

薄喬萍 著

Reliability and Validity

信度與效度



東華書局

信度與效度

薄喬萍 著

Reliability
and
validity

ISBN
957-05-1111-1

東華書局

國家圖書館出版品預行編目資料

信度與效度 / 薄喬萍著. -- 初版. -- 臺北市 :
臺灣東華, 民 99.10
面 ; 公分

ISBN 978-957-483-636-9 (平裝)

1. 質性研究 2. 信度 3. 效度

501.2

99019855



版權所有 · 翻印必究

中華民國九十九年十月初版

信度與效度

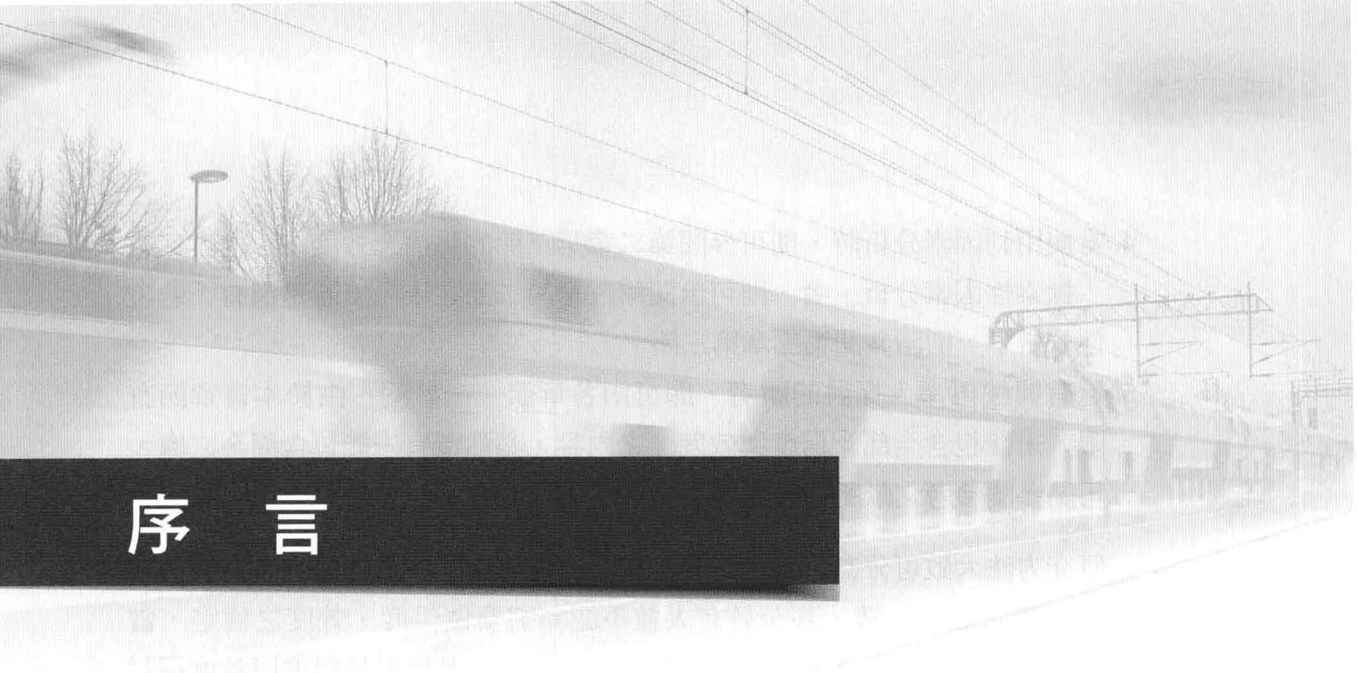
定價 新臺幣參佰伍拾元整

(外埠酌加運費匯費)

著 者 薄 喬 萍
發行人 卓 劉 慶 弟
出版者 臺灣東華書局股份有限公司
臺北市重慶南路一段一四七號三樓
電話：(02) 2311-4027
傳真：(02) 2311-6615
郵撥：0 0 0 6 4 8 1 3
網址：http://www.tunghua.com.tw

印刷者 鴻展彩色印刷股份有限公司

行政院新聞局登記證 局版臺業字第零柒貳伍號



序 言

凡是以問卷作為論文工具者，對於所設計的問題必須先經過信度與效度之檢測，檢測合格者，這份問卷才算有效、才能夠正式地發出去，實施意見調查。這個道理很簡單，大家都知道。

然而，當要執行檢測的時候，很多研究者卻不真正了解信度和效度的意義和做法，往往都是把測驗所收回的問卷分數，輸入電腦中的某種軟體，此時電腦報表會告訴我們 Cronbach's α 之值，也許很多人還不很清楚，此一 α 值究竟是如何得出來的。本書為了使初學者了解此一 α 值之由來，特以簡單明瞭的 EXCEL 舉例說明，淺顯易懂。以筆者之經驗，無論以前是否學過「統計學」，30 分鐘之內，每個人都學會如何計算 α 係數。

由於「信度」與「效度」的應用非常廣泛，各家學說並不齊一，很多內容並不容易彙整於章節之內，但是為了讀者能有更多的參考，本書將許多不易收錄在正文的內容，也設法出現在習題內。以筆者之設計，習題之作用可以視之為「Q & A」，因此，本書的習題也形成了另一種章節。

本書所包含的內容很多，對於只想學習如何使用信度、效度檢測的初學者而言，筆者建議可按以下順序閱讀較為有效：

1. 預習時，先做第一章第五節之「項目分析」。
2. 一般的內部一致性檢定，以第一章的 Cronbach's α 檢測為主。若是熟悉 SPSS 之讀者，可直接以 SPSS 評估。否則，可使用每部電腦都會配備的 EXCEL 軟體實施。信度的參考範圍請參閱第一章第三節。
3. 簡單的效度分析，通常都是以「內容效度」為主，此時可參閱第二章第三節。

4. 當使用到因素分析時，則可參閱第二章第三節的「構念效度」。內容只有「探索性因素分析」者，則可參閱第三章第二節之十一；若是還有「驗證性分析」，則請參閱第三章第三節。
5. 還有興趣再深入探討的讀者，即可由各章節一一參閱。由於本書章節有限，未能包含全部「信度與效度」之內容，必然還有未能包含周全之處，再請參閱參考文獻所列之叢書。

研究方法大致可分為「量化研究」和「質性研究」。質性研究者對於信度與效度的看法並不一致，甚至於有人並不認為需要做信度、效度之檢定，當然，這是見仁見智各有立場，筆者不擬評論。然而，基於對於信度與效度探討之立場，本書也特於第四章，撰寫了關於質性研究的信度與效度，希望藉此機會能夠對於質性研究的讀者們，也能提供一份心力與貢獻。

本書之撰寫前後費時幾近兩年，當然，一方面由於筆者才疏學淺，另一方面也是由於本書之題材較為分散，一時難以有效地蒐集足夠的資料所致。編撰之中，難免有掛一漏萬之可能，在此謹向讀者們致歉，也期望先進們不吝賜教為盼。

薄喬萍 謹識

九十九年9月於義守大學工管學系

CONTENTS

序 言

iii

1 Chapter

信 度

1

第一節 信度之意義

1

第二節 信度之內涵

6

▲ 一、從受試者內在的變異分析

6

▲ 二、從眾多受試者相互之間的變異加以分析

9

第三節 信度之種類

14

▲ 一、再測信度

14

▲ 二、複本信度

16

▲ 三、折半信度

18

▲ 四、庫-理信度

21

▲ 五、Cronbach's α 係數

23

▲ 六、評分者互評信度

31

▲ 七、訪問的信度

31

▲ 八、觀察的信度

32

▲ 九、測量標準誤

33

第四節 影響信度之因素

37

第五節 項目分析

41

▲ 一、項目分析之原理

41

▲ 二、四類、七項準則之檢定

42

2 效 度 63 Chapter

第一節	效度之意義	64
第二節	效度之評量方法	64
第三節	效度之種類	65
	一、內容效度	66
	二、效標關聯效度	83
	三、構念效度	90
	四、內在效度與外在效度	106
	五、觀察的效度	108
	六、其他的效度	110
第四節	信度與效度之關係	110

3 因素分析 117 Chapter

第一節	因素分析的基本觀念	118
第二節	探索性因素分析	119
	一、理論基礎	119
	二、主成份分析	119
	三、因素分析的基本原理	130
	四、因素分析實施之步驟	133
	五、共同性 h_i^2 之估計	136
	六、共同因素的萃取	140
	七、共同因素數目之決定	141

八、特徵值與特徵向量之內涵	143
九、因素軸之旋轉	146
十、模式適合性之評估	157
十一、統計驗證	163
第三節 驗證性因素分析	167
一、結構方程式	167
二、驗證性因素分析之理論	171
三、信度與效度	173
4 Chapter 質性研究之信度與效度	185
第一節 信 度	187
第二節 效 度	189
第三節 質性研究中三角資料檢測法	194
一、三角資料檢測法原理	194
二、三角資料檢測法	196
參考文獻	201
習題解答	205
索 引	245



信 度

第一節 信度之意義

信度 (Reliability) 即可靠性、可信的程度，若是應用在一份測驗上，是指這份測驗的一致性 (Consistency) 或是穩定性 (Stability)。

所謂的一致性，是指一份測驗問卷中，對於同一件問題內容，各試題之間的涵義是否能夠相互符合、前後是否能夠連貫一致；對於穩定性而言，則是指此份問卷若是施測多次，其測驗結果是否還能保持不變？一項研究或是研究所使用的測量工具，反覆不斷地用來實驗或是驗證，如果所得的結果不會改變，就代表這項研究或是測量的工具具有良好的「信度」。

由此可知，並不是任何的測驗皆有良好的信度，信度的產生與提升，是需要刻意地努力和修訂的。以上這些問題，不論是討論各問題間的內容是否相互符合，或是討論幾次測驗的結果，都不致有太大的變異，信度在這些問題上的討論，都是在探測「相對程度上的差異」，因為「一致」或是變異都是形容詞，或是以相對的比較結果，而不是以絕對的數字表現，是有信度或是沒有信度。

在實際的應用上，研究者要先把觀念性的問題，轉換成「操作性定義」的

實證層次，由於不是十分容易實施，因而經常導致問卷的題意不明，這就產生了「信度」上的問題。

在測驗或是問卷的調查，由於是以隨機抽樣的進行，因此每次的衡量結果或多或少總是會有些變動，這種變動是否可以諒解或允許，就要看這份問卷或測驗是否具有相當程度的「信度」。

例如，如果用一台「體重計」測量某人的體重，昨天量的結果是 70 公斤，而今天再量，卻量得 65 公斤，除非此人在一天內暴減了 5 公斤，否則斷不致此，顯然是測量工具的「穩定度」發生了問題，也就是「信度」不佳。

本書第二章將討論到「效度」的問題，再以這個「體重計」來說明：若是想要用此一「體重計」來測量某人的身高，當然是用錯了測量的工具，這種情形就是「效度」不佳的問題；也就是說，用錯了測量的工具，就會產生「效度」上偏差的問題。

編製一份測驗或是問卷並不困難，但是編成了這份問卷所測驗出的答案，是否真正是研究者所需要的？這就是研究「信度」所應具備的無扭曲性 (Lack of Distortion)，也就是說，這份問卷的文字部份，不致產生令人誤解或是看不懂的情形。談到此一論點，往往會令人對於「信度」與「效度」產生混淆。其實，信度並不討論所衡量的「真實性」，但是，有時候也會考慮到所衡量的「正確性」，此處所說的正確性，與以後談到「效度」所在意的正確性，其切入的角度並不相同。「信度」所關心的是測量工具的「正確性」，而不管其衡量的內容；若是需要知道這個測量工具，真正可以測到所需要知道的內容，這就是「效度」所要討論的問題。

「信度」也是在討論測量工具的「變異」問題。測量的變異，可分為「系統變異」和「隨機變異」。系統變異就是一種「偏離」的變異。經常見到一種評分的現象，所測出的分數，可能全是高分或者全是低分，在這種情形下，誤差是一種常數；而隨機所造成的誤差，是均勻分佈在整個領域之內，隨機誤差所造成的原因，可能是偶發的狀況、或是一些莫名的原因所造成的。隨機誤差的變動應該侷限在某一個特定的範圍內，若是所衡量的誤差，超過了某一個限度，就可以認定此一衡量工具不具有令人取信的「信度」。

「信度」也是衡量測量工具的精準性 (Precision)。所謂的精準性是指量測多次，每次所量得的結果都會一致，或是說多次量測的「變異」不大。然而，一種精準性很高的測量工具，卻不一定量測得正確。譬如，未經「歸零」的步

槍，以射手每次的射擊，都能集中同一個目標區，這個步槍可以說具有了「信度」，但是並不能說也具有「正確性」。

圖 1-1 以 A、B、C 三種步槍在靶紙上射擊的情形，說明「信度」的意義：

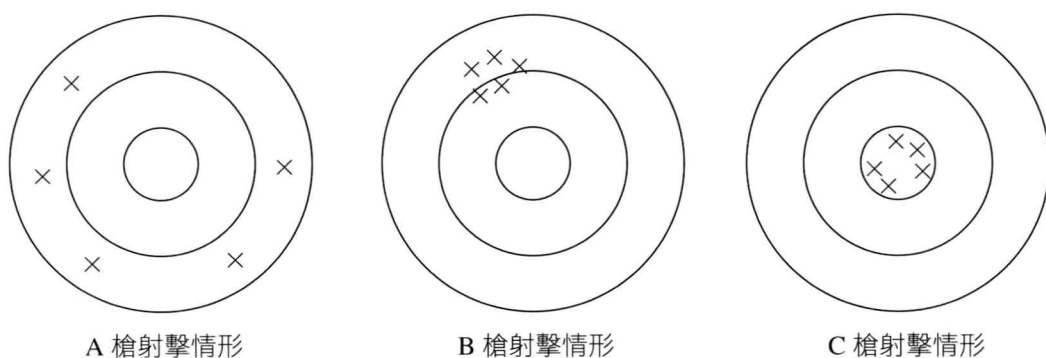


圖 1-1 A 槍、B 槍與 C 槍的信度

從 A、B、C 三槍之射擊情形看來，以三種槍的彈著得分，A 槍彈著點不但未集中於靶心，而且其彈著點的分佈又很廣、意即「變異」很大，因此並無一致性可言，結論是 A 槍之射擊既無信度亦無效度；而，B 槍的彈著點很集中，以一致性信度而言較 A 槍為佳，但是距離中心點還是很遠，因此可說是具有信度卻無效度；以 C 槍的射擊情形來看，彈著點既集中又準確，因此，既具有高信度又具有高效度。

將以上步槍打靶之例應用於一般的問卷，若是該問卷的題意非常簡明，被測之人很容易回答，則無論該項問卷施測多少次，都不會改變原來的答案，這個問卷可稱為具有很高的「信度」；但是，另有一種情形，學校經常舉行「學生對於教師教學的意見調查」，若是某位教師非常嚴格，學生因懷恨而心生報復，在意見調查表全都以「最不滿意」給分，則無論調查多少次，答案都不變，此種調查問卷的「信度」是有了，但是缺乏「精準性」，亦即，此種情形稱為沒有「效度」。



例題 1-1

某班級共有 10 位學生，這 10 位同學的程度鑑定（稱為真正程度），分別已知各為 1~10 名。如今，再以甲、乙另外兩種測驗，測試此 10 位同學之成績，其成績列於表 1-1：

表 1-1

學 生	真正程度排名	甲測驗排名	乙測驗排名
甲	1	1	5
乙	2	2	4
丙	3	3	1
丁	4	4	7
戊	5	5	9
己	6	6	2
庚	7	7	3
辛	8	8	10
壬	9	9	8
癸	10	10	6

從表 1-1 中，各同學成績情形來看，以「真正程度排名」視作基準排名，可知甲測驗的成績排名與基準排名之差異較小，因此，可認為甲測驗比乙測驗穩定，或者可說甲測驗之「信度」較高。以本例而言，由於排名情形極為顯著，很容易即可判斷兩種測驗的可靠度高低，若是排名之差異不是如此明顯，可使用「無母數統計」的「等級相關」檢定之。

假設，此 10 位同學以丙測驗所得之排名次序分別為：1，3，4，6，2，5，7，8，9，10。以 Spearman 的等級相關係數，分別計算乙測驗之排名、丙測驗之排名與「真正程度排名」之差距，計算其等級相關係數。等級相關係數之定義為：

$$rsp = 1 - \frac{6(\sum d_i^2)}{n(n^2 - 1)} \quad (1-1)$$

其中，

d_i = 「某測驗排名」與「真正程度排名」之「差距」

n = 測驗題目之題數，本例題假設 $n = 10$

以本例而言，將各種測驗排名及「差距」整理如下表：

真正程度排名	乙測驗排名	d_i	d_i^2
1	5	-4	16
2	4	-2	4
3	1	2	4
4	7	-3	9
5	9	-4	16
6	2	4	16
7	3	4	16
8	10	-2	4
9	8	1	1
10	6	4	16
合 計		$\Sigma d_i = 0$	$\Sigma d_i^2 = 102$

$$\begin{aligned} \text{rsp (乙)} &= 1 - \frac{6 \times 102}{990} \\ &= 0.3818 \end{aligned}$$

再計算「丙測驗排名」與「真正程度排名」之「差距平方和」為 $\Sigma d_i^2 = 16$ ，則

$$\begin{aligned} \text{rsp (丙)} &= 1 - \frac{6 \times 16}{990} \\ &= 0.9030 \end{aligned}$$

同理，「甲測驗排名」與「真正程度排名」之「等級相關係數」為：

$$\begin{aligned} \text{rsp (甲)} &= 1 - \frac{6 \times 0}{990} \\ &= 1.0 \end{aligned}$$

此時即可看出，丙測驗的「信度」優於乙測驗，此三種測驗「信度」之排序為：

$$\text{rsp (甲)} > \text{rsp (丙)} > \text{rsp (乙)}$$

第二節 信度之內涵

一、從受試者內在的變異分析

由於測量是以抽樣測試，其中免不了會有誤差，因此，對於受試者所獲得的實測分數，與其真正的應得分數，此兩者之間，還有一個測量誤差。意即：

實測分數 (Obtained Score)

= 真正分數 (True Score) + 測量誤差 (Error of Measurement)

以數學式表示，即為

$$X_o = X_t + X_e \quad (1-2)$$

若是測量誤差愈小，則實測分數與真正分數就愈接近。實測分數與真正分數之所以有差異，主要是由於抽樣的原因，若是以「普查」的情況測驗，其中除了工具上可能造成誤差之外，抽樣誤差幾乎已經沒有了。因此，若是樣本數愈多，或是在不改變測驗題目的品質條件下，測驗的題數愈多，此測驗的信度就會愈高，誤差自然就會愈小。上述 X_o 、 X_t 及 X_e 各有變異，在假設 X_t 與 X_e 為統計獨立的前提下，此三種變異之間的關係情形為：

實測分數之變異 = 真正分數之變異 + 誤差項之變異

其數學式表示為：

$$SS(X_o) = SS(X_t) + SS(X_e) \quad (1-3)$$

以此種觀點視之，真正分數的變異量與實測分數變異量之比，即可視為「信度」。意即：

$$\begin{aligned} r &= SS(X_t) / SS(X_o) \\ &= 1 - SS(X_e) / SS(X_o) \end{aligned} \quad (1-4)$$

由此，亦可說明「信度」是衡量測量的工具，其「隨機誤差」愈少的程度。若是隨機誤差愈少，則每次所測量的結果就會愈接近，也就是顯示出每次測量結果的「一致性」或「穩定性」就愈強。

測量中的「真正分數」，其實只是個「理想」的數字，此一「理想」值無法以某個公式計算得出，但是在測量中，對於同一個受試者，測驗無數次的平均數，庶幾可認為近似於「真正分數」。然而，由於每次測量都是隨機性的，因此，每次測量的結果與真實分數並不會相同，其間的差異稱為「測量誤

差」，這些無限多次測量誤差的機率分配，許多學者都認為近似於「常態分配」。



例題 1-2

假設某種測驗可分為甲、乙兩種。甲、乙兩種測驗皆是為了測驗 4 個母體之平均數有無差異。假設各測驗之數據皆滿足常態分配，且此 4 組母體之變異數假設相等，甲、乙測驗之資料記錄如下：

(1) 甲測驗

分 組	1	2	3	4
成 績	60	60	50	40
	40	60	50	30
	40	40	40	20
	30	10	40	20
	10	20	10	10

(2) 乙測驗

分 組	1	2	3	4
成 績	60	40	50	10
	40	10	50	40
	40	60	40	20
	30	60	40	30
	10	20	10	20

甲測驗：

$$\text{行平均 } \bar{X}_{.1} = 36 \quad \bar{X}_{.2} = 38 \quad \bar{X}_{.3} = 38 \quad \bar{X}_{.4} = 24$$

$$\text{列平均 } \bar{X}_{1.} = 52.5 \quad \bar{X}_{2.} = 45 \quad \bar{X}_{3.} = 35 \quad \bar{X}_{4.} = 25 \quad \bar{X}_{5.} = 12.5$$

$$\text{總平均 } \bar{\bar{X}} = 34$$

以下，以 $SS(T_i)$ 表示甲、乙兩測驗之總變異； $SS(C_i)$ 表示甲、乙兩測驗各行間之變異； $SS(R_i)$ 表示甲、乙兩測驗各列間之變異。 $SS(E_i)$ 是誤差之變異，可由 $SS(T_i)$ 減 $SS(C_i)$ 及 $SS(R_i)$ 獲得。

$$SS(T_1) = \sum \sum (x_{ij} - \bar{\bar{X}})^2 = 5680$$

$$SS(C_1) = \sum 5 (\bar{X}_{.i} - \bar{\bar{X}})^2 = 5 (4 + 16 + 16 + 100) = 680$$

$$SS(R_1) = \sum 4 (\overline{X_i} - \overline{\overline{X}})^2 = 4 (342.25 + 121 + 1 + 81 + 462.25) \\ = 4030$$

$$SS(E_1) = (5680 - 680 - 4030) \\ = 970$$

為便於讀者操作，乙測驗之運算，則採用 EXCEL 執行，以「雙因子變異數分析：無重複試驗」，得出 ANOVA 如下：

ANOVA					
變 源	SS	自由度	MS	F	臨界值
列	1880	4	470	1.807692	3.25916
欄	680	3	226.6667	0.871795	3.4903
錯 誤	3120	12	260		
總 和	5680	19			

$$SS(T_2) = 5680 = SS(T_1)$$

$$SS(C_2) = 680 = SS(C_1)$$

$$SS(R_2) = 1880 \neq SS(R_1) = 4030$$

$$SS(E_2) = 3120 \neq SS(E_1) = 970$$

由以上甲、乙兩種測驗得知，在 $SS(T_1) = SS(T_2)$ 及 $SS(C_1) = SS(C_2)$ 之條件下，可只考慮 R_1 與 R_2 之間的變異數差異，則此兩種測驗之信度各為：

甲測驗之「信度」：

$$r(\text{甲}) = 1 - SS(E_1) / SS(R_1) \\ = 1 - 970 / 4030 \\ = 0.7593$$

乙測驗之「信度」：

$$r(\text{乙}) = 1 - SS(E_2) / SS(R_2) \\ = 1 - 3120 / 1880 \\ = -0.6596$$

亦即可知，甲測驗之信度大於乙測驗之信度。

■ 二、從眾多受試者相互之間的變異加以分析

(一) 相關係數

以受試者相互之間的相關係數表示信度之高低，相關係數之公式為：

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \quad (1-5)$$

式中：

$$S_{xy} = \Sigma (x_i - \bar{X})(y_i - \bar{Y})$$

$$S_{xx} = \Sigma (x_i - \bar{X})^2$$

$$S_{yy} = \Sigma (y_i - \bar{Y})^2$$

其中， $S_{xx} = x$ 之變異

$S_{xy} = x$ 與 y 之共變異。

(二) 相關係數之意義

1. 相關係數表示兩個變數間的關係，其絕對值愈大，表示兩者之關係愈強，卻並不表示此二者具有因果關係。
2. 相關係數介於 +1 與 -1 之間。若是相關係數為 +1，即表示此二變數之間為「完全正相關」；若是此二變數之相關係數為 -1，則表示此二變數為「完全負相關」；若是相關係數為 0 時，稱此二變數之間為「零相關」。
3. 本節所介紹之相關係數稱為 **Pearson 積差相關係數**，其使用於兩變數間為直線相關；若是兩變數間為非線性關係時，則不宜使用此一公式。
4. 相關係數之大或小，不能由個人所決定，而應依照統計檢定公式，檢定之後方能決定該係數算是「大」或是「小」。



例題 1-3

某單位想要評量張先生之學識及品德等 5 項能力，為了更公允起見，再請同仁給予互評分數，假設 6 位同仁，以 Likert 1-5 分量表對於張先生之評分如下：