	-1.83 (0.12)	-0.09 (0.07)	1.23 (0.08)
PN4. Dependence on medication or treatment	-3.78 (0.27)	-1.23 (0.09)	0.91 (0.07)	1.69 (0.09)
PN10. Energy and fatigue	-3.73 (0.26)	-1.12 (0.08)	1.48 (0.08)	5.3 (0.25)
PN15. Mobility	-1.56 (0.14)	-1.17 (0.09)	0.79 (0.07)	3.63 (0.14)
PN16. Sleep and rest	-3.86 (0.29)	-1.5 (0.09)	1.13 (0.07)	4.88 (0.21)
PN17. Activities of daily living	-3.81 (0.31)	-2 (0.11)	1.43 (0.07)	5.21 (0.24)
PN18. Working capacity	-3.46 (0.27)	-1.81 (0.1)	1.7 (0.08)	5.26 (0.25)
Psychological domain				
YN5. Enjoyment of life	-2.74 (0.15)	-1.25 (0.08)	2.37 (0.09)	6.78 (0.41)
YN6. Spirituality, religion and personal beliefs	-4.4 (0.24)	-1.83 (0.09)	2.06 (0.09)	5.64 (0.26)
YN7. Thinking, learning, memory & concentration	-3.5 (0.19)	-1.68 (0.09)	1.86 (0.08)	5.98 (0.29)
YN11. Body image and appearance	-6.74 (0.61)	-2.79 (0.12)	1.8 (0.08)	5.64 (0.25)
YN19. Self-satisfaction	-5.9 (0.49)	-3.45 (0.14)	1.21 (0.08)	5.68 (0.24)
YN26. Negative feeling	-4.83 (0.29)	-2.09 (0.11)	-0.23 (0.08)	2.41 (0.1)
Social domain				
SN20. Personal relationship	-7.18 (0.63)	-3.9 (0.16)	1.42 (0.09)	6.84 (0.19)
SN21. Sexual activity	-6.38 (0.47)	-2.61 (0.13)	3.99 (0.11)	9.66 (0.61)
SN22. Friends' support	-8.39 (1.04)	-4.06 (0.16)	1.75 (0.09)	7.6 (0.24)
SN27. Esteemed and respected	-5.63 (0.34)	-3.08 (0.13)	2.15 (0.09)	7.83 (0.26)
Environmental domain				
EN8. Physical safety and security	-2.19 (0.18)	-1.92 (0.1)	0.41 (0.07)	4.11 (0.17)
EN9. Physical environments	-3.13 (0.3)	-2.46 (0.13)	-0.36 (0.07)	3.44 (0.13)
EN12. Financial environment	-1.88 (0.13)	-0.94 (0.08)	2.01 (0.09)	4.58 (0.25)
EN13. Opportunities	-3.33 (0.24)	-1.48 (0.09)	1.51 (0.08)	4.44 (0.22)
EN14. Participation & support of leisure activities	-1.62 (0.1)	0.03 (0.07)	1.39 (0.08)	4.04 (0.19)
EN23. Home environment	-3.21 (0.34)	-2.89 (0.14)	0.16 (0.07)	3.71 (0.15)
EN24. Health & social care: availability & quality	-3.12 (0.26)	-2.09 (0.11)	0.64 (0.07)	4.6 (0.21)
EN25. Transportation	-3.67 (0.31)	-1.96 (0.1)	0.74 (0.07)	4.75 (0.22)
EN28. Eating food	-3.72 (0.42)	-2.89 (0.15)	0.16 (0.07)	2.14 (0.09)

Estimates were obtained from partial credit model by WINSTEP.

Abbreviations: PN = physical domain; YN = psychological domain; SN = social domain; EN = environmental domain; the numbers were the item number.

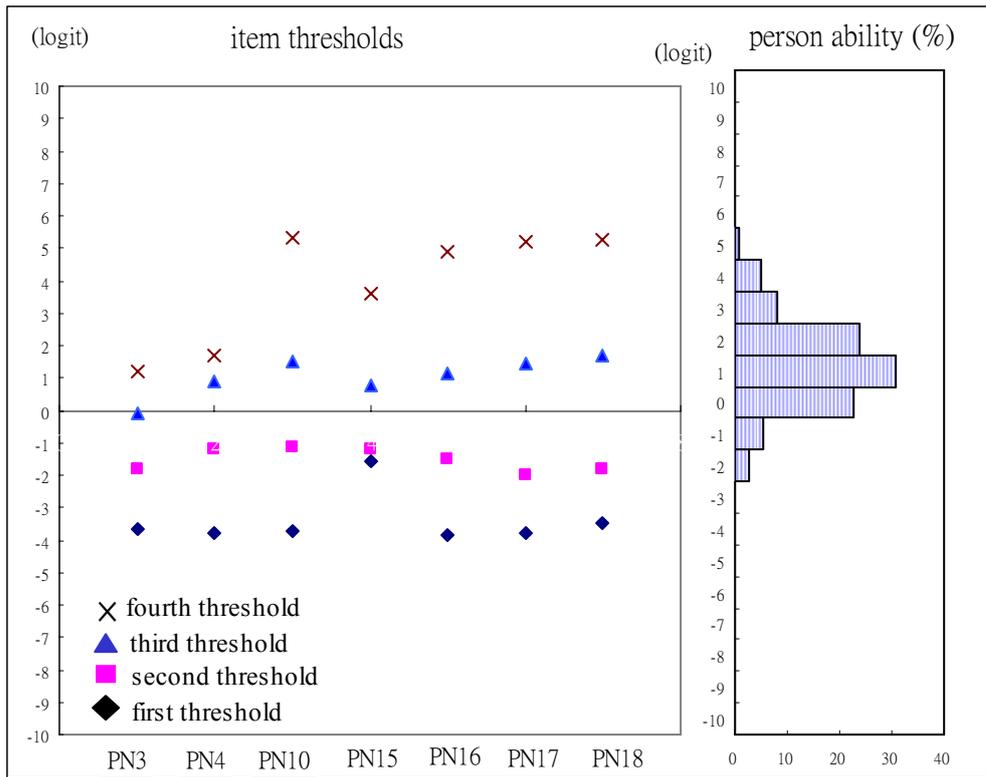


Figure 15. Person-threshold map for physical domain of WHOQOL-BREF

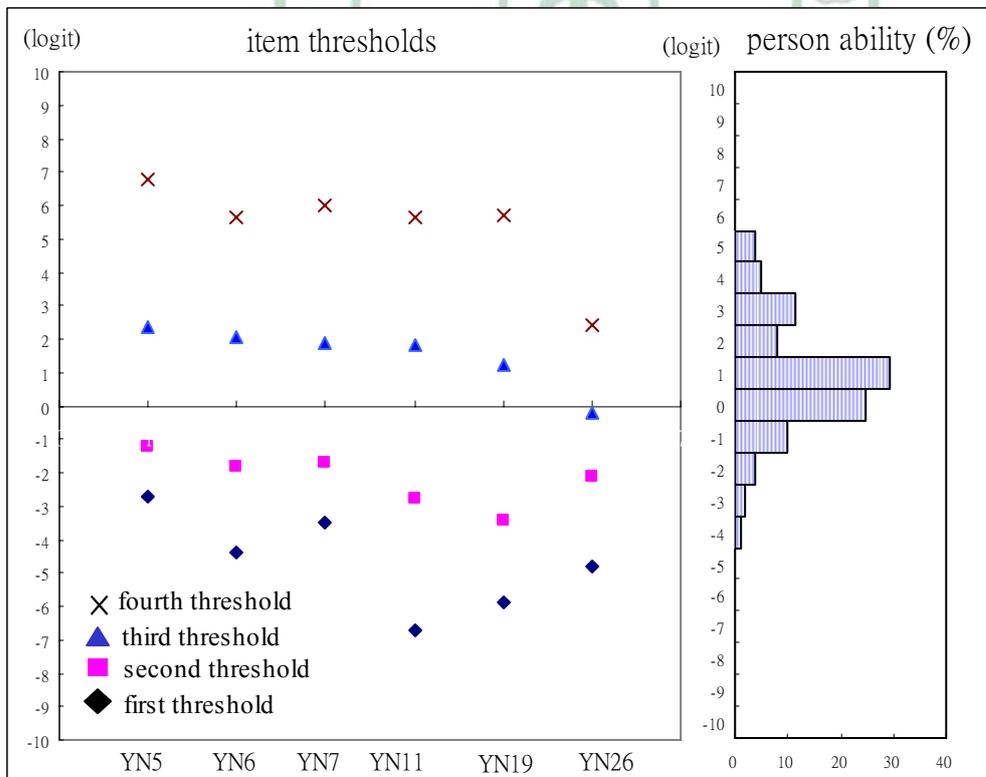


Figure 16. Person-threshold map for psychological domain of WHOQOL-BREF

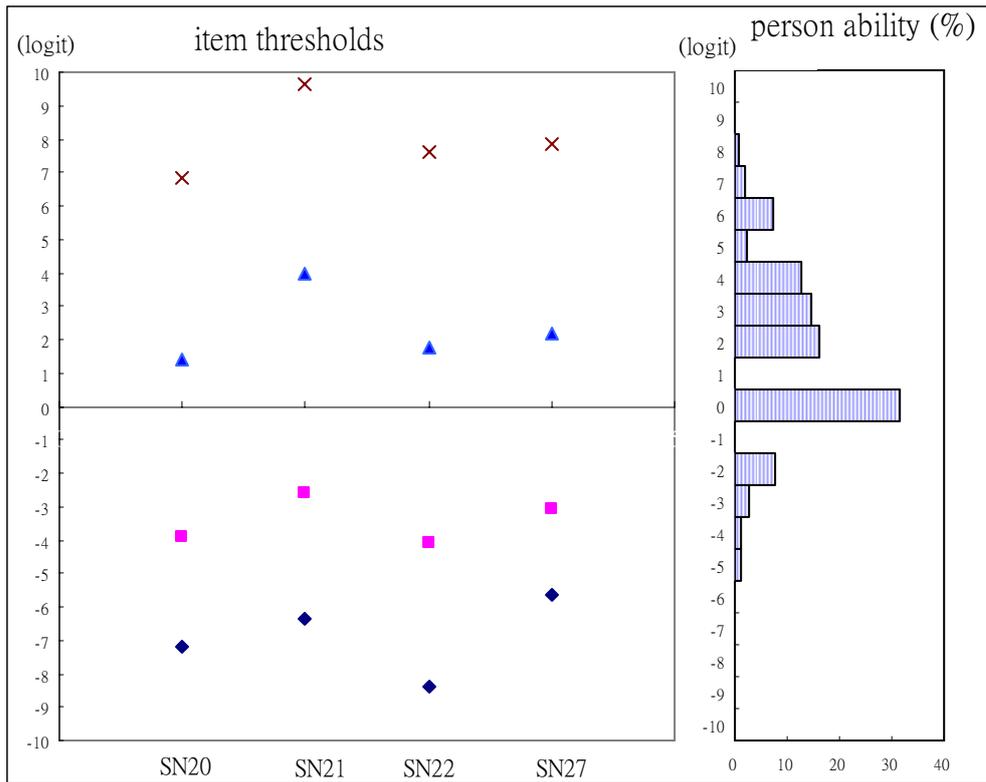


Figure 17. Person-threshold map for social domain of WHOQOL-BREF

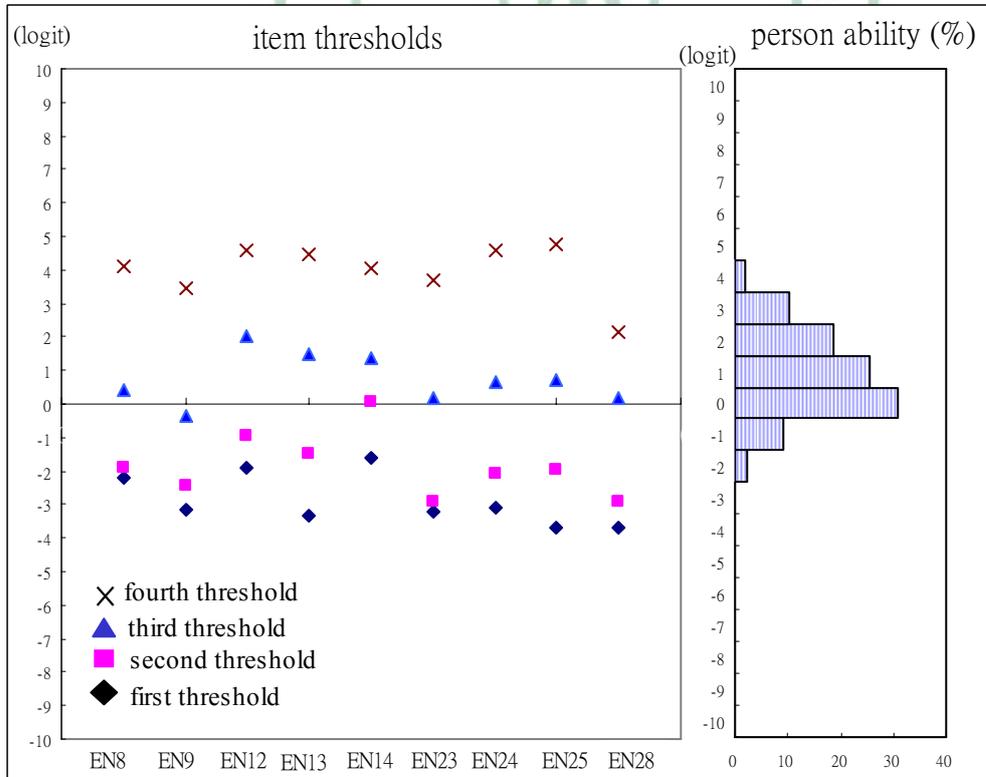


Figure 18. Person-threshold map for environmental domain of WHOQOL-BREF

第五章 討論

題目難度的分佈

Hwang 等利用傳統的心理計量評估 WHOQOL-BREF 在老人具有良好的信、效度並且是適用的¹³。但是這些都是以傳統的方法從範疇的層次來看 WHOQOL-BREF 在老人的適用性。本研究進一步利用項目反應理論從題目的角度作更深入的探討。本研究的結果證實 WHOQOL-BREF 大多數的題目都符合單一向度的假設。題目的平均難度介在+1~-1 之間，然而，WHOQOL-BREF 四個範疇的門檻值的分佈相較於社區老人族群 95% 的能力分佈是較廣的，表示題目能準確地估計大部分老人的能力，對於少數最好及最差的老人則較不準。

在題目的重覆性上，環境範疇的題目有明顯難度重覆性存在，日後若要修改問卷時可以刪減重覆性的題目，增加題目時則選擇在有較大間隔(gap)之處，以增加量尺上題目的均勻性。在題目難度的適中性上，社會範疇的題目對老人而言較易得到高分，而其他範疇的題目難度則較為適中。

問卷信度的檢驗

Hwang 等利用傳統心理計量方法來評估 WHOQOL-BREF 在新社鄉社區老人的適用性¹³，生理、心理、社會及環境範疇的信度各範疇 Cronbach's alpha 值分別為 0.80、0.81、0.73 及 0.80 相同，本研究中 Rasch 模式分析所得結果中，受測者潛在特質的信度指標與傳統的測量信度的 Cronbach's alpha 相似，受測者潛在特質的信度指標在生理、心理、社會及環境範疇分別為 0.76、0.77、0.68 及 0.78。兩個研究所得結果次序一致，社會範疇的信度指標均最差。

問卷效度的檢驗

Hwang 等利用傳統心理計量方法中各範疇分數由加總計算所得之分數(summated score)，以老人憂鬱症分數、認知分數等來檢驗 WHOQOL-BREF 的區辨效度，發現各範疇均有顯著的區辨力¹³，在表十中，我們以 Rasch 分析之得分進行老人憂鬱程度與認知程度之區辨力分析並與 Hwang¹³ 的結果作比較，F 檢定的結果發現 WHOQOL-BREF 各範疇均有顯著的區辨力，更進一步我們用相對有效性指標來評估何者更能敏感地區辨出老人憂鬱程度與認知程度，在憂鬱程度方面 Rasch 得分之相對有效性均優於 Hwang¹³ 的結果，在認知分數方面 Rasch 得分之相對有效性只有在社會範疇上優於 Hwang¹³ 的結果。

傳統方法假設的探討

傳統的方法直接將分數等級加總有兩個假設，一為每一個題目的貢獻相同，另一為題目的選項間具有相同的間距³⁰。題目的貢獻表示題目所提供的訊息，即測量的精確度。當問卷在每一個能力下測量的精確度都相同時，表示在每個能力的區間下都有題目可以來估計人的能力。影響測量的精確度的因素很多，題目難度分佈不均勻、題目難度間有較大的間距或有 DIF 存在等都會影響測量的誤差⁴⁰。本研究分析中，社會範疇有顯著的 DIF(圖七、圖十一範疇選項間並未具有相同的間距(圖十五~圖十八)，環境範疇中有較明顯的題目難度重覆性存在(表四與圖四)，因此 Rasch 分析所得分數應可較傳統加總分數來得準確，尤其是在社會範疇。此與表十一中 Rasch 分數較能區辨老人憂鬱程度之結果一致，但與表十一中老人認知程度之區辨結果不完全一致，或可解釋為認知程度本身並非一個很好的效標。

Table 11 Discriminant validity analysis: means of domain or facet score by characteristic in classical test theory and item response theory

Domains	Depression				Cognitive impairment			
	No Mean(SD)	Yes Mean(SD)	F-value	Relative validity	No Mean(SD)	Yes Mean(SD)	F-value	Relative validity
Physical	n=1135	n=63			n=980	N=220		
Rasch	0.61 (1.46)	-0.25 (0.95)	45.43***	1.22	0.74 (1.36)	-0.25 (0.95)	69.72***	0.92
Sum ^a	13.39 (2.25)	12.03 (1.69)	37.21***		13.6 (2.07)	12 (2.53)	76.04***	
Psychological								
Rasch	0.44 (1.88)	-0.26 (1.39)	14.59**	1.23	0.65 (1.73)	-0.26 (1.39)	83.91***	0.97
Sum ^a	12.8 (2.15)	12.03 (1.72)	11.90**		13.06 (1.96)	11.44 (2.4)	86.86***	
Social								
Rasch	1.07 (2.68)	0.18 (2.07)	10.63**	1.43	1.25 (2.58)	0.18 (2.07)	41.47***	1.22
Sum ^a	13.18 (1.94)	12.63 (1.54)	7.45**		13.32 (1.83)	12.42 (2.13)	33.87***	
Environmental								
Rasch	0.4 (1.31)	-0.31 (0.93)	32.83***	1.16	0.49 (1.23)	-0.31 (0.93)	42.51***	0.87
Sum ^a	13.14 (2)	12.07 (1.52)	28.41***		13.29 (1.85)	12.13 (2.29)	48.72***	

^a Based on Hwang's result¹³, represent one kind of Summated score

Based on Student's t test. *: P<0.05; **p<0.001; ***p<0.0001

Relative validity = F value (Rasch) / F value (Summed)

另一個假設為每個選項的間距都是相同的。惟有當题目的間距都是相同的時候才可以相加受測者反應數值作為受測者的能力指標。若是選項的間距不同，那就只能算為序位資料不能將其加總。但是一般在應用上為了方便性常常不去評估題目是否具有這些特質，就假設都符合了。但是在 Rasch 分析中對於每一個題目選項的門檻值進行分析，從圖十五到圖十八中我們發現每個門檻值的間距不都是相同的。因故以傳統方法中對題目生硬的假設，可能對範疇分數的估計會有較大的誤差。

鑑別力與訊息函數

Rasch 系列中的鑑別力參數是事後估計得到的，與參數系列模式中鑑別參數的估計是有所不同的。Rasch 模式會先預設鑑別參數為 1，估完受

測者能力及題目參數後，再用已估出的參數，事後估計鑑別參數⁴⁶。鑑別參數的可能範圍雖然是 $0 \sim \infty$ ，但 Rasch 所得的鑑別參數大都是在 0~2 之間。因為是事後估計的，主要反應的是原模式無法估計的部分，所以從表二中我們發現鑑別力與 misfit 統計量有一致性存在，亦即高的鑑別力與低的 misfit MNSQ 數值相一致，低的鑑別力與高的 misfit MNSQ 一致。

在參數系列模式中可發現訊息函數與鑑別力參數有很大的相關性^{3, 4}，但在本研究中卻發現在生理範疇中鑑別力最差的題目 PN4 其總訊息量反而較其他鑑別力較好的題目來得高；此外，在項目反應理論由訊息函數的公式可知當人的能力與題目難度相同時所提供的訊息量最大，但 Rasch 模式並未有此特質，例如 PN15 難度估計值為 0.42(表二)，但其訊息量最高是出現在 -2~-1 之間。

鑑別力與 infit 統計量有一致性存在，但與訊息函數結果不一致，與 Rasch 模式特質相符。Rasch 模式的訊息量的應該要配合適合度來看，若該題不適合(infit > 1.4)則訊息量的準確性會有很大的偏差。

試題差別功能 (DIF)

在評估題目特質是否會因為性別的不同，而有不同的題目特質時，心理範疇的 YN6 及 YN19 有輕微的 DIF，但在社會範疇中 SN21(對性生活的滿意)及 SN22(對朋友支持的滿意)則有顯著的 DIF 存在。SN21 在男性是最難的題目，而在女性則是容易的題目，有趣的是題目特質在不同性別中是不相同，究其原因可能是因為在傳統的中國文化下，男性偏向追求壯陽；而女性則是可能因生理狀態的改變對於性的需求降低，對性生活較容易滿意。然而，針對 SN21，在 Wang⁴⁷ 等分析國健局 20~65 歲的一般族群發現在較年輕(20-29 歲)與較年長(大於 60 歲)這兩個次族群間有 DIF 的情形，但在性別上則沒有發現。為確認性生活在性別間的 DIF 是否會有城鄉間的不同存在，故我們選取國健局中台北市的樣本，針對

社會範疇進行 DIF 分析，發現性生活這題確實在性別及年齡有 DIF 存在。以不論在都市或是鄉下地區 SN21”性生活”在性別都有 DIF。

接著在本研究中，SN22”朋友支持”在男性是較容易的題目，但在女性卻是最難的題目，造成題目特質差異存在的原因可能為本研究族群為較鄉下地區的社區老人，在傳統”男主外，女主內”的情形下，男性在外常有社交活動，但女性則是在家中相夫教子，較缺少社交活動所致。

針對社會範疇將分年齡層來看性別間的 DIF，以及分性別來評估 DIF，我們發現對性生活的滿意度與生理功能改變有很明顯的關係，表十三我們可以發現在年輕老人（65~75 歲）對性生活在性別中的差異，但是在老老人中性生活在男性與女性間都是最簡單的題目。在年輕老人中我們可以發現 SN20”人際關係”與 SN22”朋友支持”在男性是相對簡單的題目，而在女性則是相對簡單的題目。但是在老老人中我們發現不管在男性或是女性 SN22”朋友支持”都是最難的題目。朋友支持的概念較偏向與同年齡的人之互動，而社會關係則是與人的互動。到了 75 歲以後，身邊的朋友相繼的離開人世，這可能是老老人在朋友支持上難以得到滿足的原因。圖十四為分性別來看年輕老人與老老人之間的 DIF，不論在男性或是女性都可以發現 SN21”性生活”在老老人都是最容易的題目，而 SN22”朋友支持”在老老人都是最難的題目。

我們發現社會範疇的題目在性別及年齡間有明顯的 DIF 存在，可能在測量上造成的影響為題目在不同的性別（或年齡）間貢獻不同。當題目在性別（或年齡）間的貢獻不同時，卻放在同一個基礎上來比較是不適當。總之，DIF 分析可迅速地找出不同族群間題目特質的差異性，除了檢視問卷外，也可以幫助我們進一步對於不同研究族群特質的瞭解與區隔。

第六章 結論與建議

第一節 結論

以 Rasch 分析來評估 WHOQOL-BREF 應用在老人的心理計量特質時，乃是針對各個題目來探討。從我們的結果中發現 WHOQOL-BREF 各範疇的題目都符合 Rasch 模式單一向度的假設。

每個範疇題目門檻值的分布分別為-3.86~5.3、-6.74~6.78、-8.39~9.66 及-3.72~4.75，而大部分老人能力在生理、心理、社會及環境範疇分別為-1.52~3.28、-2.37~3.27、-3.29~5.2 及-1.62~2.38，可見。四個範疇最難的題目為生理範疇”足夠的精力”(0.48)、心理範疇”享受生活”(1.29)、社會範疇”性生活”(1.17)及環境範疇”有機會從事休閒活動”(0.96)。最容易的題目則分別為為生理範疇”身體疼痛”(-1.1)、心理範疇”享受生活”(1.29)、社會範疇”朋友支持”(-0.77)及環境範疇”想吃的食物的都吃的到”(-1.08)。

然而，在環境範疇的題目難度有較多重覆性的題目。在題目順序方面，我們與 Wang⁴⁷的題目順序相比較，發現在生理範疇及心理範疇有少數題目的順序不同，然而，在社會及環境範疇的題目順序則是相同的。

受測者能力估計值的平均在 WHOQOL-BREF 中生理、心理、社會及環境範疇分別為 0.57、0.44、1.02 及 0.36。除了社會範疇的題目對於老人是相對簡單的題目外，其他範疇的題目對於這些老人是適中的。

除了心理範疇的 YN11”接受自己的外表”在年齡有明顯的 DIF，以及社會範疇的 SN21”性生活”及 SN22”朋友支持”兩題在性別及年齡有明顯的 DIF 存在，其他的範疇都沒有顯著的 DIF 存在。根據區別性指標，顯示各範疇皆能將老人的能力區分為三組，例如將生活品質分為高、中及低三組。

利用傳統的方法只能從加總的分數來看人的能力，也沒有考慮到題目

的特質，題目難度分佈不均勻、題目難度間有較大的間距或有 DIF 存在等都會影響測量的誤差，加總的分數存在著許多的問題。然而，項目反應理論不僅有其嚴謹的理論架構，利用數學模式將人的能力及題目特質整合在一起，它同時可以產生許多指標如 infit、outfit、DIF、separation index、person reliability、information function 等，使我們可據以檢視問卷的心理計量特質，同時 person-threshold map 亦全面提供對題目難度及受測者能力的整合，除了方便於檢視問卷特質外，亦大大增加我們對研究族群特質的瞭解與區隔。

WHOQOL-BREF 在生理、心理、社會及環境範疇中的信度指標分別為 0.76、0.77、0.68 及 0.78，除了社會範疇外，其他的範疇皆顯示良好的內部一致性。整體而言，WHOQOL-BREF 應用於社區老人生活品質的測量有不錯的心理計量特質。本研究利用項目反應理論的 Rasch 模式分析，從結果中我們更深入發現 WHOQOL-BREF 的一些特質及可以在未來改進的地方，例如在兩個能力尺度較極端的部分增加題目、考慮刪除有難度重覆性的題目、針對 SN21”性生活”及 SN22”朋友支持”兩題在不同性別有明顯 DIF 因而影響測量品質的部份提出討論與修改建議。

第二節 研究限制

1. 新社鄉是一個鄉下的地方，保留許多傳統文化的色彩，居民單純且保守，這可能是造成社會範疇中 SN22”朋友支持”在不同的性別下明顯 DIF 的原因。

2. 在本研究中排除掉 59 位在研究訪視期間住院或是在家臥床的老人，使得本研究中缺少生活品質較差的老人，若能將能力較差的老人加到分析中可以更準確且全面性地評估社區老人的生活品質。

第三節 應用與建議

1. 由於本研究在社區中取樣，相對於在醫院取樣之研究，本研究之老人較為健康，故許多題目相對較為簡單，以致於在難度的估計上，較簡單的題目的誤差較大，由於當初收案時，臥病在床或住院之老人並未收案，建議增加此部份老人之資料，使得問卷整體的估計能夠更完整，以期能更準確的評估 WHOQOL-BREF 各種難度題目之特質。

2. 從本研究與國健局 65 歲以下族群的比較下，發現 65 歲以上老人和一般族群的生活品質的特質是不盡相同的，因老人乃是社會中需要更多關心的一個大族群。日後可以針對這兩族群間的差異進行研究，以期更深入瞭解老人不同於年輕族群之特質，能更有效地提供老人合適的社會及醫療服務的幫助。

3. 未來若要對 WHOQOL-BREF 進行修改，若要新增題目，則建議應增加題目難度的範圍，亦即加入在極端好及極端差兩端的題目，以提供全面性精確的測量。

4. 在生活品質的評估中，太長的問卷常常會使受測者失去耐心，藉由 IRT 的發展可建立生活品質題庫，算出每題的訊息，結合電腦的應用，可據此發展電腦適性試驗(computerized adapted test)提供給每個受測者最適合且能夠提供最大訊息的題目，使研究者能更夠用最少的題目準確地估計受測者的能力。

參考文獻

1. HAYS RD, MORALES LS, REISE SP. Item response theory and health outcomes measurement in the 21st century. *Med Care* 2000;38:II28-42.
2. CELLA D, CHANG CH. A discussion of item response theory and its applications in health status assessment. *Med Care* 2000;38:II66-72.
3. EMBRETSON SE, REISE SP. Item response theory for psychologists. New Jersey: Lawrence Erlbaum Associates, 2000.
4. HAMBELTON RK, SWAMINATHAN H, ROGERS HJ. Fundamentals of Item Response Theory. London: SAGE, 1991.
5. CONRAD KJ, SMITH EV, JR. International conference on objective measurement: applications of Rasch analysis in health care. *Med Care* 2004;42:II-6.
6. FITZPATRICK R, NORQUIST JM, DAWSON J, JENKINSON C. Rasch scoring of outcomes of total hip replacement. *J Clin Epidemiol* 2003;56:68-74.
7. FRALEY RC, WALLER NG, BRENNAN KA. An item response theory analysis of self-report measures of adult attachment. *J Pers Soc Psychol* 2000;78:350-65.
8. DUNCAN PW, BODE RK, MIN LAI S, PERERA S. Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Arch Phys Med Rehabil* 2003;84:950-63.
9. NAUMANN VJ, BYRNE GJA. WHOQOL-BREF as a measure of quality of life in older patients with depression. *International Psychogeriatrics* 2004;16:159-173.
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, 王榮德. 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;18:262-70.
13. HWANG HF, LIANG WM, CHIU YN, LIN MR. Suitability of the WHOQOL-BREF for community-dwelling older people in Taiwan. *Age Ageing* 2003;32:593-600.
14. MCHORNEY CA, COHEN AS. Equating health status measures with item response theory: illustrations with functional status items. *Med Care* 2000;38:II43-59.
15. ANDRICH D. Controversy and the Rasch model: a characteristic of incompatible paradigms? *Med Care* 2004;42:I7-16.
16. MALLINSON T, STELMACK J, VELOZO C. A comparison of the separation ratio and coefficient alpha in the creation of minimum item sets. *Med Care* 2004;42:II7-24.
17. NORQUIST JM, FITZPATRICK R, DAWSON J, JENKINSON C. Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care* 2004;42:I25-36.
18. TENNANT A, PENTA M, TESIO L, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Med Care* 2004;42:I37-48.
19. HALEY SM, COSTER WJ, ANDRES PL, et al. Activity outcome measurement for postacute care. *Med Care* 2004;42:I49-61.
20. COSTER WJ, HALEY SM, ANDRES PL, LUDLOW LH, BOND TL, NI PS. Refining the conceptual basis for rehabilitation outcome measurement: personal care and instrumental activities domain. *Med Care* 2004;42:I62-72.
21. NIJSTEN T, UNAEZE J, STERN RS. Refinement and reduction of the Impact of Psoriasis Questionnaire: classical test theory vs. Rasch analysis. *Br J Dermatol* 2006;154:692-700.

22. 行政院衛生署統計處. 民國 93 年衛生統計重要指標. In: 行政院衛生署統計處, ed.: 行政院衛生署, 2006.
23. LAWTON MP, WINDLEY PG, T.O. B. Aging and the Environment: Theoretical Approaches. New York, 1982.
24. MCHORNEY CA, WARE JE, JR., LU JF, SHERBOURNE CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994;32:40-66.
25. BINDMAN AB, KEANE D, LURIE N. Measuring health changes among severely ill patients. The floor phenomenon. *Med Care* 1990;28:1142-52.
26. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. *Psychol Med* 1998;28:551-8.
27. WHO. WHO Constitution. In: Organization WH, ed. Geneva: WHO, 1948.
28. FAU HI, CARR AJ. Measuring quality of life: Using quality of life measures in the clinical. *Bmj* 2001;322:1297-300.
29. TSAI S-Y, CHI L-Y, LEE L-S, CHOU P. Health-related quality of life among urban, rural, and island community. *J Formos Med Assoc* 2004;103:196-204.
30. BOND TG, FOX CM. Applying the Rasch model: Fundamental measurement in the human sciences. New Jersey: Lawrence Erlbaum Associates, Inc., 2001.
31. STAQUET MJ, HAYS RD, FAYERS PM. Quality of life assessment in clinical trials: Methods and practice. In: Hays RD, ed. Item response theory model. New York: Oxford Univeristy Press, 1999.
32. REISE SP, WIDAMAN KF, PUGH RH. Confirmatory factor analysis and item repsonse theory: two approaches for exploring measurement invariance. . *Psychological Bulletin* 1993;114:552-66.
33. HALEY SM, MCHORNEY CA, WARE JE. Evaluatione of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *Journal of Clinical Epidemiology* 1994;47:671-84.
34. RASCH G. An item analysis which takes individual difference into account. *British Journal of Mathematical and Statisitcal Psychology* 1996;19:49-57.
35. UTTARO T, LEHMAN A. Graded response modeling of the Quality of Life Interview. *Evaluation and Program Planning* 1999;22:41-52.
36. WIKIPEDIA. Georg RaschMathematics, 2006.
37. EVERETT S, RICHARD S. Introduction to Rasch Measurement: Theory, Models and Applications An overview of the family of Rasch measurement models. Sacramento, CA: Jam Press, 2004.
38. ANDERSEN EB. A goodness of fit for the Rasch model. *Psychometrika* 1973;38:123-40.
39. DOUGLAS G, WRIGHT BD. The two-category model for objective measurement Chicago: University of Chicago, MESA Psychometric Laboratory, 1986.
40. GARRATT AM. Rasch analysis of the Roland disability questionnaire. *Spine* 2003;28:79-84.
41. PRIETO L, ALONSO J, LAMARCA R. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes* 2003;1:27.
42. 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.
43. HART DL, MIODUSKI JE, STRATFORD PW. Simulated computerized adaptive tests for measuring functional status were efficient with good discriminant validity in patients with hip, knee, or foot/ankle impairments. *J Clin Epidemiol* 2005;58:629-38.

44. JENKINSON C, NORQUIST JM, FITZPATRICK R. Deriving summary indices of health status from the Amyotrophic Lateral Sclerosis Assessment Questionnaires (ALSAQ-40 and ALSAQ-5). *J Neurol Neurosurg Psychiatry* 2003;74:242-5.
45. LAI JG, CELLA DG, CHANG CHG, BODE RKG, HEINEMANN AWG. Item banking to improve, shorten and computerize self-reported fatigue: An illustration of steps to create a core item bank from the FACIT-Fatigue Scale. *Quality of Life Research* 2003;12:485-501.
46. LINACRE JM. *A user's Guide to WINSTEPS: Rasch-Model Computer Programs*. Chicago: Winsteps company, 2006.
47. WANG WC, YAO G, TSAI YJ, WANG JD, HSIEH CL. Validating, Improving Reliability, and Estimating Correlation of the Four Subscales in the WHOQOL-BREF using Multidimensional Rasch Analysis. *Qual Life Res* 2006;15:607-20.
48. 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.
49. GEORGE LK, LANDERMAN R, D.G. B, ANTHONY C. Cognitive impairment. In: Robins LN, Regier DA, eds. *Psychometric Disorders in America: The Epidemiological Catchment Area Study*. New York: Free Press, 1991.
50. WRIGHT BD, STONE MH. *Best test design: Rasch measurement*. Chicago: MESA Pr., 1979.
51. WRIGHT BD, O.N. M. *Rating scale analysis: Rasch measurement*. Chicago: MESA Pr., 1982.
52. WOLFE F, MICHAUD K, PINCUS T. Development and validation of the health assessment questionnaire II: a revised version of the health assessment questionnaire. *Arthritis Rheum* 2004;50:3296-305.

附錄一 世界衛生組織生活品質問卷（台灣簡明版）

問卷說明：

請選出最適合您在最近兩星期內對自己健康的滿意程度，如果您很滿意自己的健康，就在「極滿意」前的□內打「√」。請仔細閱讀每個題目，並評估您自己的感覺，然後就每一個題目選出最適合您的答案。謝謝您的協助！

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. 您滿意醫療保健服務的方便程度嗎？
- 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
- 完全沒有 有一點有 中等程度有 很有 極有
 27. 您覺得自己有面子或被尊重嗎？
- 從來沒有 不常有 一半有一半沒有 很常有 一直都有
 28. 您想吃的食物通常都能吃到嗎？

