

輔測定錨法與心理計量法之 效度評估： 以政治功效感的測量為例

張書憲* 黃旻華**

- 一、「項目功能差異」的問題
- 二、「輔測定錨法」和「心理計量法」
- 三、研究設計
- 四、效度評估分析
- 五、結論

在民意調查學界，長久以來「項目功能差異」(differential item functioning)的問題一直困擾研究者對於心理測量效度的信任，主要的問題出在受訪者對於問卷題目和量表的感受及回應，不但存在著「人際不可比較性」(interpersonal incomparability)的問題，同時對於題目本身是否能夠真正抓住研究者所欲測量的心理態度，往往令人感到懷疑。對於上述問題，心理及教育學界主流作法是以精巧的心理計量模型來克服測量上的問題，但在2004年的《美國政治科學論叢》(American Political Science Review)中Gary King等學者提出了「輔測定錨法」(anchoring vignettes)的

* 高雄餐旅學院餐旅行銷暨會展管理系副教授。E-mail: schang@mail.nkhc.edu.tw

** 美國德州農工大學政治系助理教授。E-mail: huang@politics.tamu.edu

投稿日期：二〇〇八年三月十一日；接受刊登日期：二〇〇八年七月十三日。

東吳政治學報/2008/第二十六卷第三期/頁 115-147。

測量方法，宣稱此法可以徹底解決「項目功能差異」的問題。為了評估兩法的相對優缺點，本文藉由三個大學系所 671 位學生為受試者，進行政治功效感的心理測驗，並以上述樣本進行兩種心理測量法的效度評估。結果發現，輔測定錨法在測量效度上並不比心理計量法為佳；更甚者，輔測定錨法存在嚴重的方法論問題，使得作者對於此法的實施抱持著相當保留的態度。

關鍵詞：輔測定錨法、心理計量學、項目功能差異、人際不可比較性

一、「項目功能差異」的問題

心理測量應用在不同的科學領域中是很普遍的，雖然具體測量標的不同，但目的都是在對於某種「潛在心理特質」（latent trait）進行測量（黃旻華，2006），比方說政治學中公民的政治功效感（黃信豪，2006）、社會學中民眾的幸福感（陸洛，1998）、公衛學界中自我主觀的健康評價（張彩秀、黃乾全，2002）、管理學界中的消費者滿意度等等（林宜靜等，2006）。顧名思義，由於潛在心理特質所指涉的是抽象層次上的主觀測量，也就是基於受訪的表意來抽離出研究者所關切的某種心理特質，因此受訪者對問題的回答變成了研究者進行測量的唯一資訊，問題的關鍵在於：研究者對於潛在心理特質的定義是客觀的，是希望達成某種人際之間共通的態度測量，但實際上不管從受訪者如何認知問卷題目、如何感受答項量表、以及許多主客觀影響心理測量的因素，都使得心理測量所得的資料本身具有高度主觀的色彩（Suchman and Jordan, 1990）。是故心理測量的原始資料本身，究竟能否達成「人際共通性」（interpersonal equivalence），攸關調查資料的測量效度良窳，以及應該如何正確使用資料等議題。¹

一般來說，針對一個潛在心理特質來進行測量，實務上多數研

1. 「人際共通性」是所有進行調查研究的學者所希望達成的目標，即儘管每個受訪者有各種不同的人格特質和心理傾向，施測的問卷題目都能將受訪者的真實態度找出來，並且反映到一個人際間可共通的量尺上，進行客觀的態度強弱比較。關於此概念，請參考 Brady（1989）。

究者會採用題組的方式來施測，² 在經過嚴謹的概念化理論探討之後，研究者會設計出題數不等的一組問卷題目，主張每個題目都反映了此潛在心理特質的程度，即假設其表面效度和內容效度成立，³ 然後對於受試者進行測量，之後再利用「因素分析法」來進行建構效度的分析。⁴ 如果建構效度符合研究者在概念化上的理論預期，則此題組被視為具有心理測量上的內在效度，反之則代表此題組至少有部份題目不適合當作測量的工具，在測量時必須剔除。⁵

上述作法是一般心理學、教育學界檢證問卷題目「內在效度」的方法。但長久以來許多學者質疑，即便「因素分析法」的結果可以

-
2. 以國內幾個政治學常用的資料庫來說，多半在問卷設計時都是以題組的方式，將問卷題目組織在幾個主要的理論概念下，可參見「台灣選舉與民主化調查」(TEDS)、「亞洲民主動態調查」(ABS)、「台灣社會變遷調查」(TSCS)、「世界價值觀調查台灣區調查」(WVS)等資料庫。
 3. 「表面效度」是指施測題目在語意表達上與研究者意圖測量的心理態度是否相符，表面效度並不需要理論的支持，但在直觀上是要合理的(Fink, 1995)；而就「內容效度」所指的是，施測的題目是否能夠真正量測到研究者所欲探求心理態度的確切內容(Fitzpatrick, 1983)。表面效度和內容效度高度相關，一個不具備表面效度的問題，很難具有內容效度，而具備內容效度的問題，往往會有相當比例的受訪者不懂或者是有不同理解。一般習用上，許多學者在提及表面效度時，同時也包含了對於內容效度的關切，所以雖然兩者指涉的意義有所不同，但很難完全分開，這一點請讀者謹記在心。
 4. 「建構效度」所指涉的是研究者基於理論而主張的心理測量模型在操作層次上被經驗資料所支持的程度(Cronbach and Meehl, 1955)，比方說，假定我們認為「社會資本」的概念是由「人際互信」、「互惠規範」和「公民參與網絡」三核心價值所構成(Putnam et al., 1993: 195-213)，因此在問卷中各設計了一題來進行測量，理論上「社會資本」作為測量模型中的潛在因子應該能夠很顯著地解釋「人際互信」、「互惠規範」和「公民參與網絡」三道題目受訪者回答的變異，如果是的話，代表測量的建構效度很好；如果不是的話，則建構效度就有問題。通常建構效度有問題不一定代表理論有誤，有可能是測量題目的操作化有問題，有可能是測量的執行過程有問題，當然也有可能就是理論的問題。
 5. 通常因素負荷是主要的判斷關鍵，因素負荷一般來說要高於0.5 或者是將近0.5 才可被接受，如果因素負荷低於0.45，則可認定測量題與潛在因子的相關度過低，測量的建構效度不佳(SAS Institute, 2000)。

自圓其說問卷設計滿足理論上的建構效度，由於「因素分析法」的數理推論是基於施測題目皆具備「表面效度」、「內容效度」，且都滿足「人際共通性」的假設上，因此其可信度受到很大的懷疑，認為心理測驗高度受到社會情境、文化因素、執行調查中不可預期因素和個人主觀特質等等的影響（Narens and Luce, 1983）。因此，任一問卷題目的回答其可信度都值得存疑，更遑論將其視為一題組來進行因素分析，這也無怪乎許多學者主觀上直接否定因素分析的效用並批評其所做的分析不過是「垃圾進，垃圾出」（garbage in, garbage out）（Wrigley, 1976: 5）。

仔細探究上面的批判，必須回歸到心理測量的核心議題：究竟應該如何克服受訪者因為對於問卷題目的不同認知及對於答項量表的不同感受，使得不同受訪者間的回答並不具備客觀的比較性，造成任一單題問卷測量皆沒有進行客觀數理推論的條件，使得任何心理計量的數學工具都失去了適用性，這正是所謂的「項目功能差異」（differential item functioning）的問題，⁶ 此問題不僅會發生在同一題不同人之間，即「人際不可比較性」（interpersonal incomparability），同時也可能發生在同一人不同題之間，即測量效度失效的問題，包括表面效度、內容效度和建構效度都可能發生錯誤。

嚴格來說，「人際不可比較性」在應用上造成的問題遠比「表面效度」、「內容效度」失效來得大，學界對於「表面效度」、「內

6. 在心理計量學中，通常項目功能差異的問題是以「團體」（group）為對象來討論，而不會下到「個人」（individual）的層次上（Embretson and Reise, 2000: 251），但 King et al. (2004: 192, 1n) 特別強調 DIF 在個人層次上的問題。關於 DIF 的定義和介紹，可參閱 King et al. 所引註的 Holland and Wainer (1993) 一書。

容效度」是否具有效力，除了對於受訪者所處社會系絡的深入理解外，就是利用「因素分析」在經驗上呈現資料符合「單面向性」(unidimensionality)的假設(Schmitt, 1996; Reckase, 1997)，或者符應原本理論上所預期的潛在因子結構。如果挑剔一些，可以利用「概似比檢驗」(likelihood ratio test)，將預期的潛在因子結構拿來跟其他可能成立的潛在因子結構相比，如果能夠在統計上展現最適性，一般認為測量效度的問題就不大(Du Toit, 2003: 796-797)。但相對地，「人際不可比較性」的問題並無法利用「因素分析」來解決，而除非心理測驗作答的主觀性能夠消除，或者找出統一的基點進行量測，否則「人際不可比較性」將使得單純利用敘述統計來描述測量結果的作法都失去意義。⁷

類似的憂慮正反映在許多學者並不相信心理測量，特別是以態度問題來進行潛在特質的分析，許多批評都直指態度問題過於抽象、題意難以理解、問題模稜兩可、量表選項難以感受並反映受訪者確切意向程度等等，因此主張心理測量時應該多使用客觀的行為指標，不然應該簡化量表選項(比方說從五點李克特量表變成二分量表)(黃旻華，2006：54-57)，或者將態度問題具體化以實例來說明。⁸ 這些作法明顯地都希望心理測量的結果能儘可能減少主觀

7. Gary King et al. (2004: 195-196) 以中國和墨西哥的調查資料，說明傳統測量方式所得的敘述統計是不可信的。首先他們主張中國是專制國家而墨西哥是民主國家，長久以來學者們都認為客觀上墨西哥人的政治功效感比中國人高，然後他們利用兩國未經調整過的自我政治功效感評估資料，進行敘述統計而發現居然中國人的政治功效感比墨西哥人高得多，因此透過「輔測定錨法」重新調整原始的測量值之後，最後果然得到墨西哥人政治功效感比中國人高的結果。這結果使得他們認為傳統測量方法的測量效度有很嚴重的問題，必須採用「輔測定錨法」來改善。關於「輔測定錨法」的介紹，正文第二節有詳細的說明。

8. 如果將問卷題目具體化，雖然可以降低題目的抽象層次，簡化語意，讓受訪者易於理解，但是往往具體化的問卷題目同時會在語義喪失原有概念中的某些元素，使其答案並不能真

性，進而降低人際不可比較性對於資料品質的影響。

下節針對「項目功能差異」問題，特別是人際不可比較性的部分來進行方法論上的討論，主要檢視的測量方法是「輔測定錨法」和「心理計量法」，我們將說明其數理推論基礎，並且解釋如何應用在實際測量上。

二、「輔測定錨法」和「心理計量法」

「輔測定錨法」(Anchoring Vignettes)是著名美國政治學者 Gary King等在2004年的《美國政治科學論叢》(American Political Science Review)中所提出，宣稱可以徹底解決「項目功能差異」問題的新測量方法(King et al., 2004; King and Wand, 2007)。⁹主要的想法是透過受訪者對於輔測問題的參照進行定錨，然後將受訪者自我陳述的態度強度置入其輔測問題的錨點中，來找尋其相對位置作為人際共通的普遍性測量。¹⁰換句話說，雖然受訪者對於輔測問題的回答也具有主觀性，但因為不牽涉到自己態度的評估，所以可以當作一客觀的參照點，然後將同時具備受訪者主觀性和真實態度的自

正能測到研究者所探求的心理態度。至於實務上究竟能否簡化和如何簡化問卷題目，必須靠研究者對其社會脈絡的理解來下判斷，並沒有一定的準則，但基本原則儘量能夠簡化越好(Parten, 1950)。

9. 關於「輔測定錨法」的介紹，除了正文中所引註的兩篇文章外，另外 Gary King 所建「輔測定錨法」的專門網站(The Anchoring Vignettes Website)，也提供許多豐富的資訊可讓讀者瀏覽和交流(King, 2007a)。
10. King et al. (2004)認為「定錨題」因為是對參照團體(第三人)所進行的政治功效感評估，所以並不會涉入自我評價時可能產生的問題，同時又可以將自我答題時的習慣辨識出來。如此一來，受訪者對參照團體的評價回答，就成了之後定位其自我陳述回答的比較基點，也就是所謂的「定錨點」。所以人際間的態度強弱要能比較，必須以這些定錨點為共同的量尺刻度來進行評量。

我陳述與之相比，則可以得出純然是真實態度的測量結果。這樣的結果在基於輔測題目具有絕對客觀性的前提下，是可以達成人際共通比較的目的，因為輔測題目的定錨正式阿基米德的支點，可以作為所有態度的客觀比較基礎 (King et al., 2004: 195)。

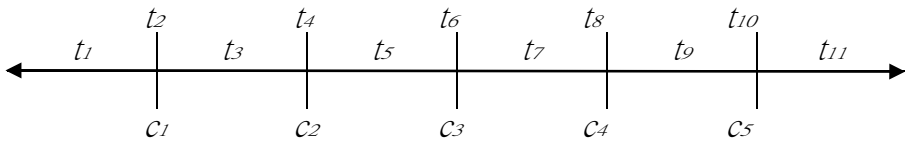
如果用符號來表示，比方說當受試者在回答某個輔測題目時受到的主觀性效果為 DIF_V ，而在回答自我態度陳述時受到的主觀性效果為 DIF_S ，而真實態度為 A ，所以現實呈現的結果為 $(DIF_S + A)$ 。若將自我態度陳述置入輔測題目的參照點時，可得態度測量為 $(DIF_S + A) - DIF_V$ 。在完美的狀態下，我們可以主張同一受訪者對於輔測題目和自我態度陳述問題有相同的主觀性效果，即 $DIF_V = DIF_S$ ，所以 $(DIF_S + A) - DIF_V = A$ ，也就是在輔測題有絕對客觀的假設前提下，我們對於每個人的態度測量都可以擺脫其主觀性，而達成人際共通的測量結果。¹¹

按此說法，一個輔測題目可以定一個測量錨點，得到三種不同結果，即大於、等於、小於此錨點。所以若一個心理特質題目的測量用了 c 個錨點，則將產生一個具有 $2c+1$ 種類別結果的量表，通常研究者會依潛在心理特質的概念將這 c 個輔測題目設計成具有不同難度，然後在設計答項時採用 k 點李克特量表，使得 $k \geq c$ 。如此一來，研究者就可利用受訪者對於 c 個輔測題目的回答，然後將其自我態度陳述題的回答以 $2c+1$ 的順序類別量表來歸類。

比方說針對一個潛在心理特質，研究者用了五個輔測題目，其難度分別為 $c_1 < c_2 < c_3 < c_4 < c_5$ ，在答項的部份也都使用相同的五點李克特量表，分別為 $k_1 < k_2 < k_3 < k_4 < k_5$ 。某位受訪者針對 $\{c_1, c_2, c_3, c_4, c_5\}$ 分別回答 $\{k_1, k_2, k_2, k_4, k_4\}$ 的答項，然後在自我態度陳

11. 這部分的說明參考自「輔測定錨法」網站 (King, 2007b)。

述題 c_6 中回答了 k_3 的答項, 那麼就代表此人的潛在心理特質是介於 c_3 和 c_4 之間, 因為 k_3 大於其對於 c_3 所回答的 k_2 , 但小於其對於 c_4 所回答的 k_4 。而這五個定錨點總共所產生的十一種順序類別 $\{t_1, t_2, \dots, t_{11}\}$ 分別圖一所示:



圖一 五個輔測題所形成的十一點李克特量表

因此對於每個受訪者的自我態度陳述, 輔測定錨法都是利用自我定錨方式對於態度測量進行調整, 最後以十一點的李克特量表來得出真實態度測量值, ¹²是故去除了使用原始資料時受到「人際不可比較性」的問題, 因為個人的主觀性已經在定錨比較的過程中消除了。¹³

輔測定錨法對於消除個人主觀性的主張, 至少必須建立在三個假設之上: (一) 每個輔測題都有相同的表面和內容效度, 即 $DIF_{v_1} = DIF_{v_2} = \dots = DIF_{v_n}$ 。(二) 輔測問題和自我陳述問題也具有相同的表面和內容效度, 即 $DIF_{v_i} = DIF_s$ 。(三) 受訪者與研究者對於

12. 在某些情況下, 受訪者的回答可能包括了這十一點中的好幾種可能性。比方說對五個輔測題回答 $\{k_1, k_2, k_3, k_4, k_5\}$ 的受訪者, 若在自己陳述題回答 k_2 , 那麼此受訪者的心理特質則介於 c_1 和 c_4 之間, 但因為此測量結果包括 t_3, t_4, t_5, t_6, t_7 五種可能的點估計值, 因此通常可以取算術平均數而取 t_5 為最後的測量值。

13. 這裏我們使用的是 Gary King et. al (2004) 所提的「無母數方法」(Non-parametric approach), 至於更複雜的統計模型, 我們認為除非是「無母數方法」有很好的施用效果, 讓我們對於資料的信心增強, 否則, 如果使用輔測定錨法之後還需要複雜的數學公式來思考, 那倒不如不用輔測定錨法, 至少還可以減低題數和簡化測量的作業時間。

輔測題目有邏輯一致的難易度感知，即至少是 $c_1 \leq c_2 \leq \dots \leq c_n$ 。如果第一點不滿足，輔測問題之間就無法比較，也不能定錨。如果第二點不滿足，自我陳述態度也無法置入輔測定錨的量表中。如果第三點不滿足，每個人的定錨將會南轅北轍，特別是在輔測題難度的排序上失去了可比較的基礎，也因此仍然存在著「人際不可比較性」的問題。

事實上這三點滿足與否完全是經驗層次的問題。Gary King等學者雖然舉出「政治功效感」在中國和墨西哥比較的例子來說明輔測定錨法的優越性（King et. al, 2004: 196），但是此法究竟能否適用在不同的社會系絡，以及是否有必要僅為了單項態度的測量而花那麼多的時間金錢成本，¹⁴ 同時仍然冒著機會不低的三項方法論風險，我們認為是有必要進行現實上的斟酌。接下來讓我們針對心理和教育學界常用的「心理計量法」來討論其如何面對「項目功能差異」的問題。

「心理計量法」是許多應用在心理測量上的數統模型統稱。基本上近二十年來在心理學和教育學界的主流典範是「項目反應理論」（item response theory），也就是將潛在心理特質的「能力得點」與問題難度的「位置參數」量表相結合而形成的相對測量，然後再輔以「鑑別度參數」和「猜測度參數」，構成了許多不同的非線性測量模型。其中單參數的模型只包含「位置參數」，又稱「拉西模式」（Rasch Mode），而雙參數的模型通常是將「鑑別度參數」導入，而可以針對不同題目進行效度分析，至於三參數的模型進一步討論到「猜測度參數」，特別是針對教育測驗中許多選擇題作答時的猜測效果進行估

14. 在 King et al. (2004) 對於政治功效感的施測中，所實行的輔測定錨法共用了五題輔測題和一題自我評估題來進行測量，等於是用了六個題目來進行原先只需單題的量測工作。

計。¹⁵

針對「項目功能差異」的問題，「二分 IRT 模型」在假設上是最能夠呼應「輔測定錨法」的心理計量模型，原因在於「二分 IRT 模型」主張任何單題問卷的態度強度測量都可能因為各種因素影響而不可信，因此在操作上無論原始問卷量表為何，都建議將其化簡成二分答項，比方說「同意\不同意」、「支持\不支持」、或「正面\非正面」等。這種對於單題測量資訊的保守運用，基本上就是將每一題定錨在態度的轉折點，然後只取此題意向在轉折點之上或下的測量資訊，進而利用二分 IRT 模型進行「單參數」、「雙參數」或「三參數」的心理計量分析（黃旻華，2006：55）。在本文中，作者將採用結合「類別因素分析」和「兩參數的二分 IRT 模型」之軟體 TESTFACT 進行心理計量法的分析，並且可以同時應用在「輔測定錨法」和傳統心理測量的施測題組中進行比較。

除此之外，另一個解決「項目功能差異」問題的想法，可藉由設定「團體間項目功能差異」的參數來評估其影響大小，但不主張對每題重新編碼成二分答項，而仍沿用原始的多分李克特量表。嚴格來說，心理計量法不可能按輔測定錨法針對每一個個人來矯正其「項目功能差異」的問題，但是可以主張這樣的問題在某種團體或特質的人中其效果比較不顯著，因此雖然無法針對每個人，但是我們可以針對團體間的「項目功能差異」問題來納入評估，所以只要 $DIF_{within} \ll DIF_{between}$ ，即組內的項目功能差異與組間相比小到一定程度，那麼基本上心理計量法的分析結果將會趨近於理想的輔助定錨法，至於這個假設會不會成立，同樣是屬於經驗層次的問題，必須取的經驗資料後才能得知。

15. 介紹心理計量學在政治學的應用，請參考 Poole (2008)、黃旻華 (2006)。

由於本文的焦點放在輔測定錨法與心理計量法的效度評估，並且主要是基於複製 Gary King 等學者的心理測量實驗，因此本文並不處理「國際間項目功能差異」在方法上多大程度能夠減輕「項目功能差異」的問題，¹⁶ 而是以主觀政治知識認知作為測量效標，來對兩種測量方法進行效標效度 (criterion validity) 的多變量迴歸分析，以評估兩測量法的內在效度。¹⁷

三、研究設計

本文將針對輔測定錨法和傳統心理測量方法同時進行施測，以取得實際的經驗資料來進行比較分析，由於兩種方法皆為強調避免「人際不可比較性」的問題，都旨在探索受訪者個人主觀性如何影響測量結果，而不同於一般心理測驗試圖對於某個群體或社會進行態度測量的點估計，加上客觀研究條件的限制，在樣本的選取上，施測的對象為台灣大學政治系、中山大學政治學研究所、以及高雄餐旅學院各系所 671 位的授課學生，群體間的同質和異質性相當明顯，可從事後分析中看出樣本特質，不過此樣本終究非代表性抽樣，因此結論僅具啟發性的價值，但希望能夠進行這兩種方法初步的評估和比較。¹⁸

16. 就作者所知，在常用的幾個 IRT 軟體中，如 TESTFACT, PARSCALE, MULTILOG, BILOG 中，並沒有同時可以執行「國際間項目功能差異」(DIF)和「因素得點」(Factor Scoring)的功能，因此這議題仍有待未來進一步的研究。

17. 「效標效度」是指以某概念的獨立測量值為標準(依變項)，然後將欲測試的量表測量值帶入迴歸模型中(自變項)，比較何者可以具有較高的預測能力(解釋力)，預測力越高的量表，代表效標效度越強(Cronbach and Meehl, 1955: 282)。

18. 讀者這裏可能對於施測對象有便利抽樣的質疑，如此的疑慮雖有其道理，但是回顧 Gary King et al. (2004) 在中國和墨西哥的主要實驗，其實都是隨 WHS (World Health Survey)

關於輔助定錨法的問卷，由於本文是重新複製 Gary King 等學者對於政治功效感的心理測驗，因此在問卷設計上，除了少數文句考量到中英文翻譯及國情差異外有些許變動，¹⁹ 整個問題是沿用 Gary King 等學者原始的設計，具體問卷題目（中文）如表一，而原始的英文問卷題目請參考 Gary King et al. (2004: 193-194)：²⁰

-
- 所進行的便利抽樣，並沒有真正嚴謹的抽樣設計以及明確的母體定義，因此本文也無從複製其樣本所欲代表的母群體。特別是中國大陸的飲用水問題，其實跟地區有很大的關係，而且一個省之內的變化會很大。如果真正在乎輔測題的脈絡意義，應該是要嚴格定義出母體，也就是對於飲用水有問題認知的人口，或者起碼明確將抽樣設計中此項因素在選取樣本時所造成的影響，比方說如何分層，是用農村和都市、行政區劃、地域、水域、抑或有不同程度缺水問題的地域群集來分類，但 Gary King et al. (2004) 並沒有說清楚。此外，在許多心理計量的方法研究中，對於施測對象的挑選，往往不以具備全國代表性的樣本為其研究目的，而是將其作為一個個具有相同生理機制的實驗對象，也因此沒有樣本代表性的顧慮。不過我們承認所使用樣本具有相當的同質性（年齡和學歷），以及明確的異質性（台大中山代表一般升學體系、高雄餐旅學院代表技職體系），也因此之後的多變量迴歸分析會加入這些變數來進行控制。基於啟發性的研究目的來說，如此的考量應可被接受。
19. 比方說台灣多數地區普遍已經沒有太多飲用水的問題，但是位於自來水管線末端的供水問題其實仍發生在不少地區，因此作者在此更動了問卷題目的文句，希望能夠符合台灣的社會狀況，類似的調整也出現在以立法委員取代原問卷中未指涉民代類型的廣義指稱。
 20. Gary King et al. (2004) 原始的文句，至少在中國的社會脈絡下是很難理解的，特別是其所宣稱的飲用水問題和趁選舉時節找民意代表陳情遊說。雖然中國許多地方的飲用水是沒有問題的，但普遍農村多飲用地下水，且基層農村選舉中的選舉和民代性質，與西方自由民主傳統下的地方自治差異甚大，因此很難想像 Gary King 等學者在中國施測的問卷題目如何傳達問題原意。關於這點，作者曾經向 Gary King 索取在中國的施測問卷，Gary King 回應其手邊並沒有問卷，但指可至 WHO (World Health Organization) 網站找到相關資訊，但經過作者再三確認之後，WHO 並沒有將施測的中文問卷公開陳列在網站上。

表一 輔測定錨法的問卷題目

請先讀完下面五段話，然後回答問卷第一至第五題：

- 一、**淑華**家位於自來水管線末端，常常飽受缺水之苦，她和她的鄰居決定在即將到來的立委選舉中支持一位承諾改善此事的反對黨候選人，由於淑華居住的區域中許多人都有相同的感受，因此看起來這位反對黨的立委候選人篤定會打敗現任者而贏得選舉。
- 二、**懷恩**家位於自來水管線末端，常常飽受缺水之苦，他和他的鄰居希望發起在請願書簽名的活動來引起地方的注意，他們計劃將此請願書在即將到來的立委選舉前分送該區的各黨參選人，並分別向他們表達嚴重關切之意。
- 三、**巴奈**家位於自來水管線末端，常常飽受缺水之苦，這是因為政府遲遲無法通過預算積極地來改善此問題，當地現任的執政黨立委對此事也漠不關心，在即將到來的立委選舉中，有一個反對黨的立委候選人已承諾改善此事，但巴奈認為投給這位候選人並不會有太大用處，因為當地的現任立委看起來應會連任成功。
- 四、**振中**家位於自來水管線末端，常常飽受缺水之苦，當地的有力人士其實有能力解決這個問題，主要是能說服自來水公司在當地重鋪管線並且建立一個加壓站，唯這些改善工程的施工會影響當地有力人士正在炒作的房價，因此他們並不想解決這個問題。
- 五、**妹芬**家位於自來水管線末端，常常飽受缺水之苦，她希望這問題能夠解決，但她並沒有投票權，而且似乎當地的民代和官員都不重視這個問題，所以她只有默默地氣在心理，希望有一天這個問題能夠好轉。

針對上面的五人[淑華、懷恩、巴奈、振中、妹芬]，請問你認為他們有多大的影響力可以引起當地政府官員正視自來水問題的嚴重性？請按下面的答項來表達你的判斷

(1) 根本沒有任何影響力 (2) 沒有什麼顯著的影響力 (3) 有一點影響力 (4) 有很大的影響力 (5) 有絕對的影響力 (6) 無法辨別其影響力。

V1 () 淑華

V2 () 懷恩

V3 () 巴奈

V4 () 振中

V5 () 妹芬

V6 () 對於你關切的議題，你認為你有多大的影響力可得到政府官員的重視。(1) 根本沒有任何影響力 (2) 沒有什麼顯著的影響力 (3) 有一點影響力 (4) 有很大的影響力 (5) 有絕對的影響力 (6) 不知道。

從上面的題目可看出，第一到第五題分別為五個輔測題，按題意研究者所設定的政治功效感程度分別為 $c_1 > c_2 > c_3 > c_4 > c_5$ ，答項為五點的李克特量表，答項類別順序為 $k_1 < k_2 < k_3 < k_4 < k_5$ ，第六題為自我態度陳述題，量表同輔測題，而所有題目都尚有「無法辯別其影響力」或「不知道」的答項以便受訪者回答。

至於政治功效感在傳統心理測驗的問卷題目，本文沿用政治學界最常使用的「台灣選舉與民主化調查」(TEDS)的題組，如表二：²¹

表二 傳統心理計量法所採的問卷題目

<p>針對下面各項陳述，請按下面的答項來表達你認同的程度</p> <p>(1) 完全同意 (2) 非常同意 (3) 同意 (4) 不同意 (5) 非常不同意</p> <p>(6) 不知道</p> <p>V7 () 我們一般老百姓對政府所作所為，沒有任何影響力。</p> <p>V8 () 政府官員不會在乎我們一般老百姓的想法。</p> <p>V9 () 台灣不管由哪個政黨來執政，結果都是一樣的。</p> <p>V10 () 我們的中央政府官員，都不了解地方基層的需要。</p> <p>V11 () 政治是少數政治人物玩的遊戲，跟我們多數人無關，不值得我們關心。</p> <p>V12 () 現在的政治人物為了爭取權位往往不擇手段。</p>

21. 「台灣選舉與民主化調查」(TEDS)的題組中，有八題有關政治功效感的問卷題目，我們認為「主觀政治知識認知」雖然與政治功效感有關，但與「是否有能力能夠影響政治」在概念上有很大差異，特別是「政治」一詞在台灣社會有「污名化」的意涵，而「不懂政治」的回應本身不見得不代表對政治冷漠或不在乎，而是避免讓人有「狂熱」、「權謀」、「激進」或「暗示難與人相處」的意義，也因此我們將之從題組中拿掉。不過如正文接下來所示，我們認為把「主觀政治知識認知」當作效標題還是合理的，因為多數人如果願意回答對政治有所理解，的確是比較容易表現出較強的政治功效感(黃信豪，2006:136)。而從事後的分析來看，「主觀政治知識認知」(表三中的V13)當作測量指標，測量效度並不佳(因素負荷為0.41)。

此題組已在歐美和台灣政治學界沿用多年，²²原為四點李克特量表，即「非常同意」、「同意」、「不同意」、「非常不同意」，但為了跟輔測定錨法的量表一致，作者在此改為五點的李克特量表，為「完全同意」、「非常同意」、「同意」、「不同意」、「非常不同意」。²³

除此之外，為了找到理論上可以客觀比較測量效度的基礎，我們以「主觀政治知識認知」的測量當作效標項目，在「台灣選舉與民主化調查」中，有兩題是有關「主觀政治知識認知」，但意義上有相當大差異，問卷題目如表三。關於「政治有時候太複雜了，所以我們一般老百姓實在搞不懂」此題，純粹是測量主觀上受訪者不認為他有能力理解政治事務，但「您覺得您對國家大事的了解，並不比一般人差」此題除了理解的面向上，主要問題的重點在「跟一般人相比是否較差」，因此受訪者可能不認為自己真正很懂複雜的政治事務，但跟一般人相比並沒有較差、甚至更好；也有可能受訪者自認很懂政治，但認為並沒有比一般人理解的多。由於後者的測量有「雙

22. 多數國內政治學者對此題組的使用是分成「內在功效感」和「外在功效感」，其中前者指涉的就是主觀政治知識的認知，而後者是主觀政治影響力的認知（黃信豪，2006：123；Madsen, 1987: 574）。在表二中，「政治是少數政治人物玩的遊戲，跟我們多數人無關，不值得我們關心」與「現在的政治人物為了爭取權位往往不擇手段」兩題都屬於「內在功效感」的測量題目，但究其語意，尚有外在功效感的意味存在，因為都意味著一般人認為其對於政治的大環境太難有任何影響力。也因此，除了純粹的主觀政治知識認知題之外，作者決定還是將這兩題與測量外在功效感的四題一起進行心理計量分析，而結果如表六所示，「政治是少數政治人物玩的遊戲，跟我們多數人無關，不值得我們關心」此題也具有相當高「外在功效感」的測量效度，但「現在的政治人物為了爭取權位往往不擇手段」顯然建構效度偏低因而剔除。

23. 另一種可能的作法，是沿用政治功效感在傳統心理測驗中所採的四分量表，而將輔測定錨法中的施測題目改成四分量表，以平衡的量表（四分）來取代現行的不平衡量表（五分）。不過這需要另起一個新的研究，才能知道兩種特性的量表會不會造成測量效度上的差別。

重問題」(double barrel) 的成份存在，²⁴所以我們選擇前者當作效標量表，假設政治功效感較高的受訪者理論上主觀認為對政治事務的理解較高。效標效度的評量方式如下：在同樣主觀政治知識認知為依變項的前提下，透過多變量迴歸分析，如果使用「輔測定錨法量表」的解釋力較高，代表其測量方法的效度較高；相反地，則「心理計量法」的測量效度較高。²⁵

表三 「主觀政治知識認知」的問卷題目

<p>針對下面各項陳述，請按下面的答項來表達你認同的程度</p> <p>(1) 完全同意 (2) 非常同意 (3) 同意 (4) 不同意 (5) 非常不同意</p> <p>(6) 不知道</p> <p>V13 () 政治有時候太複雜了，所以我們一般老百姓實在搞不懂。</p> <p>V14 () 您覺得您對國家大事的了解，並不比一般人差。</p>

問卷執行期間為民國九十六年的六月十二到十五日，在全部 671 個樣本觀察值中，由於這三個系所在學生性別上女多於男，所以男性的樣本佔 35.9%、女性佔 64.1%；年齡的部分由於包括在職專班，所以年紀最大 51 歲、最小 19 歲、平均約 22.68 歲；以學校來看，高雄餐旅學院的學生佔 55.6%、中山大學政治所佔 7.5%、台灣大學政治系佔 33.4%、台灣大學推廣進修部學生佔 3.6%；受試者皆為在校大學部或研究生，介於大學部一年到碩士班二年級之間。

24. 「雙重問題」是指問卷題目中包涵了不只一個問題，使得受訪者無從回答，或者僅選擇其一來回答，而如此得到的結果並不具備測量效度（王雲東，2007：110）。

25. 基於我們認為「政治功效感」應可解釋「主觀政治知識認知」的前提下，使用不同量表所得出的解釋力較高（R-Square），或者同時將兩量表放入迴歸模型中所得之標準迴歸係數較高，代表其效標效度較高。

至於在分析步驟上，本文擬對輔測定錨法分別進行三項檢測，都是針對其「全部樣本的無母數測量」結果：一、先檢查其受訪者對於輔測題的定錨排序是否符合 Gary King 等學者所宣稱的客觀性，如果不是的話，將視狀況剔除不好的輔測題，重新形成定錨量表。二、針對輔測與自我陳述題目進行心理計量分析，並檢定兩者在測量上的內在效度。三、進行順序類別多變量迴歸分析 (ordinal logistic regression)，使用主觀政治知識認知為效標，並加入年齡、性別及就讀學校等控制變因，來比較「輔測定錨法」和「心理計量法」的效標效度。

四、效度評估分析

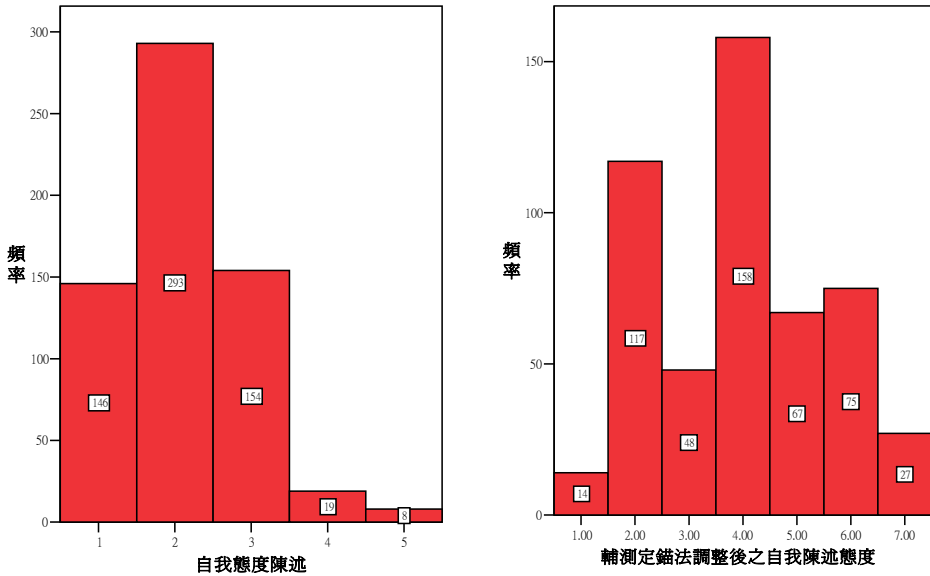
根據 Gary King et al. (2004) 的宣稱，輔測定錨法最大的 - 優點是藉由受訪者對參照對象狀態的評估而反映出其態度的主觀性，並且他們主張輔測題難度的差異設定是具有客觀性，但是根據本文的研究顯示，在全部 671 個樣本觀察值中，受訪者對輔測題難度評估符合 $c_1 \geq c_2 \geq c_3 \geq c_4 \geq c_5$ 排序的僅有 258 人，佔 38.5%，而不符合上面排序的有 413，佔 61.5%。經過進一步分析之後發現，如果將輔測題目按研究者的難度設定兩兩配對，可以看出輔測題第二和第四題的難度明顯地與問卷設計者所想的不同，如表四，以第二題和第一題相比，有 33.7% 的受訪者認為懷恩的政治功效感大於淑華，但顯然原先的設定並非如此；又如以第四題相比於第三題為例，有 20.7% 的受訪者認為振中比巴奈的政治功效感高，而且這也反映在第四題與第一、二題的錯誤排序率高於第三題與第一、二題的現象上。上述的證據顯示出，至少在這群學生當中，輔測題難度差異具有客觀性的宣稱

是站不住腳的，而最大的原因就是受訪者對於第二與第四題的認知顯然迥異於問卷設計者的認知，因而本文在下面的分析中將會將此兩題剔除，而僅使用第一、三、五題來進行輔測定錨法得測量。

表四 輔測定錨法之錯誤排序率

	V1	V2	V3	V4
V2	33.7%			
V3	4.2%	4.6%		
V4	8.2%	7.6%	20.7%	
V5	1.5%	0.8%	3.5%	4.6%

如果在僅使用第一、三、五輔測題來進行定錨法測量，則錯誤排序率將從 61.5% 大幅減少到 9.4%，在此狀態下，如果將自我態度陳述題加進比較，則可以用七點李克特量表來呈現輔測定錨法的測量結果，如圖二所示：



圖二 自我態度陳述題經過輔測修正前後的敘述統計結果

由圖二比較可以看出，在剔除了第二、四輔測題之後的定錨法測量，的確會大幅調整自我態度陳述的分配從傾向右偏變成趨近於不偏的常態分配，而如果我們先驗地認為一般大學生的政治功效感是符合常態分配的，那的確輔測定錨法可以矯正原始自我態度陳述傾向負面的分配。但究竟一般大學生的政治功效感是不是符合常態分配，其實不是「先驗的」假定為真，假使我們可以利用心理計量法來呈現出輔測題與自我陳述題具備測量的內在效度，則更可以加強研究者對於輔測定錨法在理論假設上的信心，至少是「內容效度」和「建構效度」的說服力，表五正是二分 IRT 因素分析的結果：

表五 輔測題組之建構效度分析

	因素負荷 (無遺漏值)	因素負荷 (遺漏值)
輔測題一	0.368	0.172
輔測題二	0.481	0.225
輔測題三	0.467	0.409
輔測題四	0.378	0.430
輔測題五	0.511	0.700
自我態度陳述題	0.615	0.534
解釋變異	22.79%	20.11%
樣本數	451	670

上面的數據顯示，不管要不要將帶有遺漏值的輔測題資料代入二分 IRT 因素模型中，基本上所都透露出的訊息都是輔測題並不完全符合單面向性的假設，因為因素負荷都顯著地偏低，介於 0.368 到 0.615 之間，解釋變異僅有 22.79%，並且這還是去除了近三分之一遺漏值樣本的結果，如果連遺漏值樣本都代入，那麼因素負荷不但更偏低且變動更大，介於 0.172 到 0.700 之間，而解釋變異更下降成只有 20.11%，這樣的結果說明了，輔測和自我陳述這六題之間並不具備足夠的「建構效度」可以來測量「政治功效感」這個潛在心理特質，同時由於受訪者對於第二和第四輔測題的認知也與問卷預設的落差甚大，因此其「內容效度」亦讓人相當懷疑。

如果將剔除第二、四輔測題使用定錨法所形成的測量結果，分別與去除和保留遺漏值樣本的二分 IRT 量表進行比較，結果發現不論是去除還是保留遺漏值，兩者都證明輔測定錨法和心理計量法所產生的測量結果有非常顯著的差異，相關度低到僅有 $r = 0.063$ 和 0.116 ，而基於輔測定錨法的內容效度有嚴重瑕疵（因素負荷變化頗大），且輔測題的錯誤排序率又偏高，加上建構效度又普遍偏低，

因此我們懷疑 Gary King et al. (2004) 所宣稱輔測定錨法所具備有優點是否成立。事實上從之前一連串的統計檢證看來，輔測定錨法的最大好處是將一個具有相當程度偏態的態度分配，調整成為一個接近不偏的常態分配，但是究竟這些受訪者的態度是不是常態分配，而經過輔測定錨法所調整的分配不符合真實，上面都沒有積極的證據支持這樣的判斷。

表六 輔測定錨法量表與輔測題之心理計量量表之相關性分析

	相關係數	顯著性	樣本數
去除遺漏值之輔測題二分 IRT 量表	0.063	0.239	352
保留遺漏值之輔測題二分 IRT 量表	0.116	0.009	506

既然輔測定錨法的施用並沒有辦法得到品質較佳的測量結果，我們要問的是，到底原來採用已久的政治功效感問卷設計有多糟糕？針對這個問題，我們對傳統的問卷也進行了政治功效感題組的二分 IRT 因素分析，結果如表七：

表七 政治功效感題組的二分 IRT 因素分析

	因素負荷	因素負荷
否定老百姓對政府所作所為沒有任何影響力 (V7)	0.783	0.785
否定政府官員不會在乎老百姓的想法 (V8)	0.833	0.832
否定台灣不管由哪個政黨來執政結果都是一樣的 (V9)	0.522	0.518
否定我國政府官員都不了解基層的需要 (V10)	0.602	0.601
否定政治與自己無關不值得關心 (V11)	0.578	0.577
否定現在的政治人物爭取權位往往不擇手段 (V12)	0.308	—
解釋變異	39.53%	45.41%
樣本數	669	669

由上表可以看出，在原來六題當中，多半的因素負荷都相當高，介於 0.522 到 0.833 之間，只有 V12 並不是相當適合測量政治功效感的問題，整個題組的解釋變異相當高，約為 39.53%。如果更進一步去蕪存菁將 V12 去掉，則測量的效度更大符提升，解釋變異上升到 45.41%，而因素負荷也都是介於 0.518 到 0.832 之間，這顯示了傳統政治功效感的題組其實不管是「內容效度」（因素負荷的變異數）或「建構效度」（因素負荷的係數值）來說都比輔測定錨法來得優越許多。

基於二分 IRT 因素分析法所顯示傳統問卷的優越性，讓我們回過頭來看輔測定錨法的題組是否具有內容效度，作法是將表六中去除 V12 所形成的二分 IRT 量表與六題輔測和自我陳述題進行相關分析，如果相關係數越大且符合顯著性，那麼其問題之「內容效度」就得到確證，結果如表八：

表八 傳統政治功效感問卷之二分 IRT 量表與輔測題之相關分析

	輔 測 題					自我陳述題
	一	二	三	四	五	
傳統政治功效感問卷之二分 IRT 量表	0.083* (0.035)	0.125*** (0.001)	0.013 (0.749)	-0.100* (0.014)	0.027 (0.507)	0.385*** (0.000)
樣本數	641	659	613	601	628	620

* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

基本上相關性分析顯示輔測題的「內容效度」都不佳，特別是第三、四、五題，相關係數不是不顯著，就是呈現負號，而就第一、二題來說，相關係數也偏低，唯有自我表述題的相關性尚可接受，這說明了若從心理計量學的角度來看，其實輔測定錨法不但無法客

觀的定錨，也不能解決「項目功能差異」的問題。

進一步以主觀政治知識認知為依變項，控制受訪者性別、年齡及就讀學校，以多變量順序類別迴歸法進行輔測定錨量表和心理計量量表的效標效度分析，結果如果表九所示，模型一代表單獨以輔測定錨法的政治功效感量表所進行的效標效度迴歸分析，我們發現政治功效感與主觀政治知識認知存在著顯著的正向關係（ $\beta=0.372$ ），這說明輔測定錨法的測量是有效標效度的。而模型二所顯示的是單獨將心理計量法所形成政治功效感量表帶入的迴歸分析結果，結果發現心理計量法的測量同樣具有有效標效度（ $\beta=0.566$ ）。不過比較模型一和二的結果發現，心理計量法的效度都不比輔測定錨法來得差，可以從模型二的近似解釋變異值（Pseudo R-Square）0.217 高於模型一的 0.197，以及模型二迴歸係數 0.566 大於模型一的 0.372 可以看出。而若同時將輔測定錨法和心理計量法量表帶入迴歸模型中，結果也如模型三所示，心理計量法比輔測定錨法具有較高的效標效度，這可從較大的迴歸係數和較強的顯著性看出。

表九 兩種政治功效感量表的效標效度分析
(採多變量順序類別迴歸法)

解釋變項	模型一	模型二	模型三
政治功效感 (輔測定錨法)	0.372 *** (0.089)	—	0.273 ** (0.091)
政治功效感 (心理計量法)	—	0.566 *** (0.083)	0.477 *** (0.098)
性別 (參照組：男性)	-0.080 (0.177)	-0.214 (0.155)	-0.067 (0.178)
年齡	0.07 (0.024)	-0.005 (0.020)	0.001 (0.024)
高雄餐旅學院(參照組：中山政治所)	-0.831 (0.431)	-0.415 (0.377)	-0.527 (0.436)
台大政治系 (參照組：中山政治所)	0.611 (0.424)	0.630 (0.371)	0.608 (0.426)
台大進修推廣部 (參照組：中山政治所)	0.225 (0.554)	0.591 (0.478)	0.224 (0.556)
常數項一	-2.483 ** (0.794)	-2.286 * (0.909)	-2.286 * (0.912)
常數項二	-1.333 (0.789)	-1.161 (0.904)	-1.111 (0.906)
常數項三	0.595 (0.788)	0.818 (0.903)	0.944 (0.906)
常數項四	2.238 ** (0.796)	2.427 ** (0.912)	2.569 ** (0.916)
Pseudo R-Square (Nagelkerke)	0.197	0.217	0.240
樣本數	492	648	492

註：1. 政治功效感的兩種量表皆已標準化成標準常態分配，所以兩者的迴歸係數可以直接比較。

2. 三個模型的樣本數不同是因為遺漏值的關係。

3. 常數項一指的是針對依變項問題「政治有時候太複雜了，所以我們一般老百姓實在搞不懂」回答「完全同意」，也就是以「主觀政治知識認知」的測量最低值當作二分「邏輯斯連結函數」(logistic link function)分斷點的常數估計值。至於常數項二三四相對應的是以「非常同意」、「同意」、「不同意」為分斷點的常數估計值。

五、結 論

本文探討了 Gary King 等學者近年所推廣的輔測定錨法之原理與方法，也實際以台大政治系、中山政研所和高雄餐旅學院的大學生、研究生為受試者進行政治功效感的心理測驗，結果發現輔測定錨法並不能解決 Gary King 等學者所提出的「項目功能差異」問題。更甚者，輔測定錨法在測量品質上甚至遠遠比不上使用心理計量法所形成的量表，主要的問題不僅出在輔測題的「內容效度」和「建構效度」都遠不如傳統問卷，同時更出在受試者無法與問卷設計者有相同的感知來理解輔測題的排序理則。結果不但定錨無法如 Gary King 等學者所預期的客觀，自我陳述與輔測題的相對排序又多數發生邏輯矛盾，使得輔測定錨法的施用非常困難，而研究者也無法在事先知道那些輔測題的效度有問題而剔除在測驗中。輔測定錨法唯一的好處是可以將原先看起來呈現若干程度偏態的態度分配，矯正成為偏態較為輕微的狀態，但是除非研究者可以證明本來受試者的態度分配就應該符合常態分配，但顯然這種先驗的假設是沒有任何科學基礎的。

相形之下，傳統政治功效感的問卷設計在「內容效度」和「建構效度」的表現上就比輔測定錨法好太多了，特別是經過心理計量法的檢證之後，只要在傳統題組中慎選測量題目，其實在測量品質上都有相當正面的統計指標支持，也因此不管是從施測题目的簡單性、測量結果的統計評估上，我們都認為至少在有限的證據當中，並沒有任何必要性來採行輔測定錨法，而所有 Gary King 等學者所主張的優點，雖然在學理上有其道理，但其所植根的假設在本文的

研究中都未獲支持，因此未來在大型的心理測驗或民意調查研究中，除非有更明確的經驗研究支持，否則輔測定錨法的實施尚待斟酌。

除了上述的理由外，下面還有兩個原因讓作者對於輔測定錨法的實施採取保留的態度。首先、輔測題本身的設計太過複雜，容易有「雙重問題」的缺點，假使研究者一廂情願希望受訪者能夠意會出題目設計的難易程度，是非常不切實際的。所以此法如果施用在不同年齡層與教育程度的全國民眾，而不是在本文中受過大學教育的受試對象，可以預期其測量有效性將會更低，同時測量的誤差也可能更大。其次、關於本文的施測，作者是將自我態度陳述題放在最後，同時並沒有對輔測題進行隨機排列。這個作法不同與 Gary King 等學者建議將自我態度陳述題放在輔測題前面並對輔測題進行隨機排列。此處牽涉到兩個問題，一是自我態度陳述回答是否會受到輔測題的影響，二是輔測題本身是否會給人「提示」(cuing) 其難易程度的效果。簡單來說，Gary King 等學者的作法是希望極小化上述兩因子的影響，希望在受訪者還沒有受到輔測題影響下，取得最直觀的答案，而輔測題的隨機排列，等於是讓受訪者進行心理測驗，希望測試受訪者的理性認知能力，所以這兩個作法理論上都可能讓測量的結果比較差。而作者在本文中所採的施測方式，雖然有可能會讓自我態度陳述回答受到輔測題的影響，同時輔測題也可能有「提示」的效果，但作者所得到的測量結果，應該是在所有施測方式下最樂觀的，可是從文中的發現得知，即便最樂觀的結果都尚且不盡如人意，如果改採 Gary King 等學者的作法，合理預期的結果應該會更差，基於這樣的評估，輔測定錨法的實際測量效度的確很難讓人具有信心。

本文對輔測定錨法所進行的初步評估，在研究設計上還有許多未盡完善之處，都有待未來進一步的實驗結果才能來詳盡的評斷。下面茲就可能的改進方向提出幾點建議：一、可改採 Gary King et al. (2004) 的建議，將自我評估題放在前面，然後輔測題放在後面並隨機排列，看看測量效度是否有所差異。二、對輔測題進行更口語化、適合台灣社會脈絡的修正，並簡短文句，去除可能具有的「雙重問題」毛病。三、將所有問卷的答項改成四分的平衡量表，比較與現行所採五分不平衡量表的研究結果是否不同。四、考慮到多族群文化的因素，在輔測題中針對不同族群的受訪者可使用該族群熟悉的人名來進行輔測題的陳述，儘量避免受訪者因為感受上的差別所造成測量上的誤差。

參考書目

- Brady, Henry E. 1989. "Factor and Ideal Point Analysis for Interpersonally Incomparable Data." *Psychometrika* 54, 2: 181-202.
- Cronbach, Lee J. and Paul E. Meehl. 1955. "Construct Validity in Psychological Tests." *Psychological Bulletin*, 52, 4: 281-302.
- Du Toit, M. 2003. *IRT from SSI*. BILOG-MG, MULTILOG, PARSCALE, TESTFACT. Lincolnwood, I.L.: Scientific Software International.
- Embretson, Susan E. and Steven P. Reise. 2000. *Item Response Theory for Psychologists*. Mahwah, N. J.: Lawrence Erlbaum.
- Fink, Arlene. 1995. *How to Ask Survey Questions*. Thousand Oaks, CA: SAGE Publications.
- Fitzpatrick, Anne R. 1983. "The Meaning of Content Validity." *Applied Psychological Measurement* 7, 1: 3-13.
- Holland, Paul W. and Howard Wainer. eds. 1993. *Differential Item Functioning*. Hillsdale, N.J.: Lawrence Erlbaum.
- King, Gary et al. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98, 1: 191-207.
- King, Gary and Jonathan Wand. 2007. "Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes." *Political Analysis* 15, 1: 46-66.
- King, Gary. 2007a. "The Anchoring Vignettes Website." in <http://gking.harvard.edu/vign/>. Latest update 25 August 2008.

- King, Gary. 2007b. "What Has to Go Wrong for Anchoring Vignette Corrections to Bias My Results?" in <http://gking.harvard.edu/vign/faq/node15.html>. Latest update 25 August 2008.
- Madsen, Douglas. 1987. "Political Self-Efficacy Tested." *American Political Science Review* 81, 2: 571-581.
- Narens, Louis and R. Duncan Luce. 1983. "How We May Have Been Misled into Believing in the Interpersonal Comparability of Utility." *Theory and Decision* 15, 3: 247-260.
- Parten, Mildred Bernice. 1950. *Surveys, Polls, and Samples: Practical Procedures*. New York: Harper.
- Poole, Keith T. 2008. "The Evolving Influence of Psychometrics in Political Science." in http://voteview.com/The_Evolving_Influence_of_Psychometrics_in_Political_Science.pdf. Latest updated 25 August 2008.
- Putnam, Robert D. et al. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.
- Reckase, Mark. D. 1997. "The Past and Future of Multidimensional Power of Item Response Theory." *Applied Psychological Measurement* 21, 1: 25-36.
- SAS Institute. 2000. *SAS/STAT Software: Changes and Enhancements, Release 8.1*. Cary, NC: SAS Institute.
- Schmitt, N. 1996. "Uses and Abuses of Coefficient Alpha." *Psychological Assessment* 8, 4: 350-353.
- Suchman, Lucy and Brigitte Jordan. 1990. "Interactional Troubles in Face to Face Survey Interviews (with Comments and Rejoinder)."

Journal of the American Statistical Associations 85: 232-253.

Wrigley, Jack. 1976. "Pitfalls in Educational Research." *British Educational Research Association Bulletin* 3, 1: 1-7.

王雲東。2007。《社會研究方法》。台北：威仕曼文化。

林宜靜等。2006。〈產品類型與實虛通路型態對顧客價值、顧客滿意度與忠誠度之影響〉。《顧客滿意學刊》2, 2: 121-159。

陸洛。1998。〈中國人幸福感之內涵、測量及相關因素探討〉。《國家科學委員會研究彙刊：人文及社會科學》8, 1: 115-137。

張彩秀、黃乾全。2002。〈中老年健康促進生活型態與主觀健康狀態之關係〉。《衛生教育學報》17: 15-30。

黃旻華。2006。〈態度量表的心理計量學分析：2003年TEDS統獨態度量表的研究〉。《選舉研究》13, 1: 43-86。

黃信豪。2006。〈政治功效意識的行動效果(1998-2003)〉。《台灣民主季刊》3, 2: 119-158。

A Validity Assessment of the Anchoring Vignettes vs. Psychometric Approach: The Case of Political Efficacy Measurement

Shu-Hsien Chang* Min-Hua Huang**

For a long time, the problem of differential item functioning (DIF) has troubled researchers' trust in the validity of psychological measurements in the academic community of public opinion research. The main issue is not only reflected in the phenomenon that different respondents all have their own perceptions and reactions to the questionnaire and response categories, which is called "the problem of interpersonal comparability," but also related to the serious doubt of whether the questionnaire can tap into the attitudes that researchers want to measure.

To cope with the above issues, mainstream scholars of the psychological and educational academic community apply sophisticated models to deal with the measurement problem. However, Gary King and his colleagues in their 2004 APSR article proposed the approach of anchoring vignettes and claimed that it can solve the DIF problem

* Associate Professor, Department of Hospitality Marketing and Event Management, National Kaohsiung Hospitality College.

** Assistant Professor, Department of Political Science, Texas A & M University, U.S.A.

thoroughly. This article tries to evaluate the measurement validity of the two approaches with an experimental survey of measuring the political efficacy of 671 enrolled students from three universities. The result indicates that the approach of anchoring vignettes is not superior to the approach of psychometrics at all in terms of validity assessment. Furthermore, anchoring vignettes suffers serious methodological problems which lead the authors to conclude with a considerable reservation on the application of anchoring vignettes to replace the psychometric measurements.

Key words: anchoring vignettes, psychometrics, differential item functioning, interpersonal incomparability