

行政院國家科學委員會專題研究計畫 成果報告

評定量表相關係數校正影響因素之模擬研究

計畫類別：個別型計畫

計畫編號：NSC92-2413-H-002-018-

執行期間：92年08月01日至93年07月31日

執行單位：國立臺灣大學心理學系暨研究所

計畫主持人：翁儷禎

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 11 月 3 日

行政院國家科學委員會專題研究計劃成果報告

評定量表相關係數校正影響因素之模擬研究

Factors Affecting Correction for Attenuation

With Likert-Type Rating Scales: A Monte Carlo Study

計畫編號：NSC 92-2413-H-002-018

執行期限：92年8月1日至93年7月31日

主持人：翁儷禎 國立台灣大學心理學系

ljweng@ntu.edu.tw

一、中文摘要

本研究擬以模擬研究的方法探討信度類別與量尺特性對評定量表相關係數校正的影響。模擬研究中擬操弄量表點數、量表內題目間的共變數、量表題數、受試者判斷的正確度(即個別題目之信度),跨量表題目真分數間的共變數,與點量表分配六個因素,比較以內部一致性信度 α ,再測信度,真分數與觀察分數相關的平方等三種信度係數校正之相關係數之差異。結果發現以內部一致性信度 α 校正之相關係數在題目同質性低時易高於理論值,其他兩類信度校正之相關係數的偏誤則較小。

關鍵詞：李克式量表、內部一致性信度、再測信度、效標關聯效度、相關係數校正

Abstract

The Monte Carlo experiment examined the effects of the type of reliability and the properties of Likert-type rating scales on correlations corrected for attenuation. The independent variables manipulated included number of response categories, inter-item covariance, number of items, respondents' judgment accuracy (or reliability of

individual item), correlations between true score items across measures, and distribution of scores of the rating scales. The quality of the Pearson product-moment correlations corrected for attenuation by the internal consistency reliability α , the test-retest reliability, and the squared correlation between true scores and observed scores were compared. The correlation corrected by internal consistency reliability α was found to be higher than the population value when items were less homogeneous. The correlations corrected by the other two types of reliability, in general, yielded less biased results.

Keywords : Likert-type scales, internal consistency reliability, test-retest reliability, criterion-related validity, correction for attenuation

二、前言與研究目的

評定量尺為社會科學研究常用的量化方法之一,研究者常藉評定量表測量個人特質或反應(王、翁,民91;楊、趙,民76; Likert, Roslow, & Murphy, 1993)。在量表建構中,信度與效度是研究者需要探討的兩個主要的心理計量特性。自從

Likert (1932) 提出此測量方法以來，已有許多學者探討評定量尺設計對量尺信效度的影響。在過去此方面研究中，研究者通常將量尺設計對量表信度與效度的影響分開探討，本研究則擬綜合雙方面的研究，以模擬研究探討量表特性對相關係數校正 (correction for attenuation) 的影響。

相關係數是效度研究中常用的統計量，效標關聯效度 (criterion-related validity) 一般即以測驗與效標間的相關係數進行評估。又如聚斂效度 (convergent validity) 與區辨效度 (discriminant validity)，亦是應用多個測量工具間的相關係數綜合討論 (Campbell & Fiske, 1959)。此外，在一般研究中，研究者亦常運用相關係數以瞭解變項間的關係。然而，測量工具常會因為測量誤差之故，無法具有完全的信度。此等測量誤差會影響相關係數的估計，使得所估計出來的相關係數被低估，致而無法充分反應研究者研究旨趣所在之心理構念間的理論關係。有鑑於此，Spearman (1904) 遂乃提出相關係數校正的概念與作法，嘗試估計當測量工具不受測量誤差影響時，理論構念間的相關為何。Block (1963) 認為當研究重點為評估兩測量工具在所測量的構念上是否相當時 (conceptual equivalence)，相關係數應該適度地校正，校正因測量誤差而導致的相關係數下降程度，以充分反應理論構念間的關聯程度。雖然對於相關係數校正的適用性至今仍有論爭 (Muchinsky, 1996)，相關校正依然有其適用之必要情境。以前述構念比較研究為例，當研究者的主要目的為了解不同量表所測量構念間的關係時，測量誤差乃為實徵研究中難以避免的瑕疵，如果研究者未對樣本估計之相關係數進行校正，測量誤差將使得理論構念間的關係強度被低估，而無以呈現構念間的關聯程度，甚至可能

因此錯誤推論不同量表測量之構念的關係。Block 即舉多例說明此現象。Muchinsky (1996) 並認為，效度類化 (validity generalization) 研究、整合分析 (meta-analysis)，以及效用評估 (utility assessment) 研究之發展亦使得相關係數之校正受到重視。

但是，倘若要進行相關係數校正，究竟要採用何種信度係數？各類型的信度係數乃評估不同來源的測量誤差對測量結果的影響程度 (Anastasi & Urbina, 1997; Crocker & Algina, 1986)，因此，採用何種信度宜取決於研究中對測量誤差的定義 (Guilford, 1954; Lord & Novick, 1968)。此建議就概念層次考量相當合理，然而，一個測驗的分數可能同時受各種誤差來源影響，此時，經由不同信度係數校正後的相關係數結果有何差異？何者偏差較小較接近真實的相關？而且，評定量表的設計是否會影響校正後的相關？本研究即欲以模擬研究方法回答此等問題，探討信度類別與量表特性對相關係數校正正確度的影響。

Lissitz 與 Green (1975) 率先以模擬研究的方法探討量表信度如何隨著量表點數與題間共變數而改變，冀能提供量表編製與使用者選擇適當點數。該研究探討的信度係數包括內部一致性信度係數 α ，再測信度，真分數與觀察分數相關的平方三者。Jenkins 與 Taber (1977) 繼之以 Lissitz 與 Green 的研究為基本架構，再考慮可能與量尺點數交互影響量表信度的其他因素，進一步探討量表點數、題目間的共變數、量表題數、受試者判斷的正確度 (亦即個別題目之信度) 四個因素及其交互作用對三信度係數的影響。吳 (民 85) 亦以模擬實驗探討問卷長度、項目間平均相關係數、各項目變異數變異程度三者對內部

一致性信度係數 α 之影響。但其乃估計模擬之連續資料的 α 值，未將連續資料轉換成間斷之評定量表式點資料。

Jenkins 與 Taber (1977)發現可以加成模式整合操弄之四個因素對量表信度的影響。Lissitz 與 Green (1975) 及 Jenkins 與 Taber 此二模擬研究均發現量表點數達到五點之後，就算點數增加，信度亦不會隨之而增高，因而認為以五點量尺進行測量即已足夠。然而此二研究假設回答各點數的人數均等，此種情況在一般真實資料中並不多見(Micceri, 1989)，也因此其結論難以直接類推到實際資料上(Weng, 2004a)。Jenkins 與 Taber 亦指出，假設回答各點數的人數均相等乃該研究的限制，因為大多數實際資料的分配並非如此。因此，為增進模擬研究結果在實際資料分析時的參考性，乃有研究者進一步以模擬研究方法探討評定量尺分配等因素對量表信度與效度的影響，將連續的觀察分數作不同的轉換，以形成各式的點量表分配，而非僅假設各點的機率相同(翁, 民 90, 民 91; Enders & Bandalos, 1999; Bandalos & Enders, 1996; Weng, 2004b; Weng & Cheng, 2004)。

翁(民 90, 民 91; Weng, 2004b; Weng & Cheng, 2004) 以模擬研究的方法探討影響評定量表信度與效標關聯效度的因素。模擬研究中操弄量表點數、題目間的共變數、量表題數、受試者判斷的正確度(即個別題目之信度)，與點量表題目得分分配五個變項，以探究此五因素對內部一致性信度 α ，再測信度，真分數與觀察分數相關的平方等三種信度係數，以及代表效標關聯效度之相關係數的影響。結果發現各交互作用項的影響相當小，各因素的影響可以加成模式解釋，惟各因素對不同信度指標的影響程度未必相同。內部一致性信度主要受題間共變數與量表題數影響；再

測信度主要受受試者判斷的正確度影響，量表題數次之；真分數與觀察分數相關的平方主要受受試者判斷的正確度與量表題數影響，量尺點數與分配次之。至於量表特徵與受試者特性對評定量表效標關聯效度的影響則發現，題間共變、量表題數與受試者判斷正確度對效標關聯效度的影響較大，量表點數與題目分配則影響不大。研究結果顯示，以評定量表測量時，影響信度指標與相關係數的因素並非完全一致。若是如此，採用不同的信度係數校正相關係數時，效果是否會相左？在不同的評定量表設計下，各類信度係數的校正效果為何？此即本研究欲回答的問題。

Zimmerman 與 Williams (1997)曾以模擬研究探討分數分配對相關校正的影響，發現當信度高時，校正後的相關係數非常接近真分數相關，但低信度則易導致過度校正，甚而得到高於 1 的校正後相關，母群分數的分配與相關校正之結果則無關係。然而，Zimmerman 與 Williams 的研究並未充分考量評定量表特性的影響，以致於其結果是否能類推至使用評定量表測量的情形，則有待進一步探討。言要之，本研究以模擬研究的方法，探討信度類別與量表特性對相關校正結果的影響。由於評定量表常用以收集研究資料，此研究結果將有助於瞭解考量不同誤差來源時，相關係數校正之品質。

三、研究方法

本研究共操弄六個獨變項，各獨變項包含之情境如下。

- (a) 量尺點數：量尺點數涵蓋 2 至 9 點。此設計未包括過去研究之多點數情境，主要考量以往研究結果與一般人之認知能力兩因素。過去研究發現量尺點數的

效果在五至七點後即不甚明顯(翁，民 90，民 91; Jenkins & Taber, 1977; Lissitz & Green, 1975)，而且研究參與者未必能精準地區辨多點數反應間的差異，回答高點數量表極可能反而引進不必要之測量誤差。

- (b) 量表內題目間的共變數：此變項依 Lissitz 與 Green 和 Jenkins 與 Taber 的作法，包含 0.2、0.5、0.8 三數值。
- (c) 量表題數：此變項依 Jenkins 與 Taber 的作法，包含 2、3、5、7、9、10、14 題，另加 1、12、15 與 20 題四種情形。
- (d) 受試者判斷正確度：亦即題目之信度，為題目真分數變異數對題目觀察分數變異數之比值，此變項依 Jenkins 與 Taber 的作法，包含 0.50、0.70、0.85、1.00 四種情形。
- (e) 點量表題目得分分配：本研究由檢視以往實徵資料中各點數量尺題目的頻率分配（例如翁，民 88），選取較常出現之分配，作為模擬研究中點量表題目分配選取之依據。共選取六種分配情形，包含均等分配、常態分配，以及四種不同程度偏態之分配，此四分配之偏態係數與峰度係數分別為(1.0, 1.5)、(2.0, 4.0)、(4.0, 7.0)與(0.0, 4.0)。
- (f) 兩量表各題真分數間之共變數：此變項依量表內題間共變數而異，在題間共變為 0.2, 0.5, 0.8 時，分別為 0.1, 0.1-0.4, 與 0.1-0.7。

此模擬研究之信度指標包括內部一致性信度 α ，再測信度，真分數與觀察分數相關的平方等三種信度係數。資料產生步驟乃根據 Lissitz 與 Green(1975)之作法，先依古典測驗理論建構連續觀察資料，再轉換成各種題目得分分配之點量表反應資料。每一情境重覆 100 次，每次產生 100 個受試者資料。資料收集後，即計算各情

境各信度係數、兩量表相關係數，以及分別以三種信度指標進行校正後之相關係數的平均數與標準誤等基本統計量，並計算各信度校正後相關係數的偏誤。繼以變異數分析之 η^2 探究各因素及其交互作用對各信度校正之相關係數的影響程度， $\eta^2 > .138$ 代表效果值高， $\eta^2 > .059$ 代表中等程度效果值(Cohen, 1988)。

四、結果與討論

本研究就同量表內題目間共變之三種情境分別對以三類信度校正後之相關係數進行變異數分析，結果發現兩量表各題真分數間之共變與量表題數的效果量最大，量尺點數、點量表題目得分分配、受試者判斷正確度之影響不大；而且，交互作用項的效果相當小，幾乎均未達小效果。換言之，就本研究探討的變項與情境觀之，兩量表各題真分數間之共變與量表題數是影響校正後相關係數大小之主要因素。此結果與兩量表真分數相關之理論值相符，兩量表真分數相關之理論值即為校正後相關之理論值，顯示本研究之模擬資料應為合理正確。基於本模擬研究設計之情境，以代數推導可知，兩份量表間真分數之高低與兩量表各題真分數間之共變、量表題數、同量表內各題間之共變三者有關，而且，當量表題數趨近無窮大時，量表真分數之相關趨近兩量表各題真分數間之共變與量表內題目間共變兩者之比值。由於題數為 1 之結果與題數大於 1 之組合量表分數的結果具相當差異，有鑑於一般心理學量表均為多數題之組合，本文之討論因之乃集中於題數大於一之量表情境。

為進一步了解不同類型信度校正之相關係數與理論值之差異，乃分別檢視比較不同情境下以三種信度校正之相關係數

的平均數與量表真分數相關理論值之差距，此亦即各信度校正後相關係數的偏誤。結果發現內部一致性信度係數 α 在量表內題間共變為 0.2 與 0.5 時可能過度校正量表間的相關程度，而導致校正後的相關高於理論值，在題數少如二或三題時尤然。再測信度以及真分數與觀察分數相關的平方則較少發生此現象，平均數較接近理論值，其中再測信度有時會稍微低估量表真分數間之相關。三類信度校正結果之差異主要可能肇因於三類信度估計值之數值高低。在 Lissitz & Green (1975) 與 Weng (2004b)，以及本研究的模擬研究設計下，量表各題符合平行項目之假設，此時，內部一致性信度係數 α 乃為信度之估計而非下限。然而，內部一致性信度係數 α 的數值在量表內題間共變非 0.8 時，均低於其他兩類信度，致使其校正後之相關偏高。

因此，誠如 Cronbach (1947) 與 Lord 與 Novick (1968) 所言，以不同類型之信度校正相關係數可能得到不同之結果，然而其間之比較不見的沒有意義。一個量表或測驗的分數可能同時受各種誤差來源影響，雖然採用何種信度宜取決於研究中對測量誤差的定義(Guilford, 1954; Lord & Novick, 1968)，但經由不同類別信度係數校正後之相關係數的偏誤並不一致。本研究發現，再測信度以及真分數與觀察分數相關的平方校正之相關的偏誤一般較小較接近相關係數的理論值，內部一致性信度係數 α 在量表內題間共變低時，則可能過度校正而造成嚴重偏誤，題數少時尤然。再測信度在題目反應點數少如 2 或 3 點，題目信度高或極端分配如 (0, 4) 與 (3, 7) 時，可能因校正不足而低估量表真分數間的相關。以真分數與觀察分數相關平方校正之相關係數整體結果則頗佳，只於部分極端情境具偏誤。有鑑於各類信度校正

後相關在偏誤上的差異，研究者宜於界定測量誤差後，避免可能產生偏誤的李克式量表設計，以免造成相關係數之過度校正或校正不足的情形。

五、計畫成果自評

本研究原則上依原計畫進行，另並推導兩量表真分數相關，亦即校正後相關係數之理論值以作比較。本研究之研究結果將有助於對相關係數校正影響因素之瞭解，應可於相關期刊上發表。

六、參考文獻

- 王嘉寧、翁儷禎 (民 91)。探索性因素分析國內應用之評估：1993 至 1999。「中華心理學刊」，第 44 卷，頁 239-251。
- 吳瑞屯 (民 85)。影響 α 內部一致性係數的因素。「中華心理學刊」，卷 38：51-59。
- 翁儷禎 (民 88)。「點數與標示語對評定量尺反應的影響 (I)」，國科會專題研究計劃報告：NSC 87-2413-H-002-010。
- 翁儷禎 (民 90)。「點數對評定量表信度與效度之模擬研究 (I)」，國科會專題研究計劃報告：NSC 89-2413-H-002-048。
- 翁儷禎 (民 91)。「點數對評定量表信度與效度之模擬研究 (II)」，國科會專題研究計劃報告：NSC 90-2413-H-002-022。
- 楊中芳、趙志裕 (民 76)。中國受試者所面臨的矛盾困境：對過分依賴西方評定量表的反省。「中華心理學刊」，卷 29：113-132。
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). London: Prentice-Hall International, Inc.

- Bandalos, D. L., & Enders, C. K. (1996). The effects of nonnormality and number of response categories on reliability. Applied Measurement in Education, 9, 151-160.
- Block, J. (1963). The equivalence of measures and the correction for attenuation. Psychological Bulletin, 60, 152-156.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.
- Cronbach, L. J. (1947). Test "reliability": Its meaning and determination. Psychometrika, 12, 1-12.
- Enders, C. K., & Bandalos, D. L. (1999). The effects of heterogeneous item distributions on reliability. Applied Measurement in Education, 12, 133-150.
- Guilford, J. P. (1954). Psychometric methods. New York: McGraw-Hill.
- Jenkins, G. D., Jr., & Taber, T. D. (1977). A Monte Carlo study of factors affecting three indices of composite scale reliability. Journal of Applied Psychology, 62, 392-398.
- Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 140.
- Likert, R., Roslow, S., & Murphy, G. (1993). A simplified and reliable method of scoring the Thurstone attitude scales. Personnel Psychology, 46, 689-690.
- Lissitz, R. W., & Green, S. B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. Journal of Applied Psychology, 60, 10-13.
- Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. New York: Addison-Wesley.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. Psychological Bulletin, 105, 156-166.
- Muchinsky, P. M. (1996). The correction for attenuation. Educational and Psychological Measurement, 56, 63-75.
- Spearman, C. (1904). The proof and measurement of association between two things. American Journal of Psychology, 15, 72-101.
- Weng, L.-J. (2004a). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. Educational and Psychological Measurement, 63.
- Weng, L.-J. (2004b). Factors affecting three indices of test reliability: A Monte Carlos study. Manuscript submitted for publication.
- Weng, L.-J., & Cheng, C.-P. (2004). A Monte Carlo study of factors affecting criterion-related validity with Likert-type rating scales. Manuscript in preparation.
- Zimmerman, D. W., & Williams, R. H.

(1997). Properties of the Spearman correlation for attenuation for normal and realistic non-normal distributions. Applied Psychological Measurement, 21, 253-270.