

研究論文

選項設計之調查方法差異： 以生活品質測量為例*

廖培珊**

摘要

過去研究指出數量多的選項設計在面對面訪談中較為可行，然而對於選項標示語或不平衡式量尺設計則尚無深入討論。本研究以態度量表之選項設計為主體，來分析電話訪問與面對面訪談中，不同的選項設計之結果是否有差異。資料來源包含以態度量表進行隨機分派實驗設計之 2010 年及 2011 年電訪資料，以及 2010 年社會變遷基本調查之面訪資料。

在樣本結構部分，面訪與電訪的次樣本進行內部比較時均無差異；然而比較相同選項設計的面訪與電訪資料時，則發現個人月收入及工作狀況有明顯差別。以 IRT 之等級反應模型的分析發現，不同設計在面訪的表現均較電訪要佳；若整體考量題目與選項設計結合後的表現時，則以四點平衡量尺的鑑別力最佳。此外，標示語的設計效果會因資料蒐集方法的不同而有所差異，強語氣的標示語使面訪受訪者

* 本研究為國家科學委員會補助專題研究計畫「態度量表之設計：問題困難度與方法效應」（計畫編號：NSC 100-2410-H-001-076-MY2）之部分研究成果。論文初稿曾發表於「2012 調查研究方法與應用學術研討會」，文中若有任何疏漏由作者自行負責。

** 中央研究院人文社會科學研究中心調查研究專題中心副研究員，psliao@gate.sinica.edu.tw，台北市南港區研究院路二段 128 號，02-2789-8183。

往非極端選項靠攏的情形比電訪更為明顯。進一步以人口變項對「個人生活品質」之 IRT 分數進行迴歸分析時，亦可看出方法效應。

關鍵詞：方法效應、中間選項、標示語、生活品質

Mode Effects on Response Design: An Example of Qualify-of-Life Measures

Pei-Shan Liao*

ABSTRACT

Questionnaire design often differs among various survey modes in order to obtain better data quality. Previous studies have indicated that a higher number of response categories may perform better in face-to-face surveys than in telephone surveys. Few have examined the influence of response labels and unbalanced scale design on survey answers. This study aims to compare the responses between face-to-face and telephone surveys with respect to response design of attitudinal scales. Data used for this study are drawn from the 2010 Taiwan Social Change Survey and telephone surveys collected in 2010 and 2011 using a randomized experimental design to collect data from representative samples aged 18 years or older. By measuring the construct of individual quality of life (QoL), four types of response scales with different combinations of response number and response labels were examined to distinguish both the intensity and direction of responses.

Individual characteristics between the two modes indicated that the sub-samples were different in terms of monthly income and working status. Results of the graded response model suggested that the response design used in this study performed better in face-to-

* Associate Research Fellow, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica

face surveys than in telephone surveys. Also, a balanced scale with four response categories and less intense labels had the best discriminating ability among the four designs. When compared to respondents to telephone surveys, respondents to face-to-face surveys were less likely to provide strongly-labeled answers, which can be seen a tendency of social desirability due to more interpersonal interactions. Findings of linear regression on the IRT scores of QoL, however, were in favor of telephone surveys, with higher values of chi-square. An intertwining effect of response design and data collection mode was also found. A discussion of the results is provided.

Keywords: mode effect, mid-point response, response labels, quality of life

一、前言

問卷調查被用以瞭解民眾對於主觀現象的想法或態度，常見的調查資料蒐集方法包含由訪員進行的電話訪談與面對面訪談，以及自填問卷形式的郵寄問卷與網路調查等。由於不同的資料蒐集方法可能產生方法效應(mode effect)，進一步影響了資料分布與回答類型(response pattern) (Rockwood et al. 1997; Schaeffer and Presser 2003)；而選項設計的內涵也會因應資料蒐集方法之不同而改變，以獲取較佳之資料品質(瞿海源主編 2007)，因此在調查研究的方法論(methodology)部分，方法效應與選項設計均是重要的研究議題；但是二者的交互作用對於資料品質之影響，則尚缺乏實證依據。

以問卷中常見的態度量表而言，選項設計的組成內涵包括選項的數量與標示語等。一般由資料變異量的角度來討論時，選項數量的多

寡為重要考量；面對面訪談的方法研究指出，用以評斷回應強度的量尺若具備較多的選項數量有助於提升信度，因此建議較合適的選項數量為五到九類 (Alwin 1997; Alwin and Krosnick 1991; Weng 2004)，同時也認為正向與負向選項數量相同之平衡式量尺 (balanced scale) 設計的表现應較佳。然而在無法提供視覺支援 (如提示卡) 的電話訪談中，選項過多會使得受訪者不易理解或記憶選項內容，而增加訪談的困難度，因此四個或五個選項是在電話訪談時較常被使用的量尺設計 (Dawes 2008; Schaeffer and Charng 1991; Schaeffer and Presser 2003)。

與選項數量相關之研究議題，還有中間選項 (mid-point) 包含與否，以及相對應之選項標示語 (label) 的內容。既有的文獻中雖已有不少研究成果，然而並未達成一致之結論。過去研究利用面對面訪談所蒐集之具代表性樣本的大型學術調查資料進行分析 (廖培珊 2010)，其發現當選項數量相同時，不同的選項標示語明顯造成回答分布的差異；雖無法斷言語氣較溫和的標示語 (例如「很符合」) 和語氣較強烈者 (例如「非常符合」) 何者較能明確反應受訪者的態度，但前者之回答分配較趨近於常態，符合常用統計模型之假設，而使資料較利於讓研究者直接分析。然而，另一研究以實驗設計方式所蒐集之電話訪談資料，卻有著不盡相同之發現 (Liao 2014)；其中語氣較強的標示語 (例如「非常快樂」) 之表現並不比語氣較溫和 (例如「很快樂」) 的標示語表現稍差，此一結果與過去研究並不一致。

為了確認資料蒐集方法與選項設計的不同組合可能對資料產生的影響，並避免題目差異造成的偏誤，本研究以相同的題目與選項設計，來比較不同資料蒐集方法所獲得的回答類型；除可瞭解因方法效應而產生的可能差異外，亦有機會從方法效應的角度，來檢視選項設計的類型在不同資料蒐集方法中的適合度。本研究利用隨機分派實驗

設計所獲得的調查訪問資料，來探索相同題目與選項設計之條件下，面對面訪談與電話訪談對於選項設計所可能產生的影響，以及不同選項設計（含數量及標示語）在兩種資料蒐集方法上適用性之表現。由於過去的研究已指出電話訪談的選項數量不宜過多，因此對於這二種資料蒐集方法，本研究均以限定數量的選項數來進行資料蒐集，以減低選項過多的可能影響。

二、相關文獻回顧

（一）選項數量與中間選項

有關態度量表在設計上所需要的選項數量，過去已有許多實證研究加以討論（參見 Schaeffer and Presser 2003; 廖培珊 2010），在資料變異量以及信度的考量下，一致性的結論是選項數量不應少於三個（Preston and Colman 2000），在實務中則可見到多達十一個的選項設計（參見廖培珊 2010）。但是對於認知能力較差的受訪者而言，過多的選項會讓他們不容易分辨選項間的差異，進而增加其回答的困難度（Holbrook et al. 2006），因此會改以較少的數量來設計量尺，以免影響資料內涵。一般而言，所建議的選項數量在五個到九個之間。近期以自填問卷形式所進行的研究則認為五個選項最為合適（Hamby and Levine 2016; Wakita et al. 2012），因為數量越多則受訪者對選項的理解（perception）會越不正確。

除了選項數量多寡之外，多數研究亦關切中間選項（middle response）的提供與否對於受訪者回答類型與分布的影響，以及中間選項的實質意義是否具備中間位置之意涵。中間選項的存在與否，連帶影響了其他具實質意義之選項數量的決定；以正向與負向選項數量

相同的平衡量尺而言，若包含了中間選項會使得選項數量成為奇數；若否，則數量為偶數，而相對應的選項標示語通常亦需隨之調整。

以態度量表常見的李式量表（Likert scale）或衍生形式而言，量尺的中間選項包含「沒什麼同意不同意」（neither agree nor disagree）或「普通」等；然而此一中間選項可能有兩種不同的含意：一為真實的中立（neutral）態度，另一則為某種無反應的不知道、無意見或無法選擇（Bishop 1987; Moors 2007; Raaijmakers et al. 2000; Schuman and Presser 1981）。對於中間選項的回答主要可從兩種不同的理論觀點予以解釋，第一種為最適化（optimal），是指受試者受限於能力，在搜尋思索過去經驗與記憶後，仍無法提供有意義的實質回答，因此該論點傾向於將中間選項視為無反應的一種。第二種理論觀點則為satisficing，可視為「敷衍應答」（或譯為「合理的滿意」，參見廖培珊 2010），其認為受試者僅願意稍微嘗試搜尋資訊與經驗，並迅速判斷後，再提供一個自認為訪答雙方均可接受的回答，而不會盡全力去思索問題後再予以回應（Krosnick et al. 2002; Tourangeau et al. 2000），因此沒有明確方向性的中間選項便容易被受訪者所提及。

若將中間選項視為量尺兩端點的中間點而給予計分時，由實證研究的結果可看出其在量尺中二個極端測量之間，扮演了中間位置的角色（Kroh 2007），顯示中間選項有其存在之實質意義。過去研究指出：當訪問過程中主動提供中間選項時，或者僅在敘述題目時提醒，都同樣會有較高比例的受訪者選擇中間選項（廖培珊 2010; Bishop 1987; Schuman and Presser 1981）；因此建議不主動讓受訪者有此選擇，以避免其他實質答案的回應比例降低。然而，以實驗設計進行的電話訪問結果則指出，在不主動提示中間選項的情況下，選擇此一回答的比例相對稀少，以致於不易從統計分析的角度來評量該選項是否可發揮做

為中間點之功能 (Liao 2014)。由於一般的量尺設計通常已納入「不知道」、「拒答」等無反應選項，中間選項做為量尺的中間點的角色顯而易見，因此「敷衍應答」的理論觀點應較適用於解釋受訪者對於中間選項的回答行為。由過去對於中間選項的深入討論，似乎顯示選項設計宜以具備相同數量的正向與負向回答之平衡量尺來標示，以減低測量誤差。不過，當用以評量正向的心理狀態時，則常因為回答分布的偏頗，而建議以不平衡的選項量尺來設計 (Kalmijn et al. 2011; Liao 2014)，因此對於選項的標示，仍須視問卷題目而定，因為選項標示語強度的效果亦會受到題目內容的敏感性及措辭所影響 (Lam and Stevens 1994)。

(二) 選項標示語

用於態度量表之量尺所包含的內容，除了各選項有相對應的數值外，大多亦包含選項標示語。過去的研究以探討選項的語氣強度為主，由心理語言學 (psycholinguistics) 的觀點來看，國內外均有學者探究問卷量尺所使用的選項標示語，嘗試找出具區辨力的語詞，作為建置量尺之參考。Norman Cliff (1959) 以教育程度高、認知能力較佳的大學生為對象進行實驗分組，來探究英文中不同副詞與形容詞的組合應用在問卷量尺上的效果，因為副詞可作為「放大鏡」(multipliers) 來協助標定選項間的距離接近相等，然而其研究結果在問卷設計的實用性上並不高，而未被後續學者所認可或廣為使用 (Schaeffer and Charng 1991)。在中文問卷的部分，翁麗禎 (1998) 對於國內心理學研究中量表常使用的選項用詞進行分類，同時以實驗設計控制選項的順序效應，而得出各類別中不同選項標示語的相對位置；雖然同樣以高教育程度的大學生作為受試對象，然由於分析之資料來源是既有的問卷內

容，其適用性明顯較高，而結果亦同樣指出標示語的強度主要藉由副詞的變化來操控。

部分研究觀點將量尺視為以兩端點為上下限的連續體（response continuum），此一連續體包含數值與文字標示，受訪者可依標示語的內容來回答問題；或者僅考慮數值，而將各選項的相對距離視為均等分布（Lam and Klockars 1982; Schaeffer and Charng 1991）。不過，實證研究也指出，相較於有文字標示的量尺，只標示了數值或是僅於兩端點提供文字標示的選項比較容易使受訪者假定選項之間為等距（interval）尺度（Hamby and Levine 2016; Schaeffer and Presser 2003），對年長的受訪者來說，容易造成選擇上的困難（Fox et al. 2007; Holbrook et al. 2006），且其再測信度也較均以文字標示的選項為低（Weng 2004）。此外，當量尺選項同時包含數值與文字標示時，其測量本質實為順序（ordinal）尺度；在實證分析上，將量尺視為連續體的論點也僅獲得部分支持（Schaeffer and Charng 1991），但是無論何種觀點，均確認了標示語與數值的不同配對差異會影響受訪者的回答（Kroh 2007; O’Muircheartaigh et al. 1995; Schwarz et al. 1991）。

過去研究以實驗設計固定量尺的兩端點，來探討選項標示語及相對位置的變化對於回答分布的影響（Schaeffer and Charng 1991; Wildt and Mazis 1978），結果發現量表的回答類型確實受選項的標示語及所在位置影響；且前者的影響力高於後者（Wildt and Mazis 1978）。另外的研究則變化兩端點的選項標示語，來檢視選項標示語的設計對回答類型的影響，並指出兩端點之外的其他選項標示語的語氣強度或正負方向，亦直接影響量表分數的分布（Lam and Klockars 1982）。國內的實證分析則發現，在選項數量相同的情況下，相較於量尺兩端與鄰近選項用較強的語氣標示時，語氣較弱的標示語在回答分布上較趨於

常態（廖培珊 2010），因而適用於多數以常態分配為假定的統計模型。由於在相同選項數量的情況下，對應的數值不會改變，但標示語則否；由此可知，回答分布與選項標示之文字有較明顯的關連，些微的變化即可能影響受訪者的回答方向（Schaeffer and Charng 1991）。

（三）選項設計與方法效應

由於不同資料蒐集方法所具備的特性，選項設計的內容也會隨之改變，使得二者對於資料品質有相互影響的作用（Schaeffer and Presser 2003），進而提高測量誤差的出現機率（Rockwood et al. 1997）。不過，即便在相同題目與選項設計的條件下，不同資料蒐集方法仍可能導致不同的回答類型，例如自填問卷的形式能讓受訪者看到包含中間選項在內的所有選項內容，面對面訪談（以下簡稱「面訪」）中若以提示卡列出選項亦可有相同的效用；但是電話訪談（以下簡稱「電訪」）則否，僅能依據訪員所唸出的內容來判定可選答的項目。就同樣由訪員以口述方式直接與受訪者進行訪答的面訪與電訪而言，相同的選項設計仍可能對回答分布有不同的影響。

關於面訪與電訪的優缺點與二者之比較，在許多書籍與研究文獻中已有討論（瞿海源主編 2007; Bowling 2005; de Leeuw and van der Zouwen 1988; Díaz de Rada 2011; Holbrook et al. 2003）。一般而言，面訪被視為最不容易造成受訪者負擔的方式，因為受訪者僅需具備基本的聽說能力（Bowling 2005）；若無須在訪談過程呈現文字資料給受訪者看，則連閱讀能力都不必具備。相對於電訪而言，面訪的訪談步調較緩慢，可透過面對面的互動鼓勵受訪者充分思考後再回答，並可透過肢體語言讓受訪者與訪員之間發展較密切關係（rapport）；同時能在必要時以提示卡協助說明題目，有助於提高訪談的合作率及回答率，

並可進行較長且複雜的問卷訪談 (Bowling 2005; Holbrook et al. 2003; Tourangeau et al. 2000)。不過這些面訪的優點，也可能引起社會期望性偏誤 (social desirability bias) 的問題，因為在面訪的過程中，受訪者與訪員間發展出的密切關係使得前者希望博取後者的「好印象」；亦即其回答能反映出行為與態度符合社會規範的「不逾矩」，或者可獲得一般大眾的讚賞 (傅仰止、陸洛 2012; Tourangeau and Smith 1996; Tourangeau et al. 2000)。

相對來說，在電訪進行時，通常會考量一般人對於陌生人來電的耐受性；另一方面也希望可在短期間內蒐集大量的資料，而將訪問時間限縮得較短。如此一來，受訪者的認知過程可能會受時間壓力所影響而被壓縮，而容易出現「敷衍應答」的情形；同時，在特定議題上亦較面訪容易產生社會期望偏誤 (Holbrook et al. 2003; Voogt and Saris 2005)。不過，從實證資料的分析結果來看，過去研究在社會期望偏誤的方法效應上，並未獲得一致的發現。

以選項數量而言，由於面訪能夠以載有選項內容的提示卡輔助訪談的進行，讓受訪者有機會閱讀並思考所有選項內容，因此可提供數量較多的應答選擇 (Alwin 1997; Alwin and Krosnick 1991; Weng 2004)；相對地，電訪有賴於受訪者聆聽並記憶所有可能的選擇後，再做出決定，過多的選項不僅會造成理解上的困難 (Dawes 2008; Schaeffer and Charng 1991; Schaeffer and Presser 2003)，也容易出現因選項順序所造成的訪答效應 (田芳華 1996; Bishop and Smith 2001)，使得相同的選項設計應用在面訪與電訪時，可能對回答類型產生不同的影響效果。在此類以訪員進行訪談的資料蒐集方法上，過去研究建議對於態度量表可採取先問方向、再詢問意見強度的「開展」式 (unfold) 問法，以避免前述的訪答效應 (田芳華 1996; Bishop and Smith 2001; Schaeffer

and Presser 2003)。然而，即便控制了詢問方式，選項標示語的設計在面訪與電訪的結果仍可能有所差異；語氣強的標示語雖然在面訪的效果較弱語氣的標示語要佳，但是在電訪的表現則相去不遠（廖培珊 2010; Liao 2014）。由於過去的研究所檢視的標示語並不相同，亦未直接進行面訪與電訪的比較，因此標示語強度在不同資料蒐集方法上的適用性仍需更多的實證依據。

三、研究設計

（一）樣本來源與變項說明

爲了比較不同資料蒐集方法下，相同題目與選項設計之結果是否有所差異，本研究以電訪與面訪進行隨機分派之實驗設計來蒐集相關資料。電訪資料分別於 2010 年 1-2 月與 2011 年 3-4 月蒐集調查資料，以電腦輔助電話訪問系統（CATI）對台灣地區年滿十八歲以上且家中有電話之民眾進行訪問，抽樣設計爲分層兩階段系統抽樣（stratified systematic sampling），抽取電話號碼搭配後四碼隨機產生以增加涵蓋率，並於電話接通之家戶進行戶中抽樣，來選取合格的受訪者。訪談進行時，運用 CATI 系統可使相同題目、不同選項設計各類組合隨機出現的功能，以系統設定的方法，將受訪者隨機分派到不同的選項設計情境。原始問卷各包含六組不同的選項設計，本研究則選取其中與面訪資料相同題目與選項設計的四組次樣本來進行分析。

面訪資料來源爲 2010 年「台灣社會變遷基本調查計畫」（以下簡稱「社會變遷調查」）六期一次綜合組與環境組問卷中，部分以實驗設計進行之訪談資料。「社會變遷調查」爲每年進行之大型學術面對面訪談調查，每次包含兩組問卷主題，以分層三階段不等機率的抽樣設

計，選取台灣地區年滿十八歲以上民眾為受訪對象的代表性樣本（章英華等 2011）。由於以紙本進行之面訪調查在執行上不似 CATI 般，得以用電腦程式隨機分派，因此在實驗分派的設計上，2010 年社會變遷調查的兩組問卷各包含兩種選項設計，並以受訪者編號為單號或雙號來進行隨機分派，故共計有四種設計，且題目與選項內容均與電訪相同。

本研究所檢測的主題為量測生活滿意之題組，在題目相同的條件下，包含了不同選項數量、標示語、及平衡／不平衡之設計。考量社會變遷調查歷年問卷之選項設計內容，以及在電訪的適用性，用以進行本研究之設計主要為選項數量為四類與五類之量尺。其中四類選項包含兩種不同的標示語設計；五類選項則包含平衡（含中間選項）與不平衡量尺兩種。面訪的訪談過程與電訪進行的方式相同，以「開展」式詢問法進行而無提示卡輔助；同時，平衡量尺的設計並不會主動提供中間選項給受訪者做選擇。此外，由於訪問成功之樣本組成常與母體之人口結構有所差異，為避免造成分析資料時推論的偏差，故電訪與面訪資料均以「多變數反覆加權法」（raking）對成功完訪樣本進行加權，其中電訪各次樣本的完訪數量為 560 案至 635 案，面訪則為 919 案至 1117 案不等。

在所選取的題目部分，選擇與個人生活品質相關且可構成單一因素之題目進行分析，包含「整體來說，對自己目前的生活品質滿不滿意？」、「對自己目前的健康狀況滿不滿意？」、「對自己與朋友間的關係滿不滿意？」、「整體來說，覺得目前的日子過得快不快樂？」，及「整體來說，對目前住的社區／地方附近滿不滿意？」等五題，作為「個人生活品質」題組進行分析。

以整體生活品質滿意度為例，四種選項設計分別說明如後，其中

A、B 兩類在面訪問卷的部分屬於綜合組問卷（分別稱為 FA 和 FB），C、D 兩類則為環境組問卷內容（分別稱為 FC 和 FD）；電訪問卷則選擇與面訪問卷相同選項設計的次樣本，分別取自 2010 年與 2011 年的訪問資料（分別稱為 TA、TB、TC 及 TD）。

- A. 五點平衡量尺：(1) 很滿意；(2) 還算滿意；(3) 沒什麼滿意不滿意；(4) 不太滿意；(5) 很不滿意。
- B. 四點量尺 I：(1) 很滿意；(2) 還算滿意；(3) 不太滿意；(4) 很不滿意。
- C. 四點量尺 II：(1) 非常滿意；(2) 滿意；(3) 不滿意；(4) 非常不滿意。
- D. 五點不平衡量尺：(1) 非常滿意；(2) 滿意；(3) 還算滿意；(4) 不滿意；(5) 非常不滿意。

（二）資料分析與統計模型

在資料分析方面，先以敘述性統計來瞭解電訪與面訪資料中，選項設計相同的各組次樣本之受訪者人口特徵是否有所差異，同時以偏態係數檢視其回答的分布狀況；再以項目反應理論（item response theory, IRT）的多元計分模型（polytomous model）中，適用於順序尺度的等級反應模型（graded response model, GRM）來檢視不同的回答類型。項目反應理論（或譯為「試題反應理論」，見余民寧 2009）的基本假設有二：(1) 潛藏變項具備單一向度（unidimensionality）之特性；(2) 局部獨立性（local independence）（Embretson and Reise 2000），同時所產生之潛藏變項為連續尺度；後續的研究已發展出多向度的分析模型，並且考慮潛藏變項間可能存在的相關性。

等級反應模型屬於 IRT 多元計分模式的一種，為適用於順序性尺

度（例如李式量表）之邏輯模型（Embretson and Reise 2000; Samejima 1969, 1997），同時允許各題項具備不同的量尺設計。在 GRM 模型中，每一個題項 (i) 可用一個斜率參數 (a_i) 以及選項間的難度或位置參數 (threshold parameter) (b_{ij} ，其中 $j=1 \dots m_i$) 來表示。此一斜率或鑑別參數 a_i (discrimination parameter) 說明了題項與所測量之潛藏構念 (construct) 彼此間的關係強度，同時可衡量該題項在潛藏特質 θ 的連續向度上區辨出不同位階之受訪者的能力 (Reeve and Fayers 2005)。難度參數 b_{ij} (或稱為「閾限參數」，參見余民寧 2009) 則是指稱一個選項的困難度或位置；越低的參數值表示難度越低，亦即受訪者在回答時越容易從一個選項跨到另一個選項。

除了前述的單一向度與局部獨立性的假設外，GRM 還必須符合題項鑑別度同質性 (homogeneous discrimination) 的假設。模型適配度的檢定方式包含圖形與各指標估計值，可用以檢測題項與個人層級的效果 (Reeve and Fayers 2005)，其中受訪者的回答與潛藏特質間的關係，可透過一條連續遞增的函數來加以詮釋，此函數為項目特徵曲線 (item characteristic curve, ICC)；而以項目訊息曲線 (item information curve, IIC) 所呈現的項目訊息函數則可用來鑑別題項表現的好壞。對於 IRT 模型適配度的檢定，包含對於模型假設、參數估計值、以及模型預測力的檢定 (余民寧 2009)；對於等級反應模型，則可考量概似估計值 (likelihood estimate) (Samejima 1997)，雖然此一估計量對於多元計分模型的適用性不高 (Reise et al. 1993)。

除了等級反應模型外，其他用以分析具順序性尺度之多元計分的 IRT 模式，還包含部分計分模式 (partial credit model, PCM)、一般性部分計分模式 (general partial credit model, GPCM) 與評定量尺模式 (rating scale model, RSM) 等 (Baker et al. 2000; Bond and Fox 2007;

Cook et al. 1999; Maydeu-Olivares et al. 1994; Ostini and Nering 2006)，這些模型屬於 Rasch 家族模型，被稱為除總模型 (divide-by-total model) (余民寧 2009; Ostini and Nering 2006; Thissen and Steinberg 1986)，並且容許難度參數不具順序性。同時，PCM 模型中並未考量鑑別參數，RSM 則限制反應類別的所有形式 (包含選項數量及標示語) 均必須相同，否則應視為不同的測量。

相對而言，GRM 是從瑟式的判斷定律 (Thurstone's law of categorical judgement) 發展而來 (Maydeu-Olivares et al. 1994)，屬於差異模型 (difference model) 或累積模型 (cumulative model)。等級反應模型對於回答過程的假定，在於不同受試者因為潛藏特質 θ 的個別差異，使得每個人對於反應類別的關注力 (attraction) 也不同；而受試者關注或選擇某一類別的先決條件，是立基於已經注意到該類別之前的其他反應類別 (all prior category attractions)，但是均不採納的情況下 (Ostini and Nering 2006)，因此對於反應類別或選項間既有的難度順序是確認且不可更動的。因此，GRM 是以累積的機率函數 (cumulative probability function) 來界定選項的邊界位置，並可利用累積次數拆解相鄰的類別 (Baker et al. 2000)；這些特徵與 Rasch 家族的多元計分模型有所不同，因為後者是對相鄰的回答類別進行兩兩比較，且類別無法任意拆解或組合。

考量調查訪問的回答過程、選項標示語的差異以及既定的態度 (難度) 順序等因素，本研究參考過去主觀福祉相關研究的作法 (如 Baker et al. 2000; Chen et al. 2013; Liao 2014; Orlando et al. 2000; Sijtsma et al. 2008)，採用等級反應模型，並使用統計軟體 MULTILOG 7.3 (du Toit 2003) 來進行分析，其利用邊際最大概似法 (marginal maximum likelihood estimation algorithm) 來估計參數，各題選項之數值予以重

新編碼，以高分表示「滿意／快樂／健康」來進行參數值之設定。

在利用等級反應模型估計題項之鑑別度與難度參數後，各組次樣本再以前述五個題目來計算「個人生活品質」在 IRT 模型中的潛在特質分數 (scores) 作為依變項，再利用人口特徵作為自變項，來對個人生活品質進行線性迴歸分析，以呈現不同資料蒐集方法與不同選項設計的結果是否有所差異，但並不企圖瞭解人口特徵對於這些生活品質題組之解釋力。從 IRT 模型所推估得到的潛藏特質 θ 可用以做為量尺化的回答類型分數 (scaled response pattern scores)，進行直接分析，亦或直線或非直線的轉換應用，其優點在於可擷取各題項回答的所有資訊，有助於詮釋量尺的意涵 (余民寧 2009; Thissen et al. 1995)，過去研究也已有相關應用 (Orlando et al. 2000; Thissen et al. 1995; Sijtsma et al. 2008)。本研究在 GRM 的分析中，可依據題組中受訪者對於各題目的回答類型 (例如對這五題均回答最不滿意／不健康／不快樂者，其回答類型即為「11111」)，分別計算出其 θ 值 (du Toit 2003)，故以此數值作為「個人生活品質」之 IRT 分數。人口特徵變項則包含年齡、性別、婚姻狀況、教育程度、個人月收入及工作狀況等，年齡以實際年數來測量，性別是以女性作為對照組。婚姻狀況分為未婚、已婚及其他；教育程度為小學及以下、國 (初) 中、高中職、專科、大學及以上；個人月收入則包含 NT\$9,999 及以下、NT\$10,000—NT\$29,999、NT\$30,000—NT\$49,999、NT\$50,000—NT\$69,999、與 NT\$70,000 及以上等五類；工作狀況則分為全職、兼職、與其他，除了年齡與性別外，其餘人口特徵變項均以前述的最後一個類別作為對照組；變項中含有「不知道」、「拒答」等遺漏值之樣本則於分析中排除。

四、分析結果

（一）各次樣本結構分析

為確認各組次樣本為同一問卷調查中隨機分配的受訪者，首先檢視相同資料蒐集方法之次樣本在人口變項的部分是否有所差異。在面訪資料的部分，比較社會變遷兩組問卷中的次樣本，結果發現無論是綜合組或環境組，同組問卷的兩個次樣本彼此間在性別、年齡、教育程度、婚姻狀況、個人月收入及工作狀況等變項均無顯著差異（結果未呈現）。其次則為兩個年度的電訪次樣本，除了第一年度資料中，一組次樣本的全職工作者比例較低外（本研究不含此組次樣本），其餘變項均無差異。接著再比較相同選項設計的面訪次樣本與電訪次樣本，在加權處理後的之人口變項，由表 1 的卡方檢定結果可知，四組相同選項設計之面訪與電訪的受訪者，主要在個人月收入及工作狀況等變項的分配上具有顯著差異，其中面訪樣本中個人月收入低者（NT\$9,999 及以下）以及全職工作者均較電訪樣本要多。此外，第 B 組選項設計（四點量尺 I）的婚姻狀況在電訪與面訪的樣本分布上有所不同；其中電訪樣本有稍多的已婚、同居或配偶去世者，同時其離婚、分居者的比例較少；而性別、年齡與教育程度之分布在各組面訪與電訪樣本間則無明顯差異。

接續以個人生活品質之題組進行分析，表 2 所呈現者為面訪與電訪次樣本中，各項題目之偏態係數。結果顯示在這些次樣本當中，整體而言無論是電訪或面訪資料，均以五點平衡量尺的偏態係數最高（除了 FA 的「整體生活品質」外）；而以五點不平衡量尺設計的面訪資料（FD）則最為接近常態分配，因其偏態係數的絕對值相對於其他

表 1 面訪與電話人口特徵比較表

| 變項 | 組別 | FA | TA | FB | TB | FC | TC | FD | TD |
|--------------------|----|---------------------------|------|----------------------------|------|---------------------------|------|---------------------------|------|
| | | % | % | % | % | % | % | % | % |
| 性別 | | | | | | | | | |
| 男 | | 49.9 | 49.7 | 50.0 | 49.7 | 49.9 | 49.8 | 49.9 | 49.7 |
| 女 | | 50.1 | 50.3 | 50.0 | 50.3 | 50.1 | 50.2 | 50.1 | 50.3 |
| Pearson chi-square | | $\chi^2(1)=0.011, p=.919$ | | $\chi^2(1)=.010, p=.957$ | | $\chi^2(1)=.000, p=1.000$ | | $\chi^2(1)=.004, p=.959$ | |
| 年齡組 | | | | | | | | | |
| 18-19 歲 | | 2.6 | 0.9 | 2.1 | 4.4 | 2.7 | 2.3 | 2.4 | 1.2 |
| 20-29 歲 | | 20.6 | 21.2 | 20.9 | 18.7 | 20.4 | 20.7 | 20.7 | 21.1 |
| 30-39 歲 | | 20.5 | 20.4 | 20.5 | 20.4 | 20.6 | 20.5 | 20.4 | 20.4 |
| 40-49 歲 | | 20.6 | 20.4 | 20.6 | 20.4 | 20.6 | 20.7 | 20.6 | 20.4 |
| 50-59 歲 | | 17.5 | 18.0 | 17.6 | 17.6 | 17.5 | 17.5 | 17.6 | 17.9 |
| 60 歲以上 | | 18.2 | 19.1 | 18.3 | 18.5 | 18.2 | 18.2 | 18.3 | 18.9 |
| 不知道 | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 拒答 | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Pearson chi-square | | $\chi^2(5)=5.472, p=.361$ | | $\chi^2(5)=7.503, p=.186$ | | $\chi^2(5)=2.14, p=.999$ | | $\chi^2(5)=2.875, p=.719$ | |
| 婚姻狀況 | | | | | | | | | |
| 未婚 | | 33.1 | 29.4 | 30.4 | 30.4 | 31.2 | 27.1 | 31.5 | 27.1 |
| 已婚 | | 55.7 | 60.9 | 58.4 | 60.6 | 58.6 | 62.0 | 60.2 | 66.6 |
| 離婚、分居 | | 5.2 | 4.7 | 5.4 | 2.3 | 4.5 | 4.6 | 3.1 | 2.2 |
| 同居、配偶去世 | | 5.8 | 4.9 | 5.3 | 6.7 | 5.7 | 6.2 | 5.2 | 4.1 |
| 拒答 | | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Pearson chi-square | | $\chi^2(4)=4.843, p=.304$ | | $\chi^2(4)=11.969, p=.018$ | | $\chi^2(3)=2.898, p=.408$ | | $\chi^2(3)=7.139, p=.068$ | |

表 1 面訪與電訪人口特徵比較表 (續)

| | | | | | | | | | |
|--------------------|----------------------------|------|------|------|------|------|------|----------------------------|--|
| 個人月收入 | 34.3 | 19.6 | 32.7 | 22.9 | 30.5 | 24.5 | 30.1 | 19.4 | |
| NT9,999 及以下 | | | | | | | | | |
| NT10,000~NT29,999 | 30.7 | 34.7 | 31.3 | 37.2 | 33.0 | 37.8 | 33.2 | 38.6 | |
| NT30,000~NT49,999 | 20.3 | 26.2 | 20.2 | 21.5 | 21.8 | 18.7 | 20.9 | 23.5 | |
| NT50,000~NT69,999 | 7.7 | 9.5 | 9.8 | 9.9 | 8.2 | 8.3 | 8.9 | 7.0 | |
| NT70,000 及以上 | 6.0 | 6.3 | 5.3 | 5.5 | 4.9 | 7.6 | 5.4 | 6.8 | |
| 不知道 | 0.2 | 1.6 | 0.1 | 0.7 | 0.4 | 1.6 | 0.4 | 2.6 | |
| 拒答 | 0.8 | 2.2 | 0.7 | 2.3 | 1.1 | 1.4 | 1.2 | 2.2 | |
| Pearson chi-square | $\chi^2(6)=53.810, p=.000$ | | | | | | | $\chi^2(6)=42.811, p=.000$ | |
| 教育程度 | 18.0 | 17.5 | 18.1 | 17.9 | 18.1 | 17.9 | 18.1 | 17.4 | |
| 小學及以下 | | | | | | | | | |
| 國(初)中 | 13.8 | 13.5 | 13.8 | 13.7 | 13.8 | 13.9 | 13.7 | 13.6 | |
| 高中職 | 29.9 | 29.6 | 29.8 | 29.7 | 29.8 | 29.7 | 29.8 | 29.6 | |
| 專科 | 12.1 | 12.4 | 11.2 | 10.5 | 12.7 | 12.6 | 11.1 | 10.7 | |
| 大學 | 20.5 | 20.8 | 23.5 | 23.3 | 20.4 | 20.1 | 22.0 | 24.7 | |
| 研究所以上 | 5.6 | 5.8 | 3.5 | 4.4 | 5.1 | 5.5 | 5.1 | 3.7 | |
| 其他 | 0.1 | 0.3 | 0.0 | 0.5 | 0.1 | 0.4 | 0.2 | 0.2 | |
| 拒答 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | |
| Pearson chi-square | $\chi^2(7)=1.114, p=.993$ | | | | | | | $\chi^2(6)=2.945, p=.816$ | |
| 工作狀況 | 69.1 | 57.7 | 70.7 | 56.0 | 51.3 | 52.1 | 50.9 | 55.0 | |
| 全職 | | | | | | | | | |
| 兼職 | 9.0 | 8.7 | 9.3 | 9.6 | 4.2 | 9.3 | 4.4 | 9.5 | |
| 其他 | 21.9 | 33.6 | 20.1 | 34.5 | 44.5 | 38.6 | 44.7 | 35.4 | |
| Pearson chi-square | $\chi^2(2)=27.821, p=.000$ | | | | | | | $\chi^2(2)=25.481, p=.000$ | |
| Total | 977 | 635 | 919 | 563 | 1117 | 560 | 1092 | 587 | |

選項設計或資料蒐集方法而言較接近零。由於在五點平衡量尺的電訪資料 (TA) 中,「朋友關係」一題並無受訪者選擇中間選項,為避免後續 IRT 分析出現偏誤,此一次樣本 TA 在本題的回答重新編碼為四點量尺後再進行分析;編碼後之偏態係數列於括弧中。

表 2 各次樣本生活品質題組之偏態與主成分分析結果

| 選項設計 | 五點平衡量尺 | 四點量尺 I | 四點量尺 II | 五點不平衡量尺 |
|-------------|-------------------|-------------|-------------|-------------|
| 面訪 | | | | |
| 偏態係數 | (FA) | (FB) | (FC) | (FD) |
| 整體生活品質 | 1.094 | 0.717 | 1.542 | -0.531 |
| 健康狀況 | 0.873 | 0.361 | 0.86 | -0.341 |
| 朋友關係 | 1.038 | 0.174 | -0.229 | 0.1 |
| 快樂感 | 1.149 | 0.595 | 0.663 | -0.243 |
| 社區滿意度 | 1.275 | 0.601 | 0.895 | -0.178 |
| 特徵值 1/特徵值 2 | 2.083/0.906 | 2.203/0.859 | 2.102/0.947 | 2.039/0.927 |
| 電訪 | | | | |
| 偏態係數 | (TA) | (TB) | (TC) | (TD) |
| 整體生活品質 | 0.764 | 0.45 | 0.51 | 0.397 |
| 健康狀況 | 0.811 | 0.596 | 0.432 | 0.51 |
| 朋友關係 | 1.652 (-0.805) | 0.464 | 0.361 | 0.708 |
| 快樂感 | 0.961 | 0.601 | 0.653 | 0.46 |
| 社區滿意度 | 1.173 | 0.606 | 0.701 | 0.47 |
| 特徵值 1/特徵值 2 | 2.517/0.774 | 2.133/0.904 | 2.593/0.758 | 2.772/0.761 |

（二）等級反應模型分析結果

為確認所分析之變項應可組成單一潛藏因子，亦即符合項目反應理論單一向度之假定，因此對各次樣本進行主成分分析，結果同樣呈現於表 2；其中各次樣本資料均獲得單一個主要成分；同時滿足第一個特徵值大於第二個甚多，且第二個與其他的特徵值均相近之情況下，可視為符合 IRT 模型對單一向度之要求（余民寧 2009; Baker et al. 2000; Lord 1980），各單一主成分的解釋變異量為 40%至 55%之間。

以項目反應理論之 GRM 模型對於面訪與電訪之四種選項設計所估計出的參數列於表 3；其中 a 為題項鑑別參數，b 為難度參數。由表三的結果來看，多數題項的鑑別力有所差異，此為 GRM 模型的特徵；其中參數值最高者為面訪資料中四點量尺 I 的「快樂感」(3.04)，最低者則是電訪四點量尺 I 的「朋友關係」(0.28)。從資料蒐集方法的角度來看，各類選項設計在面訪資料的鑑別力幾乎都優於電訪資料，僅五點平衡量尺的「社區滿意度」及四點量尺 II 與五點不平衡量尺的「快樂感」除外。這四類選項設計中，四點量尺 I 所設計的問卷題組在面訪資料 (FB) 中具有最佳之鑑別力，以 IIC 所呈現的項目訊息函數曲線相比較之結果，亦支持此一發現（資料未呈現）；而電訪資料的部分則以五點平衡量尺的設計較佳。

當比較不同的選項數量時，由五點平衡量尺 (FA vs TA) 與四點量尺 I (FB vs TB) 的結果發現，無論是否包含中間選項，鑑別力均以面訪資料的表現較佳，且會因題項內容的差異而各有高低；¹ 其中

1 「朋友關係」一題的比較僅限於面訪資料 (FA 與 FB)，因電訪的 TA 樣本中並無回答中間選項之受訪者。

四點量尺 I 的鑑別力在面訪資料與電訪資料的差異較大。若比較相同選項數量、不同標示語時，以四點量尺的兩種設計來看，語氣較弱的設計 I（例如「很滿意」）的整體鑑別力在面訪的表現優於電訪；但是語氣較強的設計 II（例如「非常滿意」）則在兩種資料蒐集方法的表現相近。至於五點不平衡量尺與四點量尺 II 的結果，則在不同資料蒐集方法上較為接近，僅「快樂感」一題例外；以其鑑別力的表現來看，不平衡量尺的設計在電訪（2.31）優於面訪（0.85），而平衡量尺的部分則在兩種資料蒐集方法上相近。若僅以整體選項設計的表現而言，五點平衡量尺的設計在大部分題項的鑑別力表現優於其他選項設計。

由於各組的選項數量不盡相同，因此難度參數 b 於各模型中分別為三或四個（即選項數量減一），由表 3 可知，這些難度參數在潛藏變項的各題目均妥適地由負到正分布。以包含了中間選項的五點平衡量尺來看，面訪樣本（FA）在「整體生活品質」、「健康狀況」及「快樂感」等題項的前三個難度參數值（ $b(1)$ - $b(3)$ ）比電訪樣本（TA）要高，而最後一個難度參數（ $b(4)$ ）則是電訪樣本較高；換言之，電訪受訪者對於這些題項比較容易回答「還算滿意」及「不太滿意」，同時也有較高的機率回答中間選項「沒什麼滿意不滿意」；而面訪受訪者回答「很不滿意」的機率則比電訪受訪者要高。此外，五點平衡量尺的中間選項與其鄰近選項之困難度參數變化不大（即 $b(2)$ 和 $b(3)$ ），此應為回答中間選項的數量過少所致，無論面訪或電訪資料均出現此一現象；而 TA 樣本中「朋友關係」滿意度一題並無受訪者回答中間選項，因此需將資料重新編碼再進行分析，使得其難度參數 $b(2)$ 和 $b(3)$ 的意涵與其他題項不同（表 3）。至於同樣為語氣較弱之選項標示文字但不包含中間選項的四點量尺 I，其面訪樣本（FB）與電訪樣本（TB）的回答趨勢與五點平衡量尺的結果相似，同時並不因為題

表 3 「個人生活品質」題組之 IRT 參數

| 五點平衡量尺：面訪 (FA) | | | | | | | | | | | |
|--|-------|------|--------|------|--------|------|-------|------|-------|------|-------|
| 參數 | 題目 | | 整體生活品質 | | 健康狀況 | | 朋友關係 | | 快樂感 | | 社區滿意度 |
| | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | |
| a | 2.31 | .14 | 1.11 | .08 | 0.97 | .08 | 2.45 | .15 | 0.87 | .08 | |
| b(1) | -1.83 | .14 | -2.67 | .26 | -6.52 | .16 | -2.06 | .16 | -3.84 | .43 | |
| b(2) | -0.66 | .07 | -0.88 | .11 | -3.29 | .34 | -0.98 | .08 | -2.00 | .24 | |
| b(3) | -0.44 | .06 | -0.63 | .12 | -2.25 | .23 | -0.55 | .06 | -1.56 | .21 | |
| b(4) | 1.78 | .07 | 2.24 | .14 | 1.56 | .11 | 1.49 | .06 | 2.05 | .16 | |
| 五點平衡量尺：電訪 (TA) | | | | | | | | | | | |
| 參數 | 題目 | | 整體生活品質 | | 健康狀況 | | 朋友關係 | | 快樂感 | | 社區滿意度 |
| | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | |
| a | 1.13 | .08 | 0.58 | .05 | 0.80 | .06 | 1.19 | .08 | 0.95 | .04 | |
| b(1) | -3.25 | .23 | -5.09 | .53 | -6.60 | .50 | -3.47 | .23 | -2.89 | .20 | |
| b(2) | -1.03 | .15 | -1.66 | .27 | -4.65 | .28 | -1.27 | .16 | -1.10 | .12 | |
| b(3) | -0.88 | .15 | -1.61 | .26 | 1.23 | .18 | -1.20 | .15 | -1.04 | .12 | |
| b(4) | 3.58 | .17 | 3.72 | .29 | — | — | 2.90 | .14 | 2.73 | .13 | |
| -2 Log likelihood=-14119.5 及 -24486.8 (朋友關係) | | | | | | | | | | | |
| 四點量尺 I：面訪 (FB) | | | | | | | | | | | |
| 參數 | 題目 | | 整體生活品質 | | 健康狀況 | | 朋友關係 | | 快樂感 | | 社區滿意度 |
| | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | |
| a | 1.98 | .12 | 1.43 | .10 | 1.20 | .11 | 3.04 | .21 | 0.99 | .09 | |
| b(1) | -2.19 | .14 | -2.63 | .21 | -4.61 | .55 | -2.37 | .13 | -3.99 | .42 | |
| b(2) | -1.15 | .08 | -0.87 | .08 | -2.91 | .26 | -1.37 | .07 | -2.02 | .20 | |
| b(3) | 1.19 | .08 | 1.41 | .11 | 1.03 | .12 | 0.75 | .05 | 1.24 | .14 | |
| 四點量尺 I：電訪 (TB) | | | | | | | | | | | |
| 參數 | 題目 | | 整體生活品質 | | 健康狀況 | | 朋友關係 | | 快樂感 | | 社區滿意度 |
| | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | |
| a | 0.90 | .06 | 0.33 | .04 | 0.28 | .06 | 1.11 | .10 | 0.37 | .05 | |
| b(1) | -4.16 | .27 | -8.77 | 1.06 | -17.66 | 3.04 | -3.96 | .22 | -9.11 | 1.17 | |
| b(2) | -1.30 | .19 | -3.32 | .51 | -8.85 | 1.33 | -1.92 | .17 | -4.45 | .57 | |
| b(3) | 3.99 | .26 | 5.42 | .67 | 2.55 | .47 | 3.11 | .18 | 4.68 | .63 | |
| -2 Log likelihood=-11089.3 | | | | | | | | | | | |

表3 「個人生活品質」題組之IRT參數(續)

| 四點量尺 II：面訪 (FC) | | | | | | | | | | | | |
|----------------------------|-------|------|--------|------|--------|------|-------|------|--------|------|-------|--|
| 參數 | 題目 | | 整體生活品質 | | 健康狀況 | | 朋友關係 | | 快樂感 | | 社區滿意度 | |
| | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | | |
| a | 0.99 | .08 | 0.52 | .04 | 0.49 | .05 | 1.56 | .14 | 0.45 | .04 | | |
| b(1) | -6.74 | .45 | -10.12 | .99 | -14.49 | 2.55 | -6.03 | .36 | -11.35 | 1.23 | | |
| b(2) | -3.57 | .16 | -3.60 | .27 | -7.79 | .72 | -3.82 | .12 | -5.71 | .49 | | |
| b(3) | 4.92 | .24 | 6.91 | .54 | 5.89 | .46 | 4.04 | .14 | 6.81 | .61 | | |
| 四點量尺 II：電訪 (TC) | | | | | | | | | | | | |
| 參數 | 題目 | | 整體生活品質 | | 健康狀況 | | 朋友關係 | | 快樂感 | | 社區滿意度 | |
| | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | | |
| a | 0.73 | .06 | 0.38 | .04 | 0.43 | .05 | 1.75 | .16 | 0.34 | .04 | | |
| b(1) | -4.76 | .38 | -9.52 | 1.28 | -11.91 | 1.91 | -4.05 | .17 | -10.65 | .57 | | |
| b(2) | -1.50 | .22 | -3.38 | .49 | -6.48 | .77 | -2.00 | .13 | -4.50 | .67 | | |
| b(3) | 4.74 | .38 | 5.86 | .77 | 4.24 | .57 | 3.42 | .16 | 6.84 | 1.01 | | |
| -2 Log likelihood=-23219.0 | | | | | | | | | | | | |
| 五點不平衡量尺：面訪 (FD) | | | | | | | | | | | | |
| 參數 | 題目 | | 整體生活品質 | | 健康狀況 | | 朋友關係 | | 快樂感 | | 社區滿意度 | |
| | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | | |
| a | 0.67 | .04 | 0.43 | .03 | 0.49 | .03 | 0.85 | .05 | 0.50 | .03 | | |
| b(1) | -9.33 | .78 | -10.44 | .97 | -12.96 | 1.56 | -8.38 | .72 | -12.88 | 1.40 | | |
| b(2) | -5.44 | .31 | -5.15 | .41 | -8.90 | .72 | -4.97 | .24 | -6.62 | .47 | | |
| b(3) | 1.66 | .17 | 2.05 | .25 | 0.77 | .19 | 0.61 | .13 | 0.53 | .19 | | |
| b(4) | 6.26 | .39 | 8.11 | .66 | 6.38 | .47 | 5.16 | .28 | 7.00 | .51 | | |
| 五點不平衡量尺：電訪 (TD) | | | | | | | | | | | | |
| 參數 | 題目 | | 整體生活品質 | | 健康狀況 | | 朋友關係 | | 快樂感 | | 社區滿意度 | |
| | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | Est. | S.E. | | |
| a | 0.56 | .05 | 0.31 | .05 | 0.32 | .04 | 2.31 | .20 | 0.31 | .06 | | |
| b(1) | -6.02 | .56 | -10.68 | 1.59 | -14.81 | 2.40 | -4.09 | .14 | -10.33 | 1.46 | | |
| b(2) | -2.57 | .31 | -4.21 | .68 | -9.24 | 1.31 | -2.66 | .11 | -5.42 | .78 | | |
| b(3) | 0.38 | .23 | -0.32 | .40 | -3.76 | .60 | 0.09 | .11 | 0.27 | .39 | | |
| b(4) | 5.51 | .50 | 8.12 | 1.16 | 4.93 | .79 | 3.55 | .13 | 7.57 | 1.07 | | |
| 2 Log likelihood=-16661.5 | | | | | | | | | | | | |

項的內容而有差異；在四點量尺 I 的次樣本中，電訪樣本選答「還算滿意」及「不太滿意」的機率較高，而面訪樣本則是較容易回答「很不滿意」。因此，當以五點平衡量尺與四點量尺 I 來比較不同的選項數量時，可發現難度參數的分布趨勢主要在不同的資料蒐集方法上有所差異；包含中間選項與否並不會有顯著影響。

另一方面，當選項設計為端點語氣較強的四點量尺 II 時，面訪樣本 (FC) 的難度參數 $b(1)$ 及 $b(2)$ 數值均低於電訪樣本 (TC)，表示前者比後者在各題均有較高的機率回答「滿意」或「不滿意」；而回答「非常不滿意」的難度則在兩種資料蒐集方法上相去無幾，此與四點量尺 I 的結果不同。由於快樂感或滿意度多為「偏向快樂／滿意」之回答分布，因此相較之下，可看出選項數量同為四個的時候，弱語氣的標示語設計在電訪中回答極端負選項（即「很不滿意」）與停留在極端正選項（即「很滿意」）的機會比面訪低；而強語氣的標示語雖然讓兩種資料蒐集方法的受訪者回答極端的負選項（即「非常不滿意」），以及停留在極端正選項（即「非常滿意」）的機會都相當低，但是此兩類機率在面訪中又略低於電訪。

至於五點不平衡量尺的設計，目的便是為了使回答較為接近常態分配，² 因此前兩個難度參數 ($b(1)$ 和 $b(2)$) 均為正向回答的難易度。從表 3 來看，雖然面訪與電訪的難度參數則依題項不同而各有高低，但是選擇正向回答的機率都相當高；其中「整體生活品質」、「快樂感」及「社區滿意度」三題，面訪樣本 (FD) 有較高的機率回答「滿意」或「還算滿意」；但是對於「朋友關係」則是電訪樣本 (TD) 較容易回答「滿意」、「還算滿意」、或「不滿意」，在「健康狀況」一題

2 參見 Liao (2014)。

的難度參數則無明顯差別。若與相同標示語的四點平衡量尺相比較，則同樣是回答極端負選項（即「非常不滿意」）的機率最低；並且在多數題項中，面訪回答極端負值的難度又高於電訪（亦即前者的回答機率較低）；此外，無論是否為平衡量尺，面訪從第一個選項換到第二個選項的機率，也比電訪要高，因此在難度參數的分布上，同樣呈現出方法效應，而不平衡量尺的影響則需視題項而定。從前述的結果可知，無論是否包含中間選項或搭配不同的選項標示語，電訪樣本的回答趨勢與面訪樣本不盡相同，可看出選項設計和資料蒐集方法的交互影響；整體模型的表現，則以四點量尺 I 為最佳（ $-2\log$ likelihood 值最低）。

（三）項目反應理論分數之迴歸分析結果

最後利用五個題目來計算「個人生活品質」的 IRT 分數，並以此作為依變項，利用性別、年齡、婚姻狀況、教育程度、月收入及工作狀況等人口特徵作為自變項，來對個人生活品質進行線性迴歸分析，此一分析僅用以呈現不同選項設計在電訪及面訪的推估結果是否有所差異，並不企圖瞭解人口特徵對於這些生活品質題組之 IRT 分數的解釋力。由於人口變項呈現於表 1 之部分類別比例偏低，因此排除無反應之回答後，於迴歸模型中重新歸類，除了年齡外，各變項的參照組分別為女性、其他（包含離婚、分居、配偶去世、同居及其他狀況等）、大學以上、個人月收入 NT50,000 及以上，及非全職或兼職之其他工作狀況者；同時為便於比較，所呈現之迴歸分析結果為標準化係數（表 4）。

從表 4 的結果可看出面訪資料與電訪資料在各選項設計之結果有所差異，除了工作狀況之外，其他人口變項對於個人生活品質分數的

表4 個人生活品質之IRT量表分數迴歸分析(續)

| 標準化係數 變項 | 五點平衡量尺 | | 四點量尺 I | | 四點量尺 II | | 五點不平衡量尺 | |
|-------------------|---------|--------|--------|---------|---------|------|---------|---------|
| | FA | TA | FB | TB | FC | TC | FD | TD |
| 個人月收入 | | | | | | | | |
| NT9,999 及以下 | -.29*** | -.28** | -.14 | -.48*** | -.08 | -.06 | -.05 | -.41*** |
| NT10,000—NT29,999 | -.23** | -.22* | -.10 | -.32*** | -.03 | -.01 | -.03 | -.31** |
| NT30,000—NT49,999 | -.22*** | -.18* | -.03 | -.25*** | -.06 | -.01 | -.02 | -.26** |
| NT50,000—NT69,999 | -.11* | .01 | -.02 | -.14* | .00 | .07 | .01 | -.12* |
| NT70,000 及以上 | — | — | — | — | — | — | — | — |
| 工作狀況 | | | | | | | | |
| 全職 | .04 | .04 | .05 | .02 | -.06 | -.10 | -.03 | .07 |
| 兼職 | -.03 | .04 | .00 | .02 | -.06 | .09 | .03 | -.00 |
| 其他 | — | — | — | — | — | — | — | — |
| R ² | .04 | .09 | .05 | .15 | .03 | .08 | .03 | .08 |
| n | 966 | 605 | 915 | 538 | 1100 | 540 | 1076 | 557 |

a: 此類別包含離婚、分居、配偶去世、同居及其他狀況等。

*: p<.01; **: p<.05; ***: p<.001.

影響效果，在面訪與電訪資料中均有所不同。在這些效果當中，年齡及其平方項的效果與過去研究發現一致，亦即年齡與生活品質呈 U 型的曲線關係，僅於五點平衡量尺的面訪資料（FA）中並未出現顯著效果。此外，性別與個人生活品質的關連性在各次樣本中亦呈現一致性；除了 FA 與 FD 外，男性的個人生活品質分數高於女性。婚姻狀況的影響效果則因選項設計與資料蒐集方法的不同而有所差異，同時在影響方向上並不一致；類似的結果亦可從教育程度的結果中看到，不過其影響效果較不顯著。相對地，個人月收入與生活品質的關連性在面訪資料中，僅於五點平衡量尺的 FA 次樣本中出現顯著效果；但是在電訪資料中，則是除了四點量尺 II 之外，於 TA、TB 與 TD 三個次樣本中均達到顯著，工作狀況則在不同的選項設計或資料蒐集方法中均未出現顯著效果。至於各迴歸模型之解釋力則以電訪資料優於面訪資料，無論選項設計為何；不過細看可發現各次樣本中人口變項對於個人生活品質的解釋力大多低於 .10（除 TB 外），表示人口變項對於個人生活品質的解釋力十分有限，此一研究結果實與既有生活品質相關研究之發現相符。³

五、結論與討論

選項設計為調查研究方法之重要議題，然而過去研究聚焦於選項數量及中間選項之討論，較少著眼於不同資料蒐集方法的可能影響。本研究利用樣本隨機分派之實驗設計結果，以生活品質之測量為例，

3 過去研究一致指出人口特徵對於生活品質或主觀福祉的解釋力多在 10% 以下（參見 Hagerty 2000），因此需依研究議題而另行納入健康、認知、及社會心理因素等相關變項。

來比較面對面訪談與電腦輔助電話訪談在不同選項數量及標示語之設計，對於受訪者回答類型的影響是否有所差異。分析樣本基本人口特徵的結果指出，對不同選項設計之面訪資料與電訪資料進行內部次樣本的比較時，這些相同資料蒐集方法的次樣本可視為隨機抽選出的獨立樣本；然而在比較相同選項設計之面訪與電訪資料時，則可看出部分人口特徵之分布在資料加權後仍有所差異，顯示出不同資料蒐集方法對於調查結果之影響，此實為研究者採用不同資料蒐集方法時應具備之基本概念。

以個人生活品質之題組來比較不同選項設計的結果時，IRT 的等級反應模型分析發現，除了部分題項之外，不同選項設計的鑑別力在面訪的表現均較電訪要佳；其中並以四點量尺 I 在面訪資料的整體鑑別力最高。另一方面，由難度參數來看，在不同的選項設計條件下，電訪與面訪受訪者都較容易回答量尺中非兩極端值的選項，顯示電訪與面訪受訪者在回答的類型上的差異，主要是資料蒐集方法所造成。此外，當以人口變項對於個人生活品質的 IRT 分數進行迴歸分析時，模型解釋力則是以電訪資料較高，顯示這些題項在電訪的內在效度較佳。

當考量標示語的設計在不同資料蒐集方法的效果時，研究結果指出選項設計為語氣較弱的標示語（例如「很滿意」）時，面訪樣本的題項鑑別度表現較好，且回答量尺兩極端值的機會較高；相對來說，電訪樣本則較容易選擇非兩端的答案。另一方面，在面對語氣較強的標示語時，則無論是面訪或電訪的受訪者都不太容易選擇兩極端值的回答，而使得回答分布往中間趨近。過去研究在不同資料蒐集方法對選項設計的影響上，討論了面訪與電訪在選項數量上的差異（Schaeffer and Presser 2003），但是對於標示語的研究卻相對缺乏。本研究雖然在選項標示語的分析結果呈現出方法效應的存在，ICC 的比較亦可看

出不同資料蒐集方法的差異 (Reeves and Fayers 2005)，並指出面訪之選項設計應較適用語氣較弱的標示語，有利於回答的均衡分布，在一般以資料常態分布為假設的統計模型應用中也較為適切；但是在電訪的部分則無法看出明顯趨勢，因此對於不同的資料蒐集方法是否有各自較適用的標示語文字，仍有待更多的分析資料予以佐證。

在四類的選項設計中，其中的五點平衡量尺包含了中間選項，爲了避免造成「吸引效果」(廖培珊 2010; Bishop 1987; Schuman and Presser 1981)，本研究所分析之面訪資料與電訪資料在訪談過程中均不主動提示中間選項。分析的結果指出，無論資料蒐集方法爲何，此種不主動提供中間選項的開展式 (unfold) 問法，使得該選項的回答比例極低或甚至爲零；而 IRT 的參數估計結果也可看出在這種情況下，中間選項與後一選項的難度參數數值差異甚小。因此，在訪談過程中若未主動提示中間選項，即可避免可能的「吸引效果」，與過去的研究發現相符 (Liao 2014)，也確認了調查訪問在執行時，中間選項提供與否的可預期結果，同時減低「敷衍應答」可能造成的偏誤 (Krosnick et al. 2002; Tourangeau et al. 2000)。

至於不平衡量尺的設計，雖然以「快樂感」之單一測量爲例的研究指出，不平衡量尺的表現優於常見的平衡量尺設計 (Liao 2014)，本研究的電訪資料也支持此一發現，但是面訪資料的結果卻非如此，反而是平衡量尺的表現稍佳；同時，其他「個人生活品質」相關題項亦未獲得不平衡量尺設計較佳的結論。此一發現同樣指出了相同選項設計在不同資料蒐集方法的結果有所差異，爲二者的交互影響關係提供了另一證據。

由前述結果看來，不同資料蒐集方法的差異效果大於選項設計之內涵對於回答類型的影響；由於面訪是以較長時間進行面對面的人際

互動，受訪者可能有意識地希望讓訪員有好印象(傅仰止、陸洛 2012)，而增強其合作意願；另一方面，電訪受到的時間壓力通常較大，受訪者的認知過程也進而被壓縮，因此使得相同的題項在整體鑑別力的表現上是面訪優於電訪。至於選項的難易度分布，強語氣的標示語使面訪受訪者往非極端選項靠攏的情形比電訪更為明顯，顯示受訪者在面對訪員時可能為了符合社會取向的期望(楊國樞 1982；傅仰止、陸洛 2012)，而較不願意回答強語氣的極端選項，但是對於弱語氣的標示語則無此現象。由於標示語的些微變化便可能影響受訪者的回答(Schaeffer and Charng 1991)，對於可能出現較強社會期望反應的面訪而言，選項的設計宜更為謹慎。

參考文獻

- 田芳華，1996，〈自填與訪填對答題效應之影響〉。《調查研究》2: 59-88。
- 余民寧，2009，〈試題反應理論(IRT)及其應用〉。台北：心理出版社。
- 翁儷禎，1998，〈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》40(1): 73-86。
- 章英華、杜素豪、廖培珊主編，2011，《台灣社會變遷基本調查計畫第六期第一次調查計畫執行報告》。台北：中央研究院社會學研究所。
- 傅仰止、陸洛，2012，〈陌生人互動的社會期望反應：外在情境與個人屬性〉。頁 293-346，收錄於朱瑞玲、瞿海源、張苙雲編，《台灣的社會變遷 1985~2005：心理、價值與宗教，台灣社會變遷基本調查系列三之 2》。台北：中央研究院社會學研究所。
- 楊國樞，1982，〈心理學研究的中國化：層次與方向〉。頁 153-188，收錄於楊國樞、文崇一合編，《社會及行為科學研究的中國化》。台北：中央研究院民族學研究所。
- 廖培珊，2010，〈態度量表之選項標示語：調查資料之潛藏類別分析〉，《調查研究—方法與應用》24: 91-134。
- 瞿海源主編，2007，《調查研究方法》。台北：三民書局。
- Alwin, D. F., 1997, "Feeling Thermometers versus 7-Point Scales: Which Are Better?" *Sociological Methods & Research* 25(3): 318-340. doi: 10.1177/0049124197025003003.

- Alwin, D. F. and J. A. Krosnick, 1991, "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes." *Sociological Methods & Research* 20(1): 139-181. doi: 10.1177/0049124191020001005.
- Baker, J. G., J. B. Rounds, and M. A. Zevon, 2000, "A Comparison of Graded Response and Rasch Partial Credit Models with Subjective Well-Being." *Journal of Educational and Behavioral Statistics* 25(3): 253-270.
- Bishop, G. F., 1987, "Experiments with the Middle Response Alternative in Survey Questions." *Public Opinion Quarterly* 51(2): 220-232.
- Bishop, G. F. and A. Smith, 2001, "Response-Order Effects and the Early Gallup Split-Balots." *Public Opinion Quarterly* 65(4): 479-505.
- Bond, T. G. and C. M. Fox, 2007, *Applying the Rasch Model: Fundamental Measurement in the Human Science* (2nd ed.). New York: Routledge.
- Bowling, A., 2005, "Mode of Questionnaire Administration Can Have Serious Effects on Data Quality." *Journal of Public Health* 27(3): 281-291.
- Chen, S.-K., F.-M. Hwang, and S. S. J. Lin, 2013, "Satisfaction Ratings of QOLPAV: Psychometric Properties Based on the Graded Response Model." *Social Indicators Research* 110(1): 367-383.
- Cliff, Norman, 1959, "Adverbs as Multipliers." *Psychological Review* 66(1): 27-44.
- Cook, K., B. G. Dodd, and S. J. Fitzpatrick, 1999, "A Comparison of Three Polytomous Item Response Theory Models in the Context of Testlet Scoring." *Journal of Outcome Measurement* 3(1): 1-20.
- Dawes, J., 2008, "Do Data Characteristics Change According to the Number of Scale Points Used? An Experiment Using 5-Point, 7-Point and 10-Point Scales." *International Journal of Market Research* 50(1): 61-77.
- De Leeuw, E. D. and J. van der Zouwen, 1988, "Data Quality in Telephone and Face to Face Surveys: A Comparative Meta-Analysis." Pp. 283-299 in *Telephone Survey Methodology*, edited by Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, William L. Nicholls II, and Joseph Waksberg. New York: John Wiley & Sons.
- Diaz de Rada, Vidal, 2011, "Face-to-Face versus Telephone Surveys on Political Attitudes: A Comparative Analysis." *Quality and Quantity* 45(4): 817-827.
- Du Toit, Mathilda ed., 2003, *IRT from SSI: Bilog-mg, Multilog, Parscale, Testfact*. Lincolnwood, IL: Scientific Software International.
- Embretson, S. E. and S. P. Reise, 2000, *Item Response Theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fox, M. T., S. Sidani, and D. Streiner, 2007, "Using Standardized Survey Items with Older

- Adults Hospitalized for Chronic Illness.” *Research in Nursing & Health* 3(4): 468-481.
- Hagerty, M. R., 2000, “Social Comparisons of Income in One’s Community: Evidence from National Surveys of Incomes and Happiness.” *Journal of Personality and Social Psychology* 78(4): 764-771.
- Hamby, T. and D. S. Levine, 2016, “Response-Scale Formats and Psychological Distances between Categories.” *Applied Psychological Measurement* 40(1): 73-75.
- Holbrook, A. L., M. C. Green, and J. A. Krosnick, 2003, “Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias.” *Public Opinion Quarterly* 67(1): 79-125.
- Holbrook, A. L., Y. I. Cho, and T. Johnson, 2006, “The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties.” *Public Opinion Quarterly* 70(4): 565-595.
- Kalmijn, W. M., L. R. Arends, and R. Veenhoven, 2011, “Happiness Scale Interval Study: Methodological Considerations.” *Social Indicators Research* 102(3): 497-515.
- Kroh, M., 2007, “Measuring Left-Right Political Orientation: The Choice of Response Format.” *Public Opinion Quarterly* 71(2): 204-220.
- Krosnick, J. A., A. L. Holbrook, M. K. Berent, R. T. Carson, W. M. Hanemann, R. J. Kopp, R. C. Mitchell, S. Presser, P. A. Ruud, V. K. Smith, W. R. Moody, M. C. Green, and M. Conway, 2002, “The Impact of ‘No Opinion’ Response Options on Data Quality: Non-attitude Reduction or an Invitation to Satisfice?” *Public Opinion Quarterly* 66(3): 371-403.
- Lam, T. C. M. and A. J. Klockars, 1982, “Anchor Point Effects on the Equivalence of Questionnaire Items.” *Journal of Educational Measurement* 19(4): 317-322.
- Lam, T. C. M. and J. J. Stevens, 1994, “Effects of Content Polarization, Item Wording, and Rating Scale Width on Rating Response.” *Applied Measurement in Education* 7(2): 141-158.
- Liao, Pei-Shan, 2014, “More Happy or Less Unhappy? Comparison of the Balanced and Unbalanced Designs for the Response Scale of General Happiness.” *Journal of Happiness Studies* 15(6): 1407-1423.
- Lord, F. M., 1980, “Item Response Theory—Introduction and Preview.” Pp. 11-26 in *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ.: L. Erlbaum Associates.
- Maydeu-Olivares, A., F. Drasgow, and A. D. Mead, 1994, “Distinguishing among Parametric Item Response Models for Polychotomous Ordered Data.” *Applied Psychological Measurement* 18(3): 245-256.

- Moors, G., 2007, "Exploring the Effect of a Middle Response Category on Response Style in Attitude Measurement." *Quality & Quantity* 42(6): 779-794.
- O'Muircheartaigh, C., G. Gaskell, and D. B. Wright, 1995, "Weighing Anchors: Verbal and Numeric Labels for Response Scales." *Journal of Official Statistics* 11(3): 295-307.
- Orlando, M., C. D. Sherbourne, and D. Thissen, 2000, "Summed-Score Linking Using Item Response Theory: Application to Depression Measurement." *Psychological Assessment* 12(3): 354-359.
- Ostini, R. and M. L. Nering, 2006, *Polytomous Item Response Theory Models*. Thousand Oaks: Sage.
- Preston, C. C. and A. M. Colman, 2000, "Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences." *Acta Psychologica* 104(1): 1-15.
- Raaijmakers, Q. A. W., A. van Hoof, H. Hart, T. F. M. A. Verbogt, and W. A. M. Vollebergh, 2000, "Adolescents' Midpoint Responses on Likert-Type Scale Items: Neutral or Missing Values." *International Journal of Public Opinion Research* 12(2): 208-216.
- Reeve, B. B. and P. Fayers, 2005, "Applying Item Response Theory Modeling for Evaluating Questionnaire Item and Scale Properties." Pp. 55-73 in *Assessing Quality of Life in Clinical Trials, 2nd ed*, edited by Peter Fayers and Ron Hays. Oxford: Oxford University Press.
- Reise, S. P., K. F. Widaman, and R. H. Pugh, 1993, "Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance." *Psychological Bulletin* 114(3): 552-566.
- Rockwood, T. H., R. L. Sangster, and D. A. Dillman, 1997, "The Effect of Response Categories on Questionnaire Answers: Context and Mode Effects." *Sociological Methods & Research* 26(1): 118-140.
- Samejima, F., 1969, "Estimation of Latent Trait Ability Using a Response Pattern of Graded Scores." *Psychometrika Monograph* No. 17.
- , 1997, "Graded Response Model." Pp. 85-100 in *Handbook of Modern Item Response Theory*, edited by W. J. van der Linden and Ronald K. Hambleton. New York: Springer.
- Schaeffer, N. C. and H.-W. Charng, 1991, "Two Experiments in Simplifying Response Categories: Intensity Ratings and Behavioral Frequencies." *Sociological Perspectives* 34(2): 165-182.
- Schaeffer, N. C. and S. Presser, 2003, "The Science of Asking Questions." *Annual Review of Sociology* 29: 65-88.
- Schuman, H. and S. Presser, 1981, *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.

- Schwarz, N., B. Knäuper, Hans-J. Hippler, E. Noelle-Neumann, and L. Clark, 1991, "Rating Scales: Numeric Values May Change the Meaning of Scale Labels." *Public Opinion Quarterly* 55(4): 570-582.
- Sijtsma, K., W. H. M. Emons, S. Bouwmeester, I. Nyklíček, and L. D. Roorda, 2008, "Non-parametric IRT Analysis of Quality-of-Life Scales and Its Application to the World Health Organization Quality-of-Life Scale (WHOQOL-Bref)." *Quality of Life Research* 17(2): 275-290.
- Thissen, D. and L. Steinberg, 1986, "A Taxonomy of Item Response Models." *Psychometrika* 51(4): 567-577.
- Thissen, D., M. Pommerich, K. Billeaud, and V. S. L. Williams, 1995, "Item Response Theory for Scores on Tests Including Polytomous Items with Ordered Responses." *Applied Psychological Measurement* 19(1): 39-49.
- Tourangeau, R. and T. W. Smith, 1996, "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context." *Public Opinion Quarterly* 60(2): 275-304.
- Tourangeau, R., L. J. Rips, and K. Rasinski, 2000, *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.
- Voogt, R. J. J. and W. E. Saris, 2005, "Mixed Mode Designs: Finding the Balance between Nonresponse Bias and Mode Effects." *Journal of Official Statistics* 21(3): 367 -387.
- Wakita, T., N. Ueshima, and H. Noguchi, 2012, "Psychological Distance between Categories in the Likert Scale: Comparing Different Numbers of Options." *Educational and Psychological Measurement* 72(4): 533-546.
- Weng, L.-J., 2004, "Impact of the Number of Response Categories and Anchor Labels on Coefficient Alpha and Test-Retest Reliability." *Educational and Psychological Measurement* 64(6): 956-972.
- Wildt, A. R. and M. B. Mazis, 1978, "Determinants of Scale Response: Label versus Position." *Journal of Marketing Research* 15(2): 261-267.