

# 淺談信度

王文心

## 一、何謂信度？

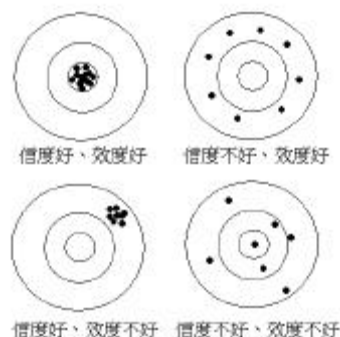
信度 (reliability) 是指對同一現象重複測量時，所得到結果一致性的高低程度。中文所稱的可靠性，以及英文的 Reproducibility、Repeatability 等，都是信度的同義複詞。影響信度的原因包括不同人、工具造成的影響 (Interobserver)，以及同一人、工具，但不同時間或環境造成的影響 (Intraobserver)。

## 二、信度與效度的關係

除了信度之外，效度 (validity) 也是我們常用來檢驗測量品質以及測量結果使用的指標。效度是指一項工具在測量其所欲測的特質或行為時所具有的真確性，又可分為內在效度 (internal validity)——結果之正確性，與外在效度 (external validity)——結果向外推論之廣度。信度與效度之間沒有絕對的相關，信度好效度有可能不好，反之，效度好，信度也不一定好。在這裡，我們用一張圖來表示信度和效度之間的關係 (圖 1)。圖中的黑點代表測量的結果，若是重複測量的結果一致性較高時，黑點就會較為集中 (信度好)，若一致性不高時，黑點則較分散 (信度不好)；而圖中最中間的圓圈代表真正的值，若黑點落在最中心的圓圈裏，才代

表測量到了真正的值，所以有可能測量結果的一致性高，但確測不到真正的結果 (圖 1 左下)。

圖 1：信度與效度示意圖



資料來源：王榮德，流行病學方法論

## 三、常用的信度指標

簡單介紹信度與效度的概念之後，我們將進入今天的主題——我們常在論文中見到作者使用的信度指標有哪些？以及這些指標是如何計算的。本文將包括測量內部一致性的 Cronbach's  $\alpha$  值；為檢驗某一種工具在重覆施測或兩種工具同時施測時，其分類結果是否一致的 Kappa 值；以及同一工具在兩個不同時間重覆測量所使用的再測信度 ICCR。我們將這三種指標的重要特性整理成表一，後文中將個別詳細介紹，並以實際的例子來介紹如何使用 SAS、SPSS 等統計軟體進行計算這些指標。

表一 常用的信度指標

指標	使用時機	資料特性
Cronbach's $\alpha$	檢驗同一測量工具的內部一致性	只須施測一次
Kappa	檢驗某一種工具在重覆施測時的一致性，或兩種測量工具在分類結果上的一致性	使用兩次測量的 <u>分類結果</u> 進行估算
ICCR	檢驗同一測量工具在兩個不同時間重覆施測時其分數的穩定性	用於基本上呈穩定趨勢的特質，如智力、性向，需要測量兩次，並且使用 <u>分數</u> 進行估算。

### (一) Cronbach's $\alpha$

Cronbach's  $\alpha$  主要是為檢驗測量工具內部的同質性、穩定度或一致性而設，所以不用重覆測量兩次。最早在測量內部一致性時，是使用庫李 (Kuder-Richardson) 信度，但是庫李信度 (公式 1) 只適用於測量結果以二分法計分之測量，例如：考試結果正確答案得分為 1，非正確答案為 0 之情況。

$$KR\ 20 = \frac{k}{k-1} \left( 1 - \frac{\sum pq}{SD^2} \right) \quad (\text{公式 1})$$

k：試題總數

$SD^2$ ：為整個測驗的變異數

p：樣本中答對該題答案之人數的百分比，代表試題的難度。

q：樣本中答錯該題的人數百分比

$\sum pq$ ：每一題試題答對與答錯人數百分比乘積的總和。

資料來源：葛樹人，心理測驗學

而為因應在測量心理或人格特質時所使用的多重評分標準需要，Cronbach (1951) 將庫李信度公式加以修改，以  $\sum (SDi)^2$  取代  $\sum pq$  而設計出下列公式 (公式 2)。KR 20 可視為 Cronbach's  $\alpha$  的一個特例，即當試題答案以二分法計分時， $\sum (SDi)^2$  即為  $\sum pq$ 。

$$\text{Cronbach's } \alpha = \frac{k}{k-1} \left[ 1 - \frac{\sum (SDi)^2}{SD^2} \right] \quad (\text{公式 2})$$

$(SDi)^2$ ：團體中所有受測者在單一試題上的變異數

SAS 和 SPSS 都可以直接計算 Cronbach's  $\alpha$  值，程式語法和視窗點選方式說明如后。

假設某一份五點量表有五個題目 (Q1-Q5)，1 分代表非常不同意，5 分代表非常同意，分別讓 19 位受試者從 1~5 分中選取其認為最符合自己狀況的

分數，最後得到的資料如下（在此列出部份資料）：

1.SAS 語法

範例一

**PROC CORR ALPHA;**

**VAR Q1-Q5;**

**RUN;**

Id	q1	q2	q3	q4	q5
001	2	3	4	2	3
002	3	2	5	3	3
003	1	4	4	2	2
004	2	3	3	1	4
005	3	3	2	2	3
006	4	4	3	3	5

執行上述程式得到的結果共有四部份：(1)這五個變項的基本統計量；(2)Cronbach's  $\alpha$  值；(3)刪除某一題目後的Cronbach's  $\alpha$  值；以及(4)皮爾森（Pearson）相關係數（在此不列出）。

### (1)基本統計量

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Q1	19	2.78947	1.18223	53.00000	1.00000	5.00000
Q2	19	2.89474	1.10024	55.00000	1.00000	5.00000
Q3	19	2.78947	1.22832	53.00000	1.00000	5.00000
Q4	19	2.42105	1.12130	46.00000	1.00000	5.00000
Q5	19	3.21053	1.03166	61.00000	2.00000	5.00000

### (2) Cronbach's $\alpha$ 值

#### Cronbach Coefficient Alpha

Variables	Alpha
Raw	<b>0.416414</b>

### (3)刪除某一問題後的Cronbach's $\alpha$ 值

Cronbach Coefficient Alpha with Deleted Variable

Deleted Variable	Raw Variables		Standardized Variables	
	Correlation with Total	Alpha	Correlation with Total	Alpha
Q1	0.411550	0.193122	0.421926	0.193376
Q2	0.278383	0.314530	0.283754	0.309385

## 研究方法專題：淺談信度

Q3	0.004178	0.523535	-.005105	0.516574
Q4	0.387659	0.224100	0.379462	0.230268
Q5	0.045195	0.472242	0.048854	0.481272

### 2.SPSS 語法：

#### RELIABILITY

/VARIABLES=q1 q2 q3 q4 q5

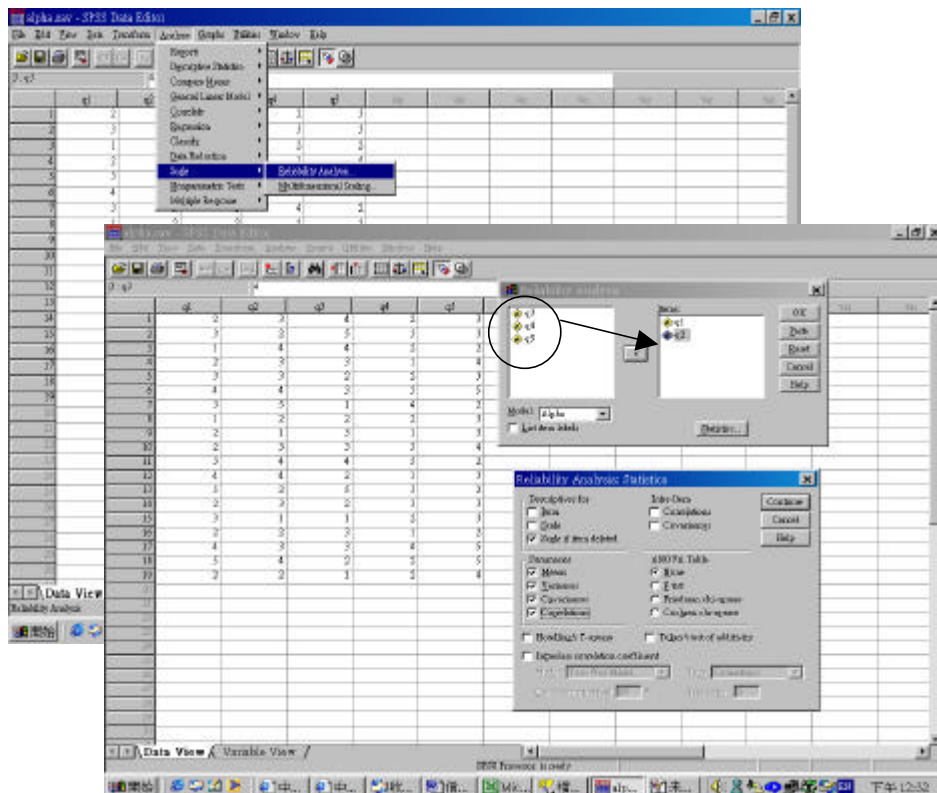
/FORMAT=NOLABELS

/SCALE(ALPHA)=ALL/MODEL=ALPHA

/SUMMARY=TOTAL MEANS VARIANCE COV CORR .

### 3.SPSS 視窗點選方式：

從功能選單的「Analyze」進入，選「Scale」，再選擇「Reliability analysis」後，便出現下面的畫面。



將同一量表的變項選到右邊，並進入「Option」設定後，執行結果依序為(1)基本統計量；(2)刪除某一變項（題目）後的 Cronbach's  $\alpha$  值；以及(3)Cronbach's  $\alpha$  值。得到的結果跟使用 SAS 統計軟體是一樣的。

(1)基本統計量

RELIABILITY ANALYSIS - SCALE (ALPHA)

Item Means	Mean	Minimum	Maximum	Range	Max/Min	Variance
	2.8211	2.4211	3.2105	.7895	1.3261	.0798
Item Variances	Mean	Minimum	Maximum	Range	Max/Min	Variance
	1.2877	1.0643	1.5088	.4444	1.4176	.0294
Inter-item						
Covariances	Mean	Minimum	Maximum	Range	Max/Min	Variance
	.1608	-.2865	.4912	.7778	-1.7143	.0568
Inter-item						
Correlations	Mean	Minimum	Maximum	Range	Max/Min	Variance
	.1252	-.2261	.3982	.6243	-1.7609	.0355

(2)刪除某一問題後的Cronbach's  $\alpha$  值

Item-total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
Q1	11.3158	5.8947	.4115	.2080	.1931
Q2	11.2105	6.8421	.2784	.1957	.3145
Q3	11.3158	8.1170	.0042	.1162	.5235
Q4	11.6842	6.2281	.3877	.2385	.2241
Q5	10.8947	8.3216	.0452	.1679	.4722

(3)Cronbach's  $\alpha$  值

Reliability Coefficients 5 items

**Alpha = .4164                      Standardized item alpha = .4171**

Cronbach's  $\alpha$  的最低接受標準值，Merchant(1985)認為是最低可接受的限度為0.5-0.6之間。

## (二) Kappa 值

Kappa 值測量的也是一致性，但與 Cronbach's  $\alpha$  不同點在於 Kappa 需要測量兩次，比較的是兩種不同工具在重覆測量，或同一種工具不同時間測量分類結果的一致性。估計 Kappa 值之前，應事先定義測量結果分類的標準，例如：如何將測量分為「及格」與「不及格」，

「陽性」或「陰性」等類別。

簡易估計 Kappa 值的方法如公式 3-1 到公式 3-3。PA 代表「同意百分比」(percent agreement)，用來計算兩測量工具（或同一工具重覆測量）在結果上完全一致的百分比。同意百分比愈高，測量之信度便愈高。

$$P_A = \sum_{i=j=k}^k N_{ij} / N \quad (\text{同意百分比})(\text{公式 3-1})$$

$$P_O = \sum_{i=j=k}^k N_{i.} N_{.j} / N^2 \quad (\text{公式 3-2})$$

$$K = \frac{(P_A - P_O)}{(1 - P_O)} \quad (\text{公式 3-3})$$

測量  
結果 1(I)

		測量 結果 2(J)				
		1	2	...	k	
測量 結果 1(I)	1	$N_{11}$	$N_{12}$	...	$N_{1k}$	$N_{1.}$
	2	$N_{21}$	$N_{22}$	...	$N_{2k}$	$N_{2.}$
	:	:	:			
	k	$N_{k1}$	$N_{k2}$		$N_{kk}$	$N_{k.}$
		$N_{.1}$	$N_{.2}$	...	$N_{.k}$	N

在此舉例說明，假設我們用兩種方法來測量某社區居民對登革熱病毒的保護力（範例二），陽性（+）代表有保護力，陰性（-）代表沒有保護力，兩種測量的結果如下所示；兩種方法測量結果皆為陽性的有 10 位，皆為陰性的

有 50 位，A 方法為陽型 B 方法為陰性的有 10 位，A 方法陰性 B 方法陽性的有 30 位。依上述公式，便可得到這兩種工具測量結果的一致性為 0.091，一致性並不理想。

### 範例二

		A 方法		
		+	-	
B 方法	+	10	30	40
	-	10	50	60
		20	80	100

$$P_A = \frac{10 + 50}{100} = 0.6$$

$$P_O = \frac{40 \times 20 + 60 \times 80}{100^2} = 0.56$$

$$K = \frac{0.6 - 0.56}{1 - 0.56} = 0.091$$

我們再來看一個複雜一點的例子，當測量的分類結果不只一類時，同樣也是按照上述的公式，將數值代入計算即可。在範例三中，A君在檢查血液中血紅素時，對同一檢體進行兩次的測量，以瞭解其操作的穩定性，結果如下表所示。第一次和第二次血紅素測量結

果都是9的有5個檢體，第一次測量為10、第二次測量得到9的有1個檢體，依此類推。最後計算得出 Kappa 值為0.33，表示同一檢體測量結果的一致性並不佳，A君可能需要對其操作過程加以檢討，以提高測量的一致性。

範例三：

		第二次測量的血紅素數值								
		9	10	11	12	13	14	15		
第一次測量的血紅素數值	9	5	2						7	
	10	1	8	3	4				16	
	11		1	6	15	3			25	
	12	2	1	2	15	5			25	
	13				9	10	4		23	
	14				4	2	7	3	16	
	15					1	8	4	13	
		8	12	11	47	21	19	7	125	

$$P_A = \frac{5+8+6+.....+7+4}{125} = 0.44$$

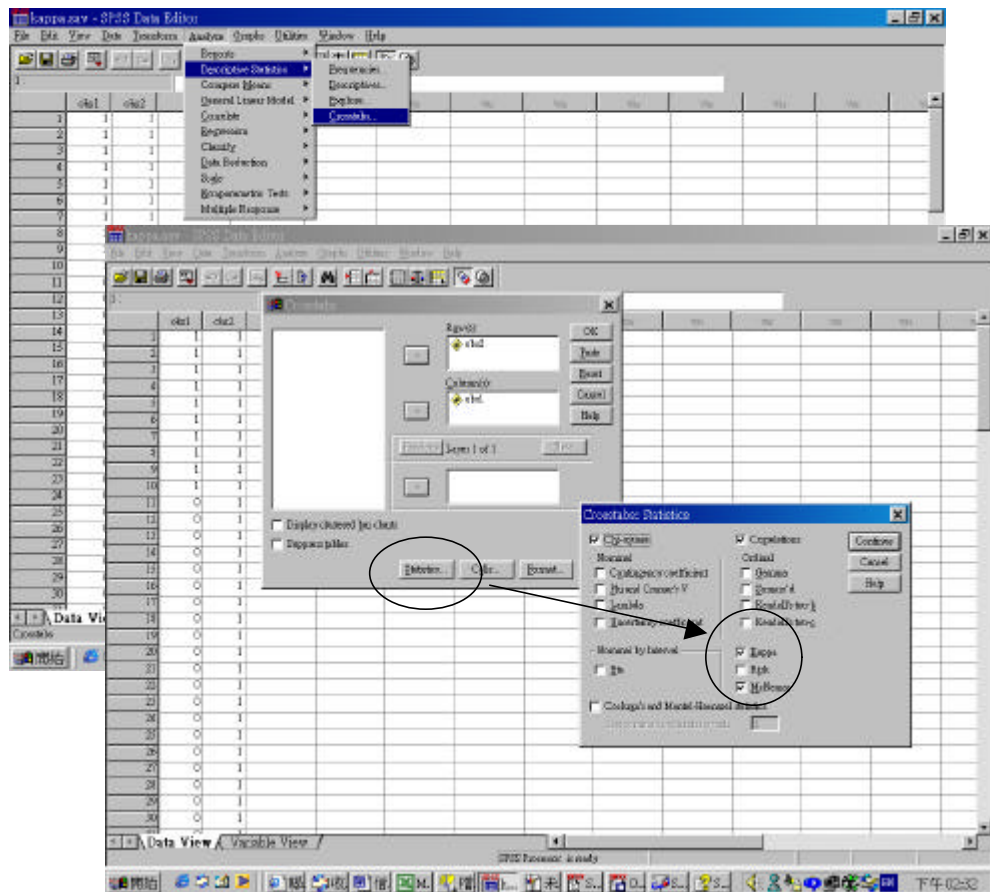
$$P_O = \frac{7 \times 8 + 16 \times 12 + ..... + 13 \times 7}{125^2} = 0.16$$

$$K = \frac{(P_A - P_O)}{(1 - P_O)} = \frac{0.44 - 0.16}{1 - 0.16} = 0.33$$

Kappa 值也可以直接用 SAS 或 SPSS 求得，但須注意原始資料的格式須為「0」與「1」。為了讓讀者瞭解資料格式，我們使用範例二的原始資料進行說明，先介紹 SPSS 視窗點選的方式，再分別示範如何撰寫 SPSS 及 SAS 的語法。

#### 1.SPSS 視窗點選

資料格式如下圖的最左方，obs1 和 obs2 分別是 A 方法和 B 方法，資料格中 0 代表陰性，1 代表陽性。計算時先從功能選單的「Analyze」進入，選「Descriptive Statistics」後再選擇「Crosstabs」，便會跳出「Crosstab」的對話視窗（如下圖）。



最後後點選「Statistics..」，勾選「Kappa」及其他相關統計值，如「McNemar」等。執行結果如下（只擇取部份），三個表依序分別是與例 1 相同的交叉列聯表、含 McNemar Test 結果的卡方檢驗結果、以及 Kappa 值。

OBS2 \* OBS1 Crosstabulation

Count		OBS1		Total
		0	1	
OBS2	0	50	10	60
	1	30	10	40
Total		80	20	100



Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.042 <sup>a</sup>	1	.307		
Continuity Correction <sup>a</sup>	.586	1	.444		
Likelihood Ratio	1.026	1	.311		
Fisher's Exact Test				.320	.221
Linear-by-Linear Association	1.031	1	.310		
McNemar Test				.002 <sup>c</sup>	
N of Valid Cases	100				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.00.

c. Binomial distribution used.

Symmetric Measures

		Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Interval by Interval	Pearson's R	.102	.102	1.016	.312 <sup>c</sup>
Ordinal by Ordinal	Spearman Correlation	.102	.102	1.016	.312 <sup>c</sup>
Measure of Agreement	Kappa	.091	.091	1.021	.307
N of Valid Cases		100			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on normal approximation.

## 2.SPSS 語法

### CROSSTABS

/TABLES=obs2 BY obs1

/FORMAT= AVALUE TABLES

/STATISTIC=CHISQ CORR KAPPA

### MCNEMAR

/CELLS= COUNT .

### The FREQ Procedure

Table of OBS2 by OBS1  
OBS2      OBS1

	Frequency		Total
	0	1	
0	50	10	60
1	30	10	40
Total	80	20	100

## 3.SAS 語法

DATA A;

INFILE 'D:\KAPPA.TXT';

INPUT OBS1 OBS2;

PROC FREQ;

TABLE OBS2\*OBS1/AGREE;

RUN;

執行後可以得到與 SPSS 一樣的結果：

Statistics for Table of OBS2 by OBS1

McNemar's Test	
Statistic (S)	10.0000
DF	1
Pr > S	0.0016

### Simple Kappa Coefficient

Kappa	0.0909
ASE	0.0911
95% Lower Conf Limit	-0.0876
95% Upper Conf Limit	0.2695

Sample Size = 100

當 Kappa 值等於 1 時，代表完全一致， $K > 0.80$  表一致性佳，而  $K < 0.4$  時，則代表一致性不好。

### 4. McNemar's test

在上面的計算中，讀者可以看到 McNemar 檢定的結果，其檢定的結果與 Kappa 值代表的意義是一樣的。McNemar 氏考驗又稱為非獨立樣本比率數的卡方考驗，或稱為相依樣本的卡方檢定，可用來檢驗重複量數之資料，例如：看不同測量工具對同一樣本測量的結果是否一致，或是看同一樣本，前後兩次反應之異同。

		Observer2		
		+	-	
Observer1	+	W	X	W+X
	-	Y	Z	Y+Z
		W+Y	X+Z	N

$$c^2_{McN} = \frac{[(X - Y)]^2}{X + Y} \quad \text{or} \quad c^2_{McN} = \frac{[(X - Y) - 1]^2}{X + Y}$$

$$c_{McN} = \frac{(X - Y)}{\sqrt{X + Y}}$$

(公式 4)

我們以先前舉過的範例二來做說明，McNemar 檢定的虛無假設 (Null

hypothesis) 是假設兩種方法測量到的結果不會有差異，但是當我們計算結果發現  $p$  值小於 0.05 時，便推翻虛無假設，亦即不同的測量方法所得到的測量結果是不一樣的，與我們之前使用 Kappa 值得到的結論—兩種方法一致性不佳是一樣的。同樣的，也與上面以統計軟體得到的計算結果是一樣的。

範例：

$H_0$ : 不同測量方法不會對測量結果造成差異。

$H_A$ : 不同測量方法對測量結果造成差異。

		A 方法		
		+	-	
B 方法	+	10	30	40
	-	10	50	60
		20	80	100

代入公式 4 計算：

$$c_{McN} = \frac{30-10}{\sqrt{30+10}} = 3.162 \quad p=0.002$$

(三) ICCR--再測信度 (test-retest reliability)

一些人類特質，例如：智力、性向、人格等基本上是穩定的，故在測量這類心理特質時，測驗分數具有高度穩定性是必要的。最適宜的相隔時間隨測驗目的和性質而異，少者兩週、多則六個月到 1、2 年。

雖然 Kappa 值也可用來當然再測信度的指標，但是 Kappa 處理的是分類過後的測量結果，像是智力、性向、人

格等就不適合用 Kappa 值來做為信度的指標。因此像是以皮爾森積差相關來處理兩次測量的分數，將所得到的係數做為再測信度，或是用我們這裡將要介紹的 ICCR ( Intraclass Correlation Coefficient Reliability ), 都可以用來評估的穩定度。

在 One-way Random Effects Model (以同一種工具評估所有病人) 的假設情況下，ICCR 值可由下列公式計算：

$$r1 = \frac{\hat{S}^2_s / (\hat{S}^2_s + \hat{S}^2_e)}{MSB + (n_0 - 1)MSW} \quad (\text{公式 5})$$

MSB : mean of between sum of square  
MSW : mean of within sum of square

因為通常是比較兩個不同時間點測量結果的一致性，所以  $n_0=2$ ，目前尚無法直接從 SAS 和 SPSS 下指令計算 ICCR，但是可先進行 ANOVA 運算，得到公式 5 中所需要的 MSB 和 MSW 數值後，再代入公式計算 ICCR 值。而在 ICCR 值的判讀上， $<0.40$  為信度不佳， $0.4-0.75$  之間良好， $>0.75$  極佳。

結語

本文中我們介紹了三種常用的信度指標，相對於沒有提到的折半信度、複本信度等，這三種指標都有計算簡便、容易實行、限制較少的特點。然而信度要達到多高才能被接受呢？這是一個不易回答的問題，主要因為研究者的

需要不同，很難訂定一個適用於全部狀況的單一標準。

測量工具本身、受測樣本的變異性、施測者、施測環境等等因素，都會影響信度的高低，大致而言，一般認為信度係數達 0.4 以上屬於良好，要達到 0.8 以上才算極佳，但是為了要將信度從 0.4 提高到 0.8，可能需要花費許多的時間與金錢去嘗試，是否有其必要性，則值得多方思考。

理想的測量應具備良好的信度和效度，藉由信度和效度的高低，研究者可以對該次測量的品質加以分析，由於目前已經發展的信度測量方式有許多種，我們在呈現研究結果時，應詳細說明測量信度的方式並列出結果，而讓其他研究者依其特定目的來決定信度是否足夠，並據此判斷如何使用資料。

參考書目：

葛樹人，民 77，心理測驗學。台北：桂冠圖書公司。

邱皓政，民 89，社會與行為科學的量化研究與統計分析 SPSS 中文視窗版資料分析範例解析。台北：五南圖書出版公司。

Shoukri, M.M. (1999) Statistical Methods for Health Sciences, 2<sup>nd</sup> ED., pp. 18-107. Boca Raton, Fla. CRC Press, New York .

Edward G. Carmines. and Richard A.Zeller (1979) "Reliability and Validity Assessment." Sage University Paper Series on Quantitative Application in the Social Sciences,07-017. Beverly Hills and London: Sage Pubns.