

行政院國家科學委員會補助專題研究計畫成果報告

點數與標示語對評定量尺反應的影響(II)

計畫類別：個別型計畫 整合型計畫

計畫編號：NSC 89 - 2413 - H - 002 - 012 -

執行期間：88年8月1日至89年7月31日

計畫主持人：翁儷禎

共同主持人：

本成果報告包括以下應繳交之附件：

赴國外出差或研習心得報告一份

赴大陸地區出差或研習心得報告一份

出席國際學術會議心得報告及發表之論文各一份

國際合作研究計畫國外研究報告書一份

執行單位：國立臺灣大學心理學系

中華民國89年10月31日

行政院國家科學委員會專題研究計畫成果報告

點數與標示語對評定量尺反應的影響 (II)

Effects of Number of Categories and Anchor Labels on Performance of Rating Scales (II)

計畫編號：NSC 89-2413-H-002-012

執行期限：88年8月1日至89年7月31日

主持人：翁儷禎 國立臺灣大學心理學系

計畫參與人員：林純純 國立臺灣大學心理學系

一、中文摘要

本研究探討評定量尺標示語與點數對受試者反應的影響。量表標示語的建構包含五種型式：兩端型、兩端加中間型、等距型、正向型與負向型；點數則包括三點至九點等七種情境，共計發展二十九種量表型式。結果發現，量表標示語對受試者反應的影響主要來自正向型與負向型之操弄，量表平均數、標準差與內部一致性信度受影響，且其效果隨量表內容或者受試者量表反應之異質性而異，受試者反應較具同質性的量表對標示語的改變可能較敏感。量表點數對內部一致性信度的影響亦隨量表內容而異，此亦可能與量表變異量有關。此研究結果顯示，評定量尺的標示語與點數可能影響受試者反應，研究者建構評定量表時，宜考慮受試者在該量表可能的反應異質程度，而選擇適用之量表型式。

關鍵詞：評定量尺、李克特式量尺、量尺標示語、量尺點數、內部一致性信度

Abstract

The purpose of the study is to investigate the effects of anchor labels and number of response categories on participant responses on Chinese rating scales and the psychometric properties of the scales. Five types of rating scales with different formats

of labels were constructed. Number of response categories varied from three to nine. A total of 29 different types of rating scales have been developed. The results indicated that the effects of anchor labels mainly came from the differences between participant responses on the positive and the negative scales. Scale means, scale standard deviations, and internal consistency reliability coefficients have been affected. The magnitude of the effects depended on the content of the scale or the heterogeneity of participant responses on the scale. The scales with homogeneous participant responses may be more sensitive to the changes of anchor labels than those with heterogeneous participant responses. The effects of the number of response categories on internal consistency reliability also depended on content of the scales or the degree of heterogeneity of participant responses on the scales. The results of the present investigation suggested that anchor labels and number of categories of a rating scale may influence participant responses and psychometric properties of the scale. Researchers better consider the heterogeneity of participant responses on the scale and select appropriate scale format.

Keywords : rating scales, Likert-type scales, anchor labels, number of response categories, internal consistency reliability

二、緣由與目的

評定量尺為社會科學研究中常用的度量方法之一，研究者常藉評定量表測量個人特質或反應。建構評定量尺時，研究者一般會考慮量尺點數與量尺標示語的選擇兩個向度。自 Likert (1932) 提出此測量方法以來，國外已有許多學者探討此類量尺的特性，相關研究議題主要包括三大方向：(1) 量尺點數對量表信度與效度的影響(如: Bendig, 1953; Champney & Marshall, 1939; Komorita & Graham, 1965; Klockars & Hancock, 1993; Lehmann & Hulbert, 1972; Lissitz & Green, 1975; Matell & Jacoby, 1971; Oaster, 1989; Symonds, 1924)。(2) 量尺標示語(如: 經常、相當符合)在受試者心目中的量尺值(scale values)(如: Cliff, 1959; Hakel, 1968; Jones & Thurstone, 1955; Spector, 1976 等)。(3) 量尺標示語對受試者反應的影響(如: Bendig, 1955; French-Lazovik & Gibson, 1984; Hancock & Klockars, 1991; Klockar & Yamagishi, 1988; Lam & Klockars, 1982; Newstead & Arnold, 1989; Wildt & Mazis, 1978 等)。

以往英文評定量尺的研究發現，受試者的反應會受量尺標示語等特徵影響(如: Bendig, 1955; Klockars & Yamagushi, 1988; Wildt & Mazis, 1978 等)，量尺的點數亦會影響量表的心理計量特徵(如: Aiken, 1983; Churcher & Peter, 1984; Halpin, Halpin, & Arbet, 1994; Jenkins & Taber, 1977; Komorita & Graham, 1965; Klockars & Hancock, 1993; Lissitz & Green, 1975; Masters, 1974 等)。相對於國外長年累積

的研究結果，國內有關中文評定量尺的研究卻頗有限(胡、彭、沈、楊, 民 78; 柯, 民 83; 翁, 民 85, 民 87; 黃, 民 75; Wong, Chuen, & Fung, 1993)。本研究擬延續過去國內評定量尺的研究，探討量尺標示語以及量尺點數對受試者反應的影響。在點數方面包含三點到九點各奇偶點數的型態，標示語則以五種方式建構，期能較全面性地探討此一議題。

過去的研究者曾設計各種不同語詞的評定量表，包括等距型、兩端型、正向型、負向型等。等距型量表的各個選項點均加上標示語，且各標示語間的量尺值間距盡量相等，使其具備等距量尺的特性。兩端型量尺則僅在量尺的最兩端註明標示語詞，其他各選項點則不提供標示語。正向型量尺的標示語以正向標示語為主建構，因此所採用的標示語並非對稱，而是將正向反應作精細區分，輔以少數負向標示語。舉例來說，「完全同意-非常同意-相當同意-大致同意-有些同意-不同意」即是一個六點的正向型反應評定量尺。負向型量尺則與正向型量尺相反，精細區辨受試者的負向反應。Dixon、Bobo 與 Stevick (1984), Frisbie 與 Brandenburg (1979), Newstead 與 Arnold (1989)的研究比較等距型與兩端型的量表型式對受試者反應的影響。Frisbie 與 Brandenburg 發現此二型式量表的受試者反應不同，其他兩個研究卻發現兩者所引發的受試者反應差不多。Lam 與 Klockars (1982)研究五點等距型、兩端型、正向型與負向型量尺，結果發現各題平均數隨量表標示語而改變，而等距型與兩端型的受試者反應相近。Klockars 與 Yamagishi (1988)、Wildt 與 Mazis (1978)、French-Lazovik 與 Gibson (1984)亦發現量表型式影響受試者反應。Hancock 與 Klockars (1991)的結果顯示五點等距型與正向型頻率量尺的分配不同，但 Klockars 與 Hancock (1993)卻發現五點等距型與正

向型評價量尺的效度係數相近。

雖然國外已累積部分量表標示語對受試者反應影響的研究結果，但誠如 Klockars 與 Yamagishi 指出，這類研究仍有待在其他母群中進行驗證，以知其研究結果的推論程度，看其是否為受試者的普遍反應。國內有關標示語對受試者反應的研究相當稀少。黃（民 75）曾探討五點、七點、九點三種奇數點量表的再測信度和內部一致性信度，以及量尺標示語對受試者反應的影響。然而該研究為小樣本研究，各情境的樣本數至多三十人。因此，其結果有待後續研究以較大的樣本檢視其結果之穩定性。由於評定量表大多數採用文字敘述，尤其是量表標示語，有鑑於國內相關研究之缺乏，以及評定量表使用之普遍，實有必要探究標示語對華文評定量表反應的影響。

過去有關量尺點數與標示語的研究多半是二者獨立進行。在量尺標示語的研究中，研究者大都直接操弄量尺標示語，再比較受試者在不同量尺型式上反應的異同，這些研究多半未考慮所採量尺或題目的信度或穩定性。然而，測量工具如果信度不高，受試者反應可能即不夠穩定；因此，受試者在不同量尺型式上反應的差異，可能源自低信度與測量誤差，而非量尺反應型式的操弄。是故，為釐清此混淆變項的影響，本研究乃以翁（民 88）的結果為基礎，進一步探討量表標示語對受試者反應及量表心理計量特性的影響。翁（民 88）探討不同點數下的分量表再測信度與內部一致性信度。本研究則進一步操弄量尺的標示語，研究其對受試者反應及量表心理計量特性的影響。由於國內評定量尺之使用以同意度評估居多（參翁，民 85，附錄一），故本研究即就同意度量尺先行探討。

既然評定量尺廣為國內研究者採用（楊、趙，民 76），探討中文評定量尺特

徵如何影響受試者反應，實屬需要。此等研究結果，配合標示語心理量尺值研究，可作為評定量尺設計之參考，協助研究者選擇合宜的量尺點數與標示語，以建構適合其研究情境與後續資料分析方法的評定量尺型式。

三、研究方法

本研究以研究者編製之關心(CO)與堅毅力(DE)兩量表為施測工具，此二量表為一甄選測驗之二分量表，作答型式為李克式五點量表，採用「完全不符合、大致不符合、有時符合有時不符合、大致符合、完全符合」的標示語，各包含 12 與 13 題題目，以一千名左右大學生施測所得的內部一致性信度分別為 0.857 與 0.915。本研究中，此二量表有三至九點之設計，並依點數有至多五種型式，包括等距、兩端、兩端加中間（即在量表之兩端與中間項加上標示語，其餘各項則未加文字說明）、正向與負向型量表。各種量表標示語乃根據翁（民 87）研究中以連續區間法估計所得的心理量尺值建構，依量表型式選取合宜之標示語。譬如五點等距量表各題反應的標示語為：完全不符合、大致不符合、普通、大致符合、完全符合。五點正向型的標示語為：完全不符合、普通、部分符合、相當符合、完全符合。五點負向型的標示語為：完全不符合、相當不符合、不太符合、普通、完全符合。共計有 29 種量表，每一量表中 25 題的順序都一樣，乃兩量表的題目交叉排列，CO 量表的題目為奇數題，DE 量表的題目為偶數題與最末一題。本研究的有效樣本共包含 3,265 名學生，其中男生 1,424 人，女生 1,839 人，有 2 人性別缺答；大學一年級 1,056 人，二年級 1,051 人，三年級 780 人，四年級 376 人，年級缺答者 2 人，分佈於 21 所大學院校之文理法等十多個學院。二十九種型式之量表乃隨機分派給所有受試者，每種量表至

少有 99 名受試者，至多有 138 受試者。除了上述型式量表外，各受試者均作各題的百分比評定，即請其自 0 至 100 中選擇一個最能描述該敘述句符合他（她）程度的整數填入。

四、結果與討論

表一列出各型式量表的受試者人數，以及量表總分的平均數和標準差。為比較評定量表型式對量表平均數之影響，乃分別就 CO 與 DE 兩量表之各點數量表進行單因子變異數分析。由於本研究共進行 14 個變異數分析，故將檢定之顯著水準設為 .01，以避免整體研究的第一類錯誤機率過高。在 CO 量表方面，除了三點 ($p > .05$) 與六點 ($p < .05$) 外，標示語型式均會影響受試者反應 ($p < .01$)。然而 DE 量表除了九點外 ($p < .01$)，受試者反應均不受標示語影響 ($p > .05$)。進一步的事後比較發現，CO 量表標示語型式的影響效果主要來自正向型與負向型量表平均數之差異，等距型、兩端型、兩端加中間型的平均數差別不大。DE 量表變異數分析的結果雖然大都不顯著，但檢視各型式量表的平均數仍可發現正向型的平均數皆比負向型低，只是其差異並不大，顯示正負型量表的操弄仍具效果，只是效果較弱。CO 正向型量表的標準差亦較其他各型式量表為高，負向型量表的標準差則偏低。由於此量表絕大多數受試者反應偏向正向反應，正向型量表的標示語設計即有效地區辨受試者在正向反應程度上的個別差異。由表一亦可看出，DE 量表的變異數比 CO 量表的變異數大，顯示受試者在 CO 量表上反應的一致性較高，在 DE 量表上反應的個別差異較大，變異數較大亦減低 DE 量表變異數分析達到統計顯著之可能性。標示語型式對受試者反應的結果顯示，量表標示語對受試者反應的影響主要來自正向型與負向型之操

弄，且其效果隨量表內容或者受試者反應之異質性而異，受試者反應較具同質性的量表對標示語的改變可能較敏感。

表二列出各型式量表的內部一致性係數 Cronbach α 。結果顯示，DE 量表的內部一致性信度並不因量表標示語型式而改變，CO 量表的 α 值則受量表標示語影響，正向型量尺的 α 值均較高，點數增加時尤然，其他四種型式量表的差異則較不明顯，此可能源於正向型設計提高了 CO 量表的受試者反應變異量。由於 CO 量表的題間相關較低，正向型量表設計提高了分數的變異量，增加題間相關，因而提高量表的內部一致性信度。DE 量表的題間相關較高，量表總分變異量亦較高，標示語型式的改變並未產生顯著影響，而維持相當恆定的內部一致性信度。表二中等距型與兩端型的結果與翁(民 88)的結果除了 CO 量表八點等距型與九點兩端型外均相差不大。顯示在量表點數高時，Cronbach α 在不同樣本的一致性可能較不穩定，此與 Lissitz 與 Green (1975) 的模擬研究結果不盡一致，兩研究之結果相左或與 Lissitz 與 Green 假設觀察之點資料為單一分配 (uniform distribution) 有關，需要再進一步探討。

此研究結果顯示，評定量尺的標示語與點數可能影響受試者反應，這些影響隨量表內容而異。而且，量尺標示語與點數對受試者反應的影響可能取決於量表總分的變異量。量表總分變異量高時，受試者反應較不受量表標示語影響，點數對量表內部一致性信度亦較無影響。量表總分變異量高表示受試者反應具異質性，或者量表題目具有高度區辨力，研究者或宜在編製量表時盡量提高題目的區辨力。量表總分變異量高可能因為各題變異量大，也有可能是由於題間相關較高，未來宜系統性地探討量表總分變異量、題目變異量、題間相關、題目區辨力等因素對受試者在評定量表反應與評定量表心理計量特徵之影

響，以了解各因素的重要性。此議題過去雖亦有研究探討 (Jenkins & Taber, 1977; Lissitz & Green, 1975; 吳, 民 85)，然其或討論連續變項，或對觀察變項的分配假設與一般實徵資料不甚相符，故實宜再深入研究探討。

五、計畫成果自評

本研究執行內容與原計畫相符，並且克服蒐集大量受試者資料之困難，除了九點負向型量表少一名受試者外，不僅收足且超出原計畫預計研究人數。預期完成之目標可謂達成，研究成果在評定量尺使用廣泛的國內兼具學術及應用價值，且可與國外研究比較，應適合於國內外學術期刊發表。

六、參考文獻

- 吳瑞屯 (民85)。影響 內部一致性係數的因素。「中華心理學刊」，卷38：51-59。
- 柯永河(民83)。同一量尺，類似受試，不同作答方式會產生甚麼測驗結果？「測驗年刊」，卷41：55-72。
- 胡志偉、彭昭英、沈永正、楊金龍(民78)。常用中文機率詞所代表的意義。「中華心理學刊」，卷31：1-6。
- 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：NSC 88-2413-H-002-010。
- 翁儷禎 (民87)。評定量表標示語之心理量尺值研究：頻率及同意度詞。「中華心理學刊」，卷40：73-86。
- 翁儷禎 (民85)。「頻率及同意度副詞的心理量尺值研究」。國科會專題研究計劃：NSC 84-2421-H-002-032。
- 黃恆獎(民75)。「問卷調查量度方法之研究 - 以Likert量表為例」。國立台灣大學商學研究所碩士論文。
- 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。
- Aiken, L. R. (1983). Number of response categories and statistics on a teacher rating scale. Educational and Psychological Measurement, 43, 397-401.
- Bendig, A. W. (1953). The reliability of self-ratings as a function of the amount of verbal anchoring and of the number of categories. Journal of Applied Psychology, 37, 38-41.
- Bendig, A. W. (1955). Rater reliability and the heterogeneity of the scale anchors. Journal of Applied Psychology, 39, 37-39.
- Champney, H., & Marshall, H. (1939). Optimal refinement of the rating scale. Journal of Applied Psychology, 23, 323-331.
- Churchill, G. A., Jr., & Peter, J. P. (1984). Research design effects on the reliability of rating scales: A meta-analysis. Journal of Marketing Research, 21, 360-375.
- Cliff, N. (1959). Adverbs as multipliers. Psychological Review, 66, 27-44.
- French-Lazovik, G., & Gibson, C. L. (1984). Effects of verbally labeled anchor points on the distributional parameters of rating measures. Applied Psychological Measurement, 8, 49-57.
- Hakel, M. (1968). How often is often? American Psychologist, 23, 533-534.
- Halpin, G., Halpin, G., & Arbet, S. (1994). Effects of number and type of response choices on internal consistency reliability. Perceptual and Motor Skills, 79, 928-930.
- Hancock, G. R., & Klockars, A. J. (1991). The effect of scale manipulations on validity: Targetting frequency rating scales for anticipated performance

- levels. Applied Ergonomics, 22, 147-154.
- 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.
- Jones, L. V., & Thurstone, L. L. (1955). The psychophysics of semantics: An experimental investigation. Journal of Applied Psychology, 39, 31-36.
- Klockars, A. J., & Hancock, G. R. (1993). Manipulations of evaluative rating scales to increase validity. Psychological Reports, 73, 1059-1066.
- Komorita, S. S., & Graham, W. K. (1965). Number of scale points and the reliability of scales. Educational and Psychological Measurement, 15, 987-995.
- Klockars, A. J., & Yamagishi, M. (1988). The influence of labels and positions in rating scales. Journal of Educational Measurement, 25, 85-96.
- Lam, T. C. M., & Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. Journal of Educational Measurement, 19, 317-322.
- Lehmann, D. R., & Hulbert, J. (1972). Are three-point scales always good enough? Journal of Marketing Research, 9, 444-446.
- Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 140.
- 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.
- Masters, J. R. (1974). The relationship between number of response categories and reliability of Likert-type questionnaires. Journal of Educational Measurement, 11, 49-53.
- Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. Educational and Psychological Measurement, 31, 657-674.
- Newstead, S. E., & Arnold, J. (1989). The effect of response format on ratings of teaching. Educational and Psychological Measurement, 49, 33-43.
- Oaster, T. R. F. (1989). Number of alternatives per choice point and stability of Likert-type scales. Perceptual and Motor Skills, 68, 549-550.
- Spector, P. E. (1976). Choosing response categories for summated rating scales. Journal of Applied Psychology, 61, 374-375.
- Symonds, P. M. (1924). On the loss of reliability in ratings due to coarseness of the scale. Journal of Experimental Psychology, 7, 456-461.
- Wildt, A. R., & Mazis, M. B. (1978). Determinants of scale response: Label versus position. Journal of Marketing Research, 15, 261-267.
- Wong, C.-S., Chuen, K.-C., & Fung, M.-Y. (1993). Differences between odd and even number of response scale: Some empirical evidence. Chinese Journal of Psychology, 35, 75-86.

表一

各量表總分之平均數與標準差(點量表)

表		分量	CO		DE	
			平均數	標準差	平均數	標準差
三點	等距(n=118)		29.2966	3.4451	29.6695	5.1492
	兩端(n=110)		29.5364	2.8532	30.1000	4.5674
四點	等距(n=112)		36.4911	4.0135	36.8214	5.4512
	兩端(n=107)		36.0841	4.4449	37.2056	5.6980
	正向(n=114)		34.7632	5.2652	36.6842	6.4332
	負向(n=119)		37.7731	4.5351	37.6723	5.4697
五點	等距(n=108)		43.9352	5.0627	45.1296	6.9700
	兩端(n=123)		45.9593	5.0072	46.3902	7.1834
	兩端+中間(n=117)		44.6325	5.6623	44.9829	6.8191
	正向(n=115)		42.3565	5.8267	44.8261	7.7598
	負向(n=117)		47.3077	4.6562	45.4103	7.8599
六點	等距(n=110)		53.2000	6.9346	54.0273	8.7425
	兩端(n=110)		52.4636	7.5001	52.8909	8.7329
	正向(n=113)		52.0088	8.2435	53.2743	9.7553
	負向(n=112)		54.6875	6.2760	54.2500	8.7760
七點	等距(n=111)		60.8468	7.9065	61.8378	10.4695
	兩端(n=112)		60.5089	8.0022	61.1786	10.8405
	兩端+中間(n=115)		59.7913	8.6820	59.6783	11.0100
	正向(n=115)		58.3913	9.3899	60.0783	11.4953
	負向(n=113)		63.9912	6.9994	63.1416	9.9318
八點	等距(n=108)		69.6852	8.2978	69.3426	12.5103
	兩端(n=109)		70.0826	10.6023	70.0092	13.6975
	正向(n=107)		64.8991	11.5998	67.8716	11.6588
	負向(n=104)		74.1538	7.9876	68.6058	11.7955
九點	等距(n=138)		76.4161	10.3821	77.2993	14.6083
	兩端(n=102)		77.1782	10.0602	77.9406	15.3908
	兩端+中間(n=121)		78.6942	11.8539	80.9917	13.0483
	正向(n=106)		73.0283	13.7207	74.6415	14.5138
	負向(n=99)		83.2222	8.6302	80.4343	11.3139

表二

各量表內部一致性信度係數 Cronbach α

量表型式	分量表	點量表		百分比量表	
		CO	DE	CO	DE
三點	等距	0.69	0.88	0.77	0.90
	兩端	0.57	0.85	0.76	0.91
四點	等距	0.69	0.85	0.81	0.90
	兩端	0.73	0.85	0.82	0.92
	正向	0.80	0.89	0.84	0.92
	負向	0.78	0.85	0.84	0.90
五點	等距	0.73	0.89	0.82	0.93
	兩端	0.65	0.85	0.79	0.91
	兩端+中間	0.77	0.85	0.80	0.90
	正向	0.76	0.88	0.82	0.93
	負向	0.68	0.90	0.76	0.91
六點	等距	0.80	0.89	0.85	0.93
	兩端	0.80	0.88	0.83	0.90
	正向	0.83	0.90	0.84	0.92
	負向	0.77	0.90	0.84	0.92
七點	等距	0.79	0.90	0.85	0.93
	兩端	0.74	0.89	0.71	0.85
	兩端+中間	0.80	0.88	0.86	0.92
	正向	0.81	0.89	0.68	0.81
	負向	0.75	0.90	0.83	0.92
八點	等距	0.74	0.88	0.79	0.93
	兩端	0.78	0.88	0.83	0.92
	正向	0.86	0.86	0.86	0.89
	負向	0.78	0.88	0.85	0.91
九點	等距	0.81	0.92	0.85	0.92
	兩端	0.73	0.90	0.80	0.94
	兩端+中間	0.81	0.86	0.85	0.91
	正向	0.84	0.86	0.85	0.91
	負向	0.72	0.83	0.84	0.90