

國立高雄師範大學教育學系

教育學刊，2008年6月，30期，頁101-140

調查方式與議題熟悉度對問卷回收率與內容的影響

余民寧^{*}、李仁豪^{**}

摘要

本研究旨在探討網路與紙本兩種問卷調查方式及對問卷所調查議題的熟悉度兩個變項，對問卷回收率及問卷調查內容是否有顯著差異的影響存在。本研究以分層隨機取樣方式抽取臺灣地區1,200位中小學教師進行研究，實際回收425份問卷，整體回收率35.4%，其中網路組195人（佔45.9%），紙本組230人（佔54.1%）。本研究以卡方檢定來進行回收率的比較，並且使用Rasch模式，將李克特式評定量尺資料轉換成具連續量尺資料後，再進行問卷內容的分析。結果發現：

（1）在問卷回收率上，調查方式與議題熟悉度之間並無交互作用效果存在，也無主要效果存在；（2）在問卷內容上，調查方式與議題熟悉度之間有交互作用效果存在，其中，針對差異試題功能（DIF）的單純主要效果檢定發現，在高議題熟悉組或低議題熟悉組中，網路組

^{*} 余民寧：國立政治大學教育學系教授

^{**} 李仁豪：中山醫學大學心理系助理教授

致謝：作者要特別感謝國科會補助本論文之部分研究經費，補助案編號為：NSC-93-2413-H-004-008及NSC-94-2413-H-004-001。

與紙本組之間幾乎沒有出現DIF的試題；但是就高低議題熟悉組之間的比較而言，在網路組出現DIF的試題與在紙本組出現DIF的試題有較大的不同；(3) 調查方式的DIF主要效果顯示，網路組與紙本組在各試題上並無DIF現象；而議題熟悉度的DIF主要效果顯示，高低議題熟悉組之間出現許多DIF試題，且高議題熟悉組的反應較低議題熟悉組極端；(4) 未來的問卷調查方式，似乎可以考慮使用網路問卷來取代紙本問卷，或者是兩者並行，因為，其效果是相同的。最後，本研究亦提出許多建議，供未來研究時參考。

關鍵詞：網路問卷、調查方式效果、議題熟悉度、Rasch 模式

隨著網路的普及，網路問卷已成為一種常見的調查方式。網路問卷具有方便、快捷、成本低廉等優點，但同時也存在著一些問題，例如：網路問卷的回應率通常較低，且容易受到網路環境的影響。此外，網路問卷的匿名性也可能導致一些不實的回答。因此，本研究旨在探討網路問卷與紙本問卷在調查效果上的差異，並分析議題熟悉度對調查結果的影響。研究結果顯示，網路問卷與紙本問卷在調查效果上存在著顯著的差異，且議題熟悉度對調查結果有著重要的影響。本研究為未來的問卷調查提供了有益的參考。



緒論

在科技時代裡，使用電子或網路問卷為工具以進行人文及社會科學的研究，正日益興盛，主要原因是其具有低成本、高效率的特性，使得愈來愈多的問卷調查改採不同於傳統郵寄紙本問卷、電話訪談、或面對面訪談的形式來進行。然而，不同調查媒介可能會導致不同資料來源的蒐集等相關方法學上的問題發生。以網路問卷與郵寄紙本問卷的比較為例，兩者在覆蓋率誤差（coverage error）、抽樣偏差（sampling bias）、不反應偏差（nonresponse bias）及測量誤差（measurement error）等因素上，都是值得關切的（Dillman, 2000; Groves, 1989）。

一、覆蓋率誤差

覆蓋率誤差係指研究所參照的母群與所要推論的母群（目標母群）之間的落差。例如，研究目的是想瞭解一般民眾對某項公共政策的意見，調查員使用電話訪談與面對面訪談方式進行，可能會獲得不同的結果，尤其在電話設施不十分普及的地區，特別明顯。同理可知，即使處在今日標榜為電子網路化世代，良好的網路設施也並非完全普及，無法全面覆蓋到所有想要推論的母群身上。因此，在進行網路問卷與紙本問卷的比較時，應該要慎選參照母群，以及所欲推論的母群，並儘量使兩者能夠重疊，才能減少覆蓋率誤差所造成無效推論的情形發生。

相對於以紙本方式為基礎的問卷調查（paper-based questionnaire，簡稱「紙本問卷」），以網路為基礎的問卷（web-based questionnaire，簡稱「網路問卷」）是一種較不昂貴且快速，並且能夠涵

蓋較廣大地理區域的理想工具。因此，網路問卷對研究者而言，是一種頗為吸引人的提議（Denscombe, 2006）。然而，在網路設備尚未如同電話般普及之前，在進行網路與紙本問卷的比較時，應該要選擇或界定出對網路有一定熟悉度及可接觸到的母群為何，才能減少網路不足所產生覆蓋率誤差的問題。由於在都會地區或高收入及高教育者，才有較多接觸網路設施的機會，因此，本研究選擇具有一定社經地位及網路使用經驗的臺灣中小學教師為母群，而非一般的民眾，如此考量應該可以避免或減少網路覆蓋率誤差的問題，當然，推論的範圍也只能僅止於此而已。

二、抽樣偏差

以完全隨機方式進行抽樣，並不適用於一般中小型研究，尤其在參照母群範圍廣大，且所要抽取的樣本數太大時，在財力及人力無法配合的情形下，很難達到完全隨機取樣的要求，因此，一般研究只能在有限的資源下，使用較不嚴格的隨機抽樣方式來進行研究。此外，若所抽取的樣本數太少，可能也會有代表性的問題存在；樣本的特性必須與母群的特性儘量相符，才能減少抽樣偏差所造成的無效推論。至於，樣本要多大才算是有代表性的問題，則是見仁見智的看法。在網路時代，有部分學者認為不存在樣本多寡的問題，因為低成本的網路問卷幾乎不必擔心沒有受試者（Dillman & Bowker, 2001），而是應該擔心受試者的代表性問題（Grandcolas, Rettie, & Marusenko, 2003）。許多調查研究的抽樣計畫中，常常使用所謂的自願受試者（volunteers）或者是提供酬賞以招募受試者參與，這些作法都不是符合隨機取樣原則的。另外，國外的調查方式也常會事先以電話或其他方式徵求受試者同意後，再進行後續問卷調查的作業，這些方式都可能使願意接受調查的樣本是否具有母群的代表性問題，受到質疑。

抽樣偏差常與調查方式效果混淆，而無法獲得較佳的結論。抽樣偏差的來源之一，即是自我選擇偏差（self-selection bias）所造成樣本代表性的問題；選擇或願意以某種方式來反應的受試者，顯然與其個人的經驗及相關背景有關（Grandcolas et al., 2003; Hayslett & Wildemuth, 2004）。Shaw與Davis（1996）在一項調查中發現，電子問卷組與紙本問卷組在問卷內容上有顯著的差異，而這些差異恰巧反映出使用電子問卷的受試者有較多的電子科技經驗，他們更常擁有人電腦、更常使用電子郵件等。Synodinos、Papacostas與Okimoto（1994）曾經比較自我選擇的受試者與隨機選擇的受試者，在一項電腦上所進行的問卷調查，結果也是顯現這兩者具有顯著不同的調查結果。Grandcolas等人比較網路與紙本問卷的差異後發現，填答網路問卷的受試者對於使用網路的態度較為正向；而這些較偏好網路的受試者，通常是來自接受良好教育和擁有高社經地位者（Morrel-Samuels, 2003）。Bandilla、Bosnjak與Altdorfer（2003）比較網路問卷與紙本問卷的差異，紙本組的樣本是來自一群有代表性的德國民眾，而網路組樣本則是來自一群有代表性的德國網際網路使用者，起初研究結果指出網路與紙本組有許多不一致的發現，研究者將之歸因為社會人口變項所造成的結果，因此，他們將兩組樣本的社會人口變項控制在高教育水平上，結果發現受試者在網路問卷與郵寄紙本問卷上的反應基本上是完全相同的。Leung與Kember（2005）以香港七個學院共2,786位大學生為研究對象，所有樣本先以郵遞方式寄送紙本問卷，幾天後再寄送電子郵件並附上網路問卷網址，三週後又同時以上述兩種方式進行催收，最後整體回收率為63.8%；研究發現，除了機械工程學院的受試者較偏好網路問卷外，其他包括藝術、商學、教育、醫學、科學及社會科學院的學生，在網路問卷與紙本問卷的回收率上並無差異；另外，兩者在18個量尺上的信度係數也沒有差異存在，最後，以

這18份量尺形成的完整結構方程式模型也獲得十分適配的結果。

因此，有效地控制樣本的背景變項，會使調查方式的比較具有更可信的結果。Hayslett與Wildemuth（2004）謹慎地選擇背景相似的300位學術圖書館館員為對象，比較郵寄紙本問卷、紙本郵寄上網填答通知、電子郵寄上網填答通知這三種問卷內容的反應結果發現，三者幾乎沒有任何差異存在。McCabe（2004）以自陳報告方式調查7,000名大學生非法藥物使用的經驗，這些學生被隨機分為兩組，並分別使用網路問卷與郵寄紙本問卷進行調查，結果發現這兩組具有相似的藥物使用經驗。Huang（2006）在商業學校及管理學校各抽取兩個班為樣本，以每個學校的一個班為網路組，另一個班為紙本組方式進行配對，比較網路問卷與紙本問卷在一般背景變項、行為量表、態度量表、適應性問題、敏感性問題及開放性問題上的差異，研究結果顯示這兩組受試者在這些問題上的比較，幾乎沒有差異存在。Denscombe（2006）以同質性高的15歲中學生為樣本，使用網路及紙本問卷調查青少年的健康行為，結果在23題中，只有1題出現有明顯差異。

因此，本研究為了避免抽樣偏差造成的影響，在研究樣本的取得上，乃是直接根據教育部網站所公布的中小學學校網址，來進行隨機點選取樣，因此，較能減少偏差樣本所產生的代表性問題。此外，由於中小學教師屬於社經地位中上的一群，對於網路使用的經驗與知識也較為豐富和整齊，因此，較不會在進行網路組與紙本組問卷調查比較時，出現因為社經背景因素而造成誤差結果的結論。

三、不反應偏差

當選定好樣本，將問卷送出後，實際上可回收的樣本數不太可能有100%，即使回收率高達80%，在文獻上也是少有記載的，因此，

未回收的問卷可能與已回收的問卷之間會有實際上的差異，因而影響結果的推論。Grandcolas等人（2003）的研究中，網路組的問卷回收率只有4%，Tse（1998）的研究中，電子問卷的回收率僅有7%，而Bason（2000）的研究中，網路問卷回收率也只有16%。當然，這些例子應該有其特殊情況下可以說明其代表性無虞時方可被接受，否則，太低的回收率將會令人質疑為什麼會有這麼多人不回應。如果只是單純隨機地不願回應，而非出於系統性的偏差，那麼不反應偏差對研究結果的推論所造成的影響就可以忽略。由於低回收率代表不反應的受試者比例很高，對研究結果容易造成不反應偏差效果。例如：在Grandcolas等人的研究中，藉由網路發出電子邀請函給4,000位學生，並提供線上問卷網址，結果網路問卷回收率只有4%，在比較網路問卷與紙本問卷的作答結果後，發現兩組有許多的差異存在。此外，網路組的受試者在某些試題上傾向於選擇量尺的極端類別選項，特別是那些對網路較有經驗者比較不怕暴露隱私而提出較強烈的意見。因此，研究者認為此項差異主要是來自樣本偏差所造成，並非不同調查方式所產生的效果。

Grandcolas等人（2003）認為，儘管網路媒介具有新穎性，但是過度調查效果（over-surveying effect）可能也是導致回收率下降的原因之一。雖然這個看法似乎有點牽強，畢竟網路電子興起也才近十年左右的事，不過，由此更可以看出電子網路時代的垃圾資訊已經多到多麼令人厭惡的地步。此外，回收率下降的情形也同樣出現在紙本問卷中，Goho（2002）即發現紙本問卷回收率有逐年下降的趨勢，在一項機構的例行調查中，回收率從1995～1996年的54.3%，下降到1999～2000年的35.2%。另一項美國海軍例行的問卷調查研究也發現，紙本問卷回收率由1989年的60%，下降至1999年的30%（Newell, Rosenfeld, Harris, & Hindelang, 2004）。在網路問卷與紙本問卷回收率

的比較上，Katja與Vasja（2004）進行兩者在回收率上的後設分析，結果發現網路問卷確實有較低的回收率，他們認為這可能是因為網路問卷的調查方式還未發展到像其他調查方式一樣程度，或者是對網路調查方式逐漸失去好奇心，以及隱私性受到侵犯等問題，都是可能原因之一。

Curtin、Presser與Singer（2000）評閱早期文獻後發現，回收率從20%增加到40%，比回收率從40%增加到60%對研究結果有更大的衝擊，儘管如此，他們在一項長期例行性調查的資料庫中研究回收率高低對調查內容的影響，藉由回歸分析比較一開始允諾配合調查的受試者與經過多次電話催收後允諾配合的受試者對問卷內容的預測效果，在控制住受試者的背景變項及年代的影響後，發現並無顯著預測效果，亦即經過多次催收後增加的回收率，並不會影響未經催收前的問卷調查結果。另外，在Keeter、Miller、Kohut、Groves與Presser（2000）的研究中，比較一般正常的五天調查期的36.0%回收率，與八週長期調查的60.6%回收率，在91題問卷上的反應差異情形，結果發現只有14題出現顯著差異，其中有7題與背景變項有關，包括收入、教育水準、種族，但是兩者在性別、年齡上並無差異，此外，在91題的比較中，最高差異量為9%，平均差異量只有2%，整體而言，五天期回收率與八週期回收率的問卷反應是十分相似的。

由上述可知，較高的回收率（50%以上）與中等的回收率（30~50%），通常在人口背景變項上有些微的差異，但是對研究結果的影響並不嚴重，有時甚至微小到可以忽略。至於那些不反應的受試者，通常對研究結果的影響並非系統性的，反而有相當大的比例是對問卷調查沒有好感所致，尤其在今日自由民主的社會裡，有用資訊與垃圾資訊充斥整個網路世界，受試者個人有選擇是否做出反應的自由，因此，問卷回收率太低，應該是單純地隨機選擇下的產物。然而，

Groves、Singer與Corning（2000）在其所提出的槓桿——突顯理論（leverage-salience）中指出，問卷的主題可能是影響樣本組成的重要因素之一，例如：認為某個問卷主題是吸引人的一群受試者，將藉由該研究而突顯出來；如此一來，這些對該項主題有興趣的一群受試者，便會在問卷測量變項中顯現出與未回覆該問卷者之間的差異，導致嚴重的不反應偏差。這時，回收率高低與不反應偏差的關係，似乎就不如問卷主題吸引力與不反應偏差的關係來得重要，儘管此時，對受試者而言，會產生較高興趣的問卷主題之回收率，會遠高於較低興趣的問卷主題之回收率。Groves、Presser與Dipko（2004）在一項5（五種不同母群）×5（五種問卷主題）的實驗設計中，以問卷主題與母群的相似性來假定母群對問卷的興趣高低，結果發現受試者對於與其母群有關的問卷主題的反應率，明顯高於與其母群無關的其他四種問卷主題之反應率，其中，尤以教師母群的受試者對「教育與學校」問卷的反應率，與對其他四種問卷主題的反應率差異最大（達13.9%）；但是政治捐獻母群受試者對「選舉」問卷的反應，與對其他四種問卷主題在反應率上並無差異，這可能是因為對政治捐獻者而言，任何問卷主題都與政治有關所致。不過，Groves等人並無針對問卷內容作比較，因此，無法獲知反應率差異對問卷內容影響是否嚴重的訊息。

承上所述，本研究將在共同問卷試題的基礎上做一番差異檢定。在本研究中，我們加入了問卷議題熟悉度的概念，來探討網路問卷與紙本問卷之間的差異是否會受影響，這裡定義的議題熟悉度與前述問卷主題是否符合受試者興趣的概念有異曲同工之妙，亦即對該議題愈有興趣者對該議題的熟悉度也應該較高，因此，可能造成回收率上的差異。但有一點不同的是，興趣一般是經由事先設計不同議題來決定或假設問卷主題對受試者的吸引力，而熟悉度在本研究中是藉由問卷

回收後來界定（請參見研究方法一節說明）。除了回收率外，我們想瞭解問卷議題的熟悉度高低對問卷內容作答的影響，以及其對問卷調查方式在問卷內容上的調節效果為何？網路問卷與紙本問卷之間是否會因為受試者對某個問卷議題的熟悉度不同，而在問卷內容上產生實質的影響！

四、測量誤差

廣義來說，測量上的誤差包括施測時違反標準化程序，以及採用不一致或不適當的計分方式等；狹義來說，測量誤差可以專指使用的量表計分形式及配合的統計分析方式是否合理，不合理的計分方式無法表達受試者特質的真正程度，會增加測量上的誤差。就以心理計量學領域最常採用的李克特式評分量尺（Likert-type rating scale）而言，原屬於次序量尺（ordinal scale），並不適合直接當作連續量尺（continuous scale）來使用，必須經過適當的轉換程序，將其轉換成具備連續量尺的性質後，才能適當地用來表徵受試者的心理特質傾向。

目前，試題反應理論（Item Response Theory, IRT）中的Rasch模式，主要是由廿世紀中葉丹麥數學家Rasch所提出（Rasch, 1960），可以用來解決受試者心理特質測量的難題；在模式與資料適配的情形下，它能夠有效地將次序量尺轉換成具備連續量尺特質的邏輯尺度（logits scale），因此，可以有效地描述不同個體對同一份李克特式評定量表的潛在特質程度，而非單純且不精確地將李克特式評定量表上的各量尺反應點相加來表示個體特質高低而已。簡單地說，Rasch模式是將受試者填答某一反應類別的機率相對於填答下一反應類別的機率，取自然對數，以用來表示受試者的潛在能力與該題反應類別閾限（category threshold）的差異程度。機率值經邏輯斯（logits）轉換

後，便具有等距或比率量尺的特性，其單位是以logits來表示，具有統計分析上的優越性。

綜合上述覆蓋率誤差、抽樣偏差、不反應偏差、測量誤差的探討，本研究在經費成本的考慮下，擬針對兩種問卷調查方式與議題熟悉度高低這兩個獨立因子，進行實驗設計，將上述可能影響研究結果的偏差或誤差加以控制或減低其效果，期望研究結果能有助於未來網路問卷的推行，並提供實務上有關問卷議題熟悉度對調查方式可能影響的相關建議。

研究方法

一、研究設計

本研究採二因子獨立樣本的設計架構，其中一個獨立因子是「問卷調查方式」(modes of survey)，分網路問卷與紙本問卷兩組。網路問卷組係以書面信函方式寄送上網填答邀請函，並要求受試者根據邀請函上的網址及一組密碼上網填答問卷；而紙本問卷組，則直接將紙本問卷寄送給受試者，請其直接在紙本問卷上填答，填寫完畢後以所附回郵信封寄回即可。至於問卷內容，則兩組皆相同。另一個獨立因子則是「問卷議題熟悉度」(issue-familiarity of questionnaire)，根據填答者對問卷所附議題之導讀文章，在第1題的作答情形來決定其對該議題的熟悉度高低。

本研究在問卷設計上，除了開頭的指導語說明外，另包含三題背景變項的調查，分別為性別、學歷及年齡，接著是一篇短文，取材自2006年3月8日聯合報A11版所引述商業週刊的專題報導，受試者在閱讀完該短文後，必須回答13題研究者自行擬定的問題，每個問題後都

有四個反應類別，其形式與李克特式評定量表相似，屬於次序量尺，得分從1分到4分，依序來計算，問卷設計內容詳如附錄。其中，值得一提的是，在13題預擬待答的問題中，第1題的目的即是用來判定受試者在接受問卷前對該篇短文議題的熟悉程度，選項1為「之前沒看過」；選項2為「並未詳看」；選項3為「看過，有點瞭解」；選項4為「看過，非常熟悉」。本研究將選擇1及2的受試者歸類為「低議題熟悉組」，而選擇3及4的受試者歸類為「高議題熟悉組」，以作為本研究有關這個獨立因子的兩個實驗水準（treatment levels）。

網路問卷的設計，節錄部分如圖1所示，所有問卷內容包含指導語、基本資料、短文及13題，全都設計在同一頁面中，受試者無須翻頁，可以自由移動捲軸進行問卷的填答。

項目。

全國各級公立學校總經費現在一年要七百二十億元，主計處官員透露，地方政府共需六百一十一億元，未來十年內，估計一年要一千四百億元。「教育總經費一年才四百二十億元，未來總經費占三分之一，教育經費不增不降？」前全國教師協會會長，向於駐地山國小教師劉政雄如是說。

因為教育經費不足，政府想出來另一個高招，是縮班減校，偏遠小孩被遣散併，偏遠地區弱勢兒童受教權受影響，未來學運經費能力較差的孩子，恐難教育翻身機會愈來愈薄。

根據上海市政府所發表的數字，二〇〇三年上海中小學每生實際經費分別達到約台幣二萬五千元，比前一年增加二成。若認真正用對學生身上的費用，再以當地物價換算，已和台灣相去不遠。

貳、正式問卷：（下列問題係基於上述真實性報導而擬定，請您依據個人的身歷來回答！）

- 您對上述一則有關教育經費問題的新聞報導，是否熟悉？
1. 〇之前沒看過 2. 〇並未詳看 3. 〇看過，有點瞭解 4. 〇看過，非常熟悉
- 您對報導中「我國每位小學生平均分到的教育經費是偏低的區內小兒之說」的看法？
1. 〇非常不擔心 2. 〇不擔心 3. 〇擔心 4. 〇非常擔心
- 您對報導中「扣除人事費後，用到學生身上的「剩錢」僅為總額的十分之一」的看法？
1. 〇非常不擔心 2. 〇不擔心 3. 〇擔心 4. 〇非常擔心
- 您對報導中「官員認為教育經費仍仍遠遠低於預算，是見政府對教育重視」的看法？
1. 〇非常不同意 2. 〇不同意 3. 〇同意 4. 〇非常同意
- 您對報導中「學者認為政府所編列教育經費的數字都是虛的！虛假！吹牛！」的看法？
1. 〇非常不同意 2. 〇不同意 3. 〇同意 4. 〇非常同意
- 您對報導中「政府用各種名目，將財源挪去充當其他教育經費項目」的看法？
1. 〇非常不恰當 2. 〇不恰當 3. 〇恰當 4. 〇非常恰當
- 您對報導中「未來十年內，「總經費」即占三分之一，將會擠現教育經費？」的看法？
1. 〇非常不擔心 2. 〇不擔心 3. 〇擔心 4. 〇非常擔心

圖 1 關於「教育經費問題」的網路問卷設計節錄

在問卷調查研究中，獲得的調查結果是否能夠推論到母群是決定研究成敗的重要因素，因此，樣本的取得是否來自隨機抽樣（random sampling），是一大關鍵要素所在。而在網路上進行問卷調查，最容易產生令人質疑其推論效度的問題，即前述覆蓋誤差（係指目前網路普及率並非百分之百，故採用網路進行問卷調查，容易出現參照群體（即研究者可接近的母群）並非目標群體（即研究者欲推論的母群）的差異問題（Couper, 2000）），以及自我選擇偏差（亦即能夠上網填答問卷者，多半具有性別、學歷、年齡、社經地位等因素的限制，而有較多自願性樣本（Birnbaum, 2004; Buchanan & Smith, 1999; Grandcolas, Rettie, & Marusenko, 2003; Hayslett & Wildemuth, 2004; Riva, Teruzzi, & Anolli, 2003））等問題，故取樣之樣本未必真正具有母群的代表性，因而影響到研究結果的推論效度。因此，本研究在考量上述干擾因素後，擬以全臺灣中小學教師作為本研究的母群，其理由有三：（1）絕大多數的中小學教師具備一定的資訊素養，對電腦與網路的使用並不陌生，較不會造成因資訊能力不足而影響回覆率的問題；（2）中小學教師母群體較易精確掌握，有助「推論」的效度；因為各級學校的名單及所屬的網站或網頁，可從教育部的網站下載，所以，從這些網路上已公開的資訊取得全部教師母群體的資訊並不困難；（3）本研究採分層隨機抽樣，並以隨機分派方式將樣本分成實驗組（即網路組受試者）與對照組（即紙本組受試者），以進行實驗研究。綜合上述，本研究將可改進前人研究方法設計的缺失，而能更嚴謹地控制干擾因素，更能審慎地分析、比較與推論研究結果。故本研究所獲得的結果，應該能有效地推論至中小學教師族群。

本研究的調查對象，係以分層隨機抽樣方式抽取全國1,200名中小學教師，進行有關「教育經費問題」議題之問卷調查。本研究採用的分層方式，是將臺灣分成北、中、南、東四大地理區塊，再依教

育部當年公告之國小、國中、高中、高職學校數，即國小數目為2,657所、國中736所、高中314所及高職164所，不分公私立學校別，進行學校數的隨機抽樣，共計抽得國小70所、國中18所、高中職12所等共100所學校，抽樣結果各約佔各級總學校數的2.4%~2.6%之間。另外，各級學校在北中南東四區的抽樣數及母群數詳如表1所示，卡方考驗結果皆不顯著，表示對每一級學校而言，四區的抽樣比例是一致的。之後再經由上述被抽中之學校網頁內的記載，隨機抽取每校12名教師作為樣本，並隨機分派其中一半（即6名）教師接受「網路問卷調查」，另一半接受「紙本問卷調查」，這兩組即構成本研究的一個獨立因子「調查方式」中的兩個組別，即網路問卷組及紙本問卷組各600名教師當作預定的受試者樣本。值得注意的是，本研究由於受研究經費限制，無法考慮所有分層因子，只能藉由隨機取樣方式來獲得有限可接受的代表性樣本，因此，學校的城鄉位置等因子，無法同時納入考慮。

表1 抽樣學校數與母群學校數之分布

	北區	中區	南區	東區	卡方值 (顯著性)
國小	20/771 [0.0259]	22/831 [0.0265]	21/790 [0.0266]	7/265 [0.0264]	0.007 (1.00)
國中	6/256 [0.0234]	5/200 [0.025]	5/210 [0.0238]	2/70 [0.0286]	0.065 (0.996)
高中	3/141 [0.0213]	2/73 [0.0274]	2/85 [0.0235]	1/15 [0.0667]	1.120 (0.772)
高職	1/55 [0.0182]	1/40 [0.0250]	1/56 [0.0179]	1/13 [0.0769]	1.656 (0.647)

註：斜線前表示抽樣數，斜線後表示母群數，中括號內為相除的結果。

問卷自2006年4月15日寄出，經三個星期的回收等待後，發現回

收率太低，因此，決定進行一次催收工作，於5月15日依同樣的流程及問卷，再寄送一份問卷給未填答的受試者進行催收，截至6月底止，雖然回收情形仍然不佳，但由於研究經費及本研究所操弄的議題有其時效限制，故不擬再進行二次催收，僅根據前兩次寄發問卷後之回收情形為準。剔除少數作答遺漏嚴重的問卷後，最後，總計共回收有效問卷425份（佔35.4%），其中，網路組問卷回收195份（佔45.9%），紙本組問卷回收230份（佔54.1%）。另外，根據正式問卷第1題填答情況的歸類結果，低議題熟悉組有207人（佔48.7%），高議題熟悉組有218人（佔51.3%）。

二、資料分析

由於本研究重點在於比較網路與紙本兩種問卷調查方式在回收率及問卷調查內容上的差異情形，並考慮議題熟悉度對問卷內容作答的影響，以及其對兩種問卷調查方式的調節效果。因此，資料分析主要集中在兩個部分，第一部分是回收率的分析，第二部分是問卷內容的分析。

由於回收率具計數資料的屬性，因此，使用卡方適合度檢定來比較網路組與紙本組回收率差異情形，以及比較高低議題熟悉組回收率差異情形。此外，使用卡方百分比獨立性檢定來考驗本研究中各獨立因子與受試者背景變項的關係情形，由於樣本回收率不高，因此，有關網路組與紙本組回收樣本之間，以及高低議題熟悉組之間，在背景屬性上必須能夠一致的要求更是不可忽略。

在問卷內容的比較上，由於正式問卷各題是以李克特氏量尺的形式來計分，因此，本研究使用試題反應理論中適合此類量尺的評定量尺模式（Rating Scale Model, RSM）來進行資料的分析工作。由於本研究可視為二因子獨立樣本設計的應用，因此，兩個獨立因子的交互

作用效果是必須先探討的，並輔以每個獨立因子的主要效果分析，但是，不同於一般變異數分析的是，本研究是以差異試題功能（Differential Item Functioning, DIF）方式，來進行單純主要效果及主要效果的考驗工作，而不是以受試者的能力參數估計值作為依變項來進行變異數分析。所謂差異試題功能，係指檢驗具有相同能力的受試者，是否會因不同身分或背景變項的干擾，而在試題反應機率上出現差異。我們假定無論是調查方式組別的不同或是對議題熟悉度的高低，都不應該影響對問卷試題的作答形態。值得注意的是，在傳統測驗理論中，雖然也有類似DIF的概念，但並未提出一套較為可行的檢定方式，本研究利用試題反應理論中強調機率的觀點及相關軟體提供的DIF功能，來檢定組別變項對受試者群體在每一試題上的反應機率之差異情形。此外，因為整體樣本回收率並不是很高，無法保證各組回收的樣本具有全然的代表性，因此，若直接使用能力估計值在各組的比較作為依據的話，則可能會導致錯誤的結論。我們假定在大樣本條件下，各組的能力分布是一致的，並以此為基礎來比較各組在問卷各試題上的差異情形。

本研究中有關描述統計及回收率卡方檢定的工作，是使用SPSS 13.0版套裝軟體程式來進行，而有關DIF的檢定則是使用FACETS 3.61版軟體程式來進行。FACETS適用於多相（many-facet）資料，雖然本研究資料並非是一般常見的三相資料，但可利用多相的技巧進行分析，即在本研究中除了將受試者及試題各視為一個相外，另外暫時先將組別也視為第三個相，但是將第三相的組別元素（elements）定錨在平均數為0，標準差為1的條件下，使第三相不會進入模式的估計，但仍可獲得適配統計量及DIF檢定結果；這種作法在Linacre（2006）的使用手冊中有詳細介紹。由於本研究重點在於DIF的檢定，因此，與本研究較無關的報表將省略不列出。有關多相IRT模式

的進一步說明，讀者可以參考Linacre (1989) 及FACETS軟體程式的使用手冊。

在應用多相IRT評定量尺模式對資料進行分析前，我們必須先確認問卷資料的單向度假定 (unidimensionality assumption)，以及各試題與評定量尺模式的適配度是否可以被接受。根據Linacre (2006) 的建議，可進行殘差的主成分分析來提供單向度假設的訊息，其原理是針對資料中無法被模式單向度解釋的變異進行主成分分析，以找出其他潛在的未知向度之數量及大小。殘差主成分分析的結果顯示，原資料的總變異之特徵值為34.8，能夠被轉換後測量值解釋到的特徵值有22.8，佔65.5%，而無法被解釋的殘差特徵值有12.0，佔34.5%；Linacre認為被解釋的特徵值超過60%即可以視為符合單向度假設的情形是好的。另外，在無法被解釋的殘差特徵值中有7.0可以被另外3個未知因子所解釋，其特徵值分別為4.8、1.2、1.0，剩下的特徵值5.0可視為隨機誤差。儘管Linacre認為第一個未知因子特徵值小於3.0是好的，但也暗示有例外發生。綜上所述，雖然第一個未知因子的特徵值4.8超過3.0，但並不嚴重，而被解釋的特徵值也超過所建議的60%，因此，整體而言，單向度假設大致是可以接受的。

接著，進行各試題與評定量尺模式間的適配度檢定 (test of goodness-of-fit)，表2所示即提供兩類適配度指標，分別是INFIT及OUTFIT，但兩者的焦點不一樣。以受試者能力估計值的適配情形而言，前者強調受試者能力在與之接近的試題上之表現情形，後者則強調受試者能力在與之遠離的試題上之表現情形；而以試題難度估計值的適配情形而言，其概念可類推。每類指標皆計算MNSQ及ZSTD兩種值；這裡的MNSQ，係指所有受試者在該試題上的實際觀察值與模式期望值之間差異的標準化殘差均方和，即標準化殘差變異的平均值，以1為標準，大於1即表示資料填答情形超出模式的預期，小於1

表 2 問卷試題與評定量尺模式適配度

試題	INFIT		OUTFIT	
	MNSQ	ZSTD	MNSQ	ZSTD
2	1.08	1.2	1.06	0.8
3	0.97	-0.4	0.88	-1.6
4	1.02	0.4	1.07	1.1
5	0.88	-2.0	0.9	-1.6
6	0.94	-0.9	0.96	-0.6
7	0.93	-1.2	0.91	-1.4
8	1.20	3.0	1.25	3.5
9	1.24	3.7	1.26	3.9
10	1.09	1.4	1.04	0.6
11	1.05	0.8	0.98	-0.3
12	0.66	-6.0	0.67	-5.8
13	0.79	-3.6	0.81	-3.2

即表示低於模式的預期；而ZSTD則是將MNSQ進行Wilson-Hilferty轉換後，以方便對照常態Z分布機率表來判定MNSQ是否達顯著水準（Bond & Fox, 2001）。根據Linacre與Wright（1994）的看法，對於評定量尺而言，MNSQ在0.6~1.4之間是合理的，代表受試者在試題上的反應符合模式預期的範圍。因此，由表2中的MNSQ值可知，本研究問卷內容各試題與模式的適配度良好。但是根據轉換後的ZSTD值來看，有幾題絕對值大於2或3，表示各該題與模式間的適配度不良。然而，根據Linacre（2006）的說法，當MNSQ值可以被接受時，ZSTD值是可以被忽略的。綜合上述，本研究的問卷資料符合單向度假設，各試題與模式的適配度亦佳，因此，有利後續分析工作的繼續進行。

結果與討論

本節主要分為三個部分，分別針對回收率、交互作用效果及主要效果進行分析，詳述如下。

一、卡方檢定

本節的主要目的，在檢定各組回收率的差異及兩個獨立因子與背景變項間在回收率上的關聯性。由表3可知，紙本組的回收率略高於網路組，但卡方適合度檢定結果顯示兩組在回收率上並無差異存在；另外，高議題熟悉組雖然較低議題熟悉組有些微高的回收率，但差異也不顯著。這裡值得探討的是，本研究原先預期對議題熟悉度的操作型定義可能會導致高議題熟悉組較願意填答，而在回收率上出現高於低議題熟悉組的結果。由於受試者對議題的熟悉度高低無法事先操弄，況且實務上為了對問卷內容在各組作有意義的比較，必需要有一套共同試題內容才可以進行DIF的檢定比較，因此，本研究乃採事後藉由受試者填答正式問卷中的第1題來決定受試者對本研究所設定的議題之熟悉度，以判定受試者歸屬於低議題熟悉組或高議題熟悉組，而這種事後界定方式有利於後續問卷內容的比較。儘管採用事後界定的方式，回收率檢定結果仍顯示，高議題熟悉組雖然在回收率上較低議題熟悉組稍微高些，但差異不顯著，這點發現說明這兩組受試者在背景變項上是具同質性的。

從表3各組在背景變項上的獨立性考驗結果顯示，並無出現任何顯著性的關聯存在，也就是說，調查方式或議題熟悉度高低的各組回收率，並沒有受到受試者服務單位、學校區域、性別、學歷、年齡的影響。這個結果意謂著，各組回收樣本在各種可能的背景屬性上是一

表 3 調查方式及議題熟悉度與背景變項的卡方檢定摘要

服務單位 (%)	調查方式		卡方檢定		議題熟悉度		卡方檢定
	網路問卷組 n=195	紙本問卷組 n=230	$\chi^2=2.88$ p=0.09	低熟悉組 n=207	高熟悉組 n=218	$\chi^2=0.29$ p=0.59	
國小	70.8	70.0	$\chi^2=0.20$ p=0.91	69.1	71.6	$\chi^2=0.57$ p=0.75	
國中	18.5	20.0		20.8	17.9		
高中職	10.8	10.0		10.1	10.6		
學校區域 (%)							
北區	30.3	25.7	$\chi^2=1.66$ p=0.65	29.0	26.6	$\chi^2=2.04$ p=0.57	
中區	31.8	33.5		34.8	30.7		
南區	27.2	31.3		27.5	31.2		
東區	10.8	9.6		8.7	11.5		
性別 (%)							
男	54.9	50.9	$\chi^2=0.68$	50.7	54.6	$\chi^2=0.64$	
女	45.1	49.1	p=0.41	49.3	45.4	p=0.43	
學歷 (%)							
專科	1.5	1.8	$\chi^2=1.57$ p=0.46	1.0	2.3	$\chi^2=3.06$ p=0.22	
大學	80.0	74.9		80.6	74.1		
研究所	18.5	23.3		18.4	23.6		
年齡 (%)							
25歲以下	3.1	2.7	$\chi^2=3.63$ p=0.46	4.4	1.4	$\chi^2=9.15$ p=0.06	
26-35歲	49.7	43.8		50.2	43.1		
36-45歲	31.3	30.5		29.8	31.9		
46-55歲	14.4	20.4		14.6	20.4		
56歲以上	1.5	2.7		1.0	3.2		

致的，這對於需要假定各組能力估計值為一致之後續問卷內容的DIF檢定工作而言，是有幫助的。

值得注意的是，由於本研究樣本的回收率並不高，因此，背景變項上的檢定變得十分重要。儘管回收率不高會導致不反應偏差的機率上升，但本研究的目的並非是單獨進行某一種調查方式下的調查結果探討，而是強調兩種調查方式的比較或議題熟悉度高低的比較，因

此，已回收的各組樣本在背景變項上分布的一致性，將有助於本研究對無關因素的控制，對後續研究結果的推論有重大幫助。

二、交互作用效果的DIF分析

這裡的交互作用效果主要是藉由單純主要效果來進行，若出現單純主要效果則必然有交互作用產生。表4所示，是將樣本單純地在網路組中進行高低議題熟悉組之間的主要效果DIF考驗結果，表中的測量值即是指在考量該試題的整體難度值（overall difficulty measure）與該題的偏差值（bias size）後所獲得的局部難度值（local difficulty measure），而偏差值的概念與實際觀察值及理論預期值間的差異有正向關係（Linacre, 1989, 2006）。因此，當實際值與預測值差異愈大時，則偏差值亦愈大，本研究並未列出這些偏差值。

表 4 網路組中的高低議題熟悉組之 DIF 分析

試題	低議題熟悉組		高議題熟悉組		低議題熟悉組減高議題熟悉組		t	自由度	顯著性
	測量值	標準誤	測量值	標準誤	差異值	標準誤			
2	-1.45	0.16	-1.76	0.16	0.31	0.22	1.37	193	0.1730
3	-1.64	0.17	-1.66	0.15	0.02	0.23	0.10	193	0.9205
4	1.73	0.17	2.02	0.16	-0.29	0.23	-1.28	193	0.2035
5	-1.03	0.15	-0.99	0.13	-0.04	0.20	-0.22	193	0.8299
6	2.00	0.18	2.25	0.17	-0.26	0.24	-1.06	193	0.2920
7	-0.76	0.14	-1.08	0.13	0.32	0.20	1.63	193	0.1044
8	1.19	0.15	1.60	0.14	-0.40	0.21	-1.92	193	0.0559
9	0.64	0.14	0.54	0.12	0.10	0.19	0.54	193	0.5930
10	-1.64	0.17	-1.40	0.14	-0.24	0.22	-1.09	193	0.2781
11	-1.61	0.17	-1.74	0.16	0.12	0.23	0.54	193	0.5927
12	1.31	0.16	1.60	0.14	-0.28	0.21	-1.35	193	0.1802
13	1.15	0.15	0.71	0.13	0.44	0.20	2.24	193	0.0264

關於測量值（局部難度值）正負號所代表的意義，亦值得探討。

由於在FACETS這套軟體程式中，我們將試題難度以相對於受試者能力方式來設定，因此，這裡的測量值愈大表示受試者愈傾向於在該試題中勾選較低等第的類別，測量值愈小則表示受試者傾向於在該題中勾選較高等第的類別。為方便說明起見，茲以表4第2題為例說明如下：您對報導中「我國每位小學生平均分配到的教育經費是居於亞洲四小龍之末」的看法？1.非常不擔心；2.不擔心；3.擔心；4.非常擔心。該題在低議題熟悉組的測量值為-1.45，而在高議題熟悉組的測量值為-1.76，顯然後者比前者小，因此，我們可以斷言高議題熟悉組比低議題熟悉組傾向於勾選「擔心」或「非常擔心」，但是兩組在該題的測量值差異檢定結果，並未達顯著差異水準。

由表4所示可知，網路組上的高低議題熟悉組在所有試題中的DIF檢定結果，其中的第2~12題皆顯示高低議題熟悉組之間在這些試題上的DIF並不顯著，第13題雖然顯著水準在0.05以下，但仍未達0.01以下的顯著水準。因此，在網路問卷環境下，高議題熟悉組與低議題熟悉組之間並無實質DIF現象出現。

表5所示，是將樣本單純地在紙本組中進行高低議題熟悉組的主要效果DIF考驗結果，表中自由度在各題中稍有差異，此乃因為紙本組受試者在部分試題上的填寫有遺漏值現象所致，但遺漏情形並不嚴重，所以對研究結果並不會有什麼影響。本研究以0.01顯著水準為基礎來決定差異性是否存在，因此，低議題熟悉組減去高議題熟悉組後，出現正向顯著性差異的試題有第2、3、10題，而出現負向顯著性差異的試題有第6、9、13題。

根據上述顯著性結果，可以將意義陳述如下：高議題熟悉組較低議題熟悉組更擔心「我國每位小學生平均分配到的教育經費是居於亞洲四小龍之末」、「扣除人事費後，用到學生身上的『活錢』僅為南韓的十分之一」、「政府財政困難而導致教育經費不足」；而高議題熟悉

表 5 紙本組中的高低議題熟悉組之 DIF 分析

試題	低議題熟悉組		高議題熟悉組		低議題熟悉組		高議題熟悉組		自由度	顯著性
	測量值	標準誤	測量值	標準誤	差異值	標準誤	差異值	標準誤		
2	-1.36	0.14	-2.10	0.18	0.74	0.23	-3.27	0.228	228	0.0012
3	-1.46	0.14	-2.35	0.20	0.89	0.24	-3.71	0.228	228	0.0003
4	1.66	0.14	2.20	0.16	-0.54	0.21	-2.58	0.228	228	0.0105
5	-0.79	0.12	-1.00	0.14	0.21	0.18	-1.12	0.227	227	0.2626
6	1.99	0.15	2.58	0.18	-0.59	0.23	-2.60	0.227	227	0.0098
7	-0.91	0.13	-1.24	0.14	0.33	0.19	-1.75	0.225	225	0.0823
8	1.30	0.13	1.40	0.14	-0.10	0.19	-0.53	0.227	227	0.5964
9	0.24	0.12	0.93	0.13	-0.69	0.17	-3.97	0.226	226	0.0001
10	-1.18	0.13	-2.01	0.17	0.82	0.22	3.77	0.228	228	0.0002
11	-1.40	0.14	-1.92	0.17	0.52	0.22	-2.37	0.228	228	0.0184
12	1.25	0.13	1.70	0.14	-0.45	0.19	-2.33	0.226	226	0.0206
13	0.81	0.12	1.29	0.13	-0.47	0.18	-2.61	0.225	225	0.0096

組較低議題熟悉組認為「政府把興建體育館、籌辦世運會等費用也算入教育經費項目」是更不恰當的、認為「未來家庭經濟能力較差的孩子，想靠教育翻身的可能性」是更不可能的、並對「整體來說，雖然教育經費不足，但對未來教育品質」的看法更沒信心。

綜合表4及表5的結果可知，在紙本組中的高低議題熟悉組之間的DIF現象，較網路組中的高低議題熟悉組之間的DIF現象嚴重。且高議題熟悉組的反應較低議題熟悉組的反應極端（由測量值的絕對值得知），這意謂著在進行高低議題熟悉組在問卷內容的比較時，單獨選用某一種調查方式所獲得的結論將會有很大的差異；例如：如果只選用紙本方式調查時，可能會出現熟悉該議題者較不熟悉該議題者更擔心國家教育經費的運用，或認為國家教育經費的運用更不恰當等；而如果只選用網路問卷進行調查的話，則會顯示無論熟不熟悉該議題者，對國家教育經費運用的看法均為一致的現象。

表6所示，是將樣本單純地在低議題熟悉組中進行網路組與紙本

表 6 低議題熟悉組中的網路組與紙本組之 DIF 分析

試題	網路組		紙本組		網路組減紙本組				
	測量值	標準誤	測量值	標準誤	差異值	標準誤	t	自由度	顯著性
2	-1.65	0.17	-1.52	0.14	-0.13	0.22	-0.57	205	0.5689
3	-1.86	0.18	-1.63	0.14	-0.23	0.23	-1.02	205	0.3111
4	-1.94	0.17	-1.84	0.14	-0.11	0.22	0.48	205	0.6317
5	-1.17	0.16	-0.90	0.13	-0.28	0.20	-1.35	204	0.1801
6	2.23	0.18	2.19	0.15	0.04	0.24	0.17	204	0.8647
7	-0.86	0.15	-1.02	0.13	0.16	0.20	0.81	203	0.4214
8	-1.37	0.16	-1.44	0.13	-0.08	0.21	-0.37	204	0.7090
9	0.74	0.15	0.26	0.13	0.48	0.20	2.46	204	0.0148
10	-1.86	0.18	-1.33	0.14	-0.53	0.22	-2.37	205	0.0187
11	-1.83	0.18	-1.56	0.14	-0.26	0.23	-1.16	205	0.2462
12	1.50	0.16	1.39	0.13	0.11	0.21	0.50	204	0.6165
13	1.32	0.16	0.91	0.13	0.41	0.20	1.99	203	0.0482

組的主要效果DIF考驗結果。本研究係以0.01顯著水準為基礎，來決定差異性是否存在，由表中顯示可知，並無任何一題達到顯著水準。故對低議題熟悉組而言，使用網路或紙本調查方式所獲得的結果之間是沒有差異的。

表7所示，是將樣本單純地，在高議題熟悉組中進行調查方式的主要效果DIF考驗結果。本研究係以0.01顯著水準為基礎，來決定差異性是否存在，由表中顯示可知，網路組減去紙本組並沒有出現任何正向顯著性差異的試題，但出現負向顯著性差異的試題則只有第13題。在第13題中，網路組減去紙本組的差異值為負值，但網路組的測量值為0.65，紙本組的測量值為1.15，兩者皆為正值，因此，此題DIF現象的意義，可以陳述如下：紙本組較網路組就「整體來說，雖然教育經費不足，但對未來教育品質」的看法更沒信心。這樣的陳述方式，會比如下的陳述為佳：網路組較紙本組就「整體來說，雖然教育經費不足，但對未來教育品質」的看法較有信心。

表 7 高議題熟悉組中的網路組與紙本組之 DIF 分析

試 題	網路組		紙本組		網路組減紙本組				
	測量值	標準誤	測量值	標準誤	差異值	標準誤	<i>t</i>	自由度	顯著性
2	-1.61	0.15	-1.88	0.17	0.27	0.23	1.18	216	0.2384
3	-1.52	0.15	-2.11	0.19	0.59	0.24	2.47	216	0.0144
4	1.86	0.15	2.00	0.15	-0.14	0.22	-0.65	216	0.5189
5	-0.89	0.13	-0.87	0.13	-0.02	0.18	-0.09	216	0.9289
6	2.08	0.16	2.37	0.17	-0.29	0.24	-1.21	216	0.2261
7	-0.97	0.13	-1.09	0.14	0.12	0.19	0.64	215	0.5239
8	1.46	0.14	1.25	0.13	0.21	0.19	1.10	216	0.2741
9	0.49	0.12	0.83	0.12	-0.34	0.17	-2.01	215	0.0460
10	-1.27	0.14	-1.79	0.17	0.52	0.22	2.39	216	0.0177
11	-1.59	0.15	-1.71	0.16	0.13	0.22	0.57	216	0.5724
12	1.46	0.14	1.53	0.14	-0.07	0.19	-0.35	215	0.7295
13	0.65	0.12	1.15	0.13	-0.51	0.18	-2.90	215	0.0042

綜合表6及表7的結果可知，高議題熟悉組與低議題熟悉組之間的DIF現象，無論對網路組或紙本組而言，都幾乎是相同的；亦即，無論對該議題熟悉度高或低的教師進行意見蒐集，使用網路及紙本方式來調查，其結果都大致可獲得一致看法的結論。既然，使用網路問卷與紙本問卷的結果是那麼的一致，這似乎暗示著我們可以使用較為經濟快速的網路調查，來取代費時費工的紙本問卷調查的可行性，因為，這兩種調查效果幾乎是一模一樣的。

三、主要效果的DIF分析

這一節要進行的，是調查方式與議題熟悉度在DIF的主要效果分析。就一般的二因子變異數分析而言，當交互作用顯著後，便不需再進行各因子的主要效果考驗，而是直接進行單純主要效果的考驗。但此處爲了要更能看出網路問卷與紙本問卷在問卷內容上的相似性，以及對議題熟悉度高低的受試者之間在各試題看法的差異，因此，

列出表8及表9的考驗結果。

由表8所示可知，網路組與紙本組之間並無DIF的主要效果，顯見網路問卷與紙本問卷在問卷內容上的效果幾乎是相同的。這項結果隱含對未來調查研究效率大幅提升的價值，亦即，使用較經濟快速的網路問卷調查方式，能獲得與傳統紙本調查方式相似的結果，足以大大提升研究的效率。

表8 網路組與紙本組的DIF分析

試題	網路組		紙本組		網路組減紙本組				
	測量值	標準誤	測量值	標準誤	差異值	標準誤	t	自由度	顯著性
2	-1.64	0.11	-1.66	0.11	0.02	0.16	0.11	423	0.9121
3	-1.68	0.11	-1.80	0.11	0.12	0.16	0.76	423	0.4454
4	1.91	0.11	1.90	0.10	0.02	0.15	0.10	423	0.9183
5	-1.02	0.10	-0.88	0.09	-0.14	0.13	-1.01	422	0.3143
6	2.16	0.12	2.25	0.11	-0.09	0.17	-0.52	422	0.6033
7	-0.94	0.10	-1.05	0.09	0.11	0.14	0.82	420	0.4144
8	1.44	0.10	1.34	0.09	0.10	0.14	0.70	422	0.4828
9	0.59	0.09	0.55	0.09	0.04	0.13	0.32	421	0.7517
10	-1.53	0.11	-1.51	0.10	-0.01	0.15	-0.10	423	0.9217
11	-1.70	0.11	-1.61	0.11	-0.09	0.16	-0.59	423	0.5523
12	1.49	0.11	1.45	0.10	0.04	0.14	0.31	421	0.7601
13	0.91	0.10	1.03	0.09	-0.12	0.13	-0.87	420	0.3837

表9所示，是低議題熟悉組與高議題熟悉組之間的DIF主要效果。這裡同樣以0.01顯著水準作為判斷顯著性的標準，低議題熟悉組減去高議題熟悉組後，出現正向顯著差異性的試題有第2、3題，而出現負向顯著差異的試題有第4、6、9、12題。其意義陳述如下：高議題熟悉組較低議題熟悉組更擔心「我國每位小學生平均分配到的教育經費是居於亞洲四小龍之末」、「扣除人事費後，用到學生身上的『活錢』僅為南韓的十分之一」；而高議題熟悉組較低議題熟悉組傾向更

表9 低議題熟悉組與高議題熟悉組的DIF分析

試題	低議題熟悉組		高議題熟悉組		低議題熟悉組減高議題熟悉組				
	測量值	標準誤	測量值	標準誤	差異值	標準誤	t	自由度	顯著性
2	-1.41	0.10	-1.93	0.12	0.52	0.16	3.31	423	0.0010
3	-1.54	0.11	-1.97	0.12	0.43	0.16	2.67	423	0.0078
4	1.69	0.11	2.12	0.11	-0.43	0.15	-2.77	423	0.0058
5	-0.90	0.10	-1.00	0.10	0.10	0.13	0.76	422	0.4450
6	2.00	0.11	2.43	0.12	-0.43	0.17	-2.59	422	0.0100
7	-0.84	0.09	-1.16	0.10	0.32	0.14	2.32	420	0.0207
8	1.26	0.10	1.50	0.10	-0.24	0.14	-1.72	422	0.0855
9	0.40	0.09	0.73	0.09	-0.33	0.13	-2.59	421	0.0100
10	-1.37	0.10	-1.68	0.11	0.31	0.15	2.02	423	0.0435
11	-1.49	0.11	-1.83	0.12	0.34	0.16	2.16	423	0.0316
12	1.28	0.10	1.65	0.10	-0.37	0.14	-2.65	421	0.0084
13	0.95	0.09	1.00	0.09	-0.05	0.13	-0.38	420	0.7024

不同意「官員認為教育經費編列仍超過法定預算，足見政府對教育重視」的說法、更認為「政府把興建體育館、籌辦世運會等費用也算入教育經費項目」是不恰當的、並認為「未來家庭經濟能力較差的孩子，想靠教育翻身的可能性」是更不可能的、且更認為「整體來說，我國教育經費編列項目所涵蓋內容」為不合理的。

綜合表8及表9所述，以及上一節的單純主要效果分析可知，網路與紙本兩種調查方式的差異不受議題熟悉度高低的調節，亦即議題熟悉度不會影響網路與紙本問卷兩種調查方式的比較；然而，對議題熟悉度高低不同的受試者之調查結果的差異，卻會受到使用網路或紙本問卷的影響，其中，當使用網路問卷時，高議題熟悉組與低議題熟悉組結果之間並無太大差異，但是使用紙本問卷時，高低議題熟悉組的問卷結果之間卻出現不少的差異，且高議題熟悉組的反應通常較低議題熟悉組為極端。

結論與建議

本研究使用IRT的方法學及應用DIF的檢定概念，針對調查方式及議題熟悉度在問卷內容的影響效果做深入的探討與分析，主要關心的是網路與紙本問卷兩種調查方式、調查議題的熟悉度兩個變項在問卷內容上的主要效果是否顯著不同，以及議題熟悉度對兩種調查方式的調節效果是否明顯；此外，本研究亦關心使用網路問卷與紙本問卷，對問卷回收率的高低是否有影響。最後，本研究綜合前述研究結果與討論，獲得下列重要結論，同時，亦提出建議供未來研究參考。

一、結論

歸納前述結果與討論，獲得幾點結論如下：

(一) 在問卷回收率上，調查方式與議題熟悉度之間並無交互作用效果存在，也無主要效果存在。亦即，網路問卷與紙本問卷在回收率上沒有差異，而高議題熟悉組與低議題熟悉組的回收率也無差異。

(二) 在問卷內容上，調查方式與議題熟悉度之間有交互作用效果存在，其中，針對差異試題功能（DIF）的單純主要效果檢定發現，在高議題熟悉組或低議題熟悉組中，網路組與紙本組之間幾乎沒有出現DIF的試題；但是就高低議題熟悉組之間的比較而言，在網路組出現DIF的試題與在紙本組出現DIF的試題有所不同。

(三) 調查方式的DIF主要效果顯示，網路組與紙本組之間並無顯著差異存在；但高低議題熟悉組的DIF主要效果顯示，高低議題熟悉組之間有許多DIF試題出現，且高議題熟悉組的反應較低議題熟悉組為極端。

(四) 未來的問卷調查方式，似乎可以考慮使用網路問卷來取代

紙本問卷，或者是兩者並行，因為，其效果是相同的。此項結論至少在教師母群的推論上是可以被接受的，本研究也相信隨著網路更加普及後，此項結果可以推論的範圍將更廣。

二、建議

根據結論，提出幾項建議供未來研究時參考：

(一) 未來的問卷調查研究，似乎可以考慮使用較為節省人力財力的網路問卷來代替傳統的紙本問卷。因為，無論在低議題熟悉組或高議題熟悉組，網路與紙本問卷的結果均為非常相似，因此，不論採用哪一種調查方式，都不會影響對推論結果的判斷。另外，在回收率方面，網路問卷雖然略少於紙本問卷，但是差異並不顯著，因此，使用網路問卷並不會有實質較低的回收率問題。最重要的一點是，在樣本組成的背景屬性上，網路組與紙本組之間也是沒有顯著差異的，即使回收率較低，對研究結果的推論也不會造成太大的衝擊。

(二) 除了本研究針對網路與紙本調查方式的比較探索外，對於其他不同調查方法間的差異研究也是值得關注的；例如：選舉前的民意調查研究常使用電話訪談，它與網路問卷、紙本問卷或面對面訪談是否有差異也應當研究。此外，未來研究者若對於議題熟悉度高低在問卷內容上的比較研究有興趣，應當小心使用調查方式，不同的調查方式可能會獲得不同的結論，因此，應該謹慎使用特定的調查方式，必要時應使用多種調查方式，以免誤導結論。

(三) 由於對議題熟悉度的不同會產生問卷作答上的差異，而熟悉度高低與受試者對該議題的興趣有一定的關聯，因此，受試者對議題興趣的高低也可能會影響問卷內容的回答。這似乎意味著未來在進行有關時事議題的調查研究時，調查員有必要確認受訪者對該議題的興趣或熟悉程度，在相似的熟悉背景下，所填答的結果應該更為可

信，尤其對政府某些重要決策而言，不應該受不熟悉相關議題者的影響才是。

(四) 由於本研究所使用的樣本是中小學教師，並非一般大眾或其他群體的受試者；亦即，教師群體所具有的特徵屬性，可能不完全與一般大眾或其他群體相似，因此，本研究結果的推論，目前僅能適用於中小學教師母群體。未來的研究，可以考慮針對不同背景屬性差異較大的群體進行分層取樣，然後重複本研究及其研究方法，以進行確認。不過，應注意的一點是，當所界定目標母群背景範圍愈大時，所應該回收的樣本數就必須愈大，才能讓不同背景的次樣本都具有代表性，使所獲得的結果更為可信。此外，本研究結果亦奠基於特定的問卷議題，未來的研究亦可嘗試其他議題，以提供更多網路與紙本問卷調查方式的比較結果。最後，問卷議題與研究對象群體之間的交互作用，可能也會影響研究結果，亦值得未來進一步探討。

參考文獻

- Bandilla, W., Bosnjak, M., & Altdorfer, P. (2003). A comparison of web-based and traditional written self-administered surveys using the ISSP environment module. *Social Science Computer Review, 21*, 235-243.
- Bason, J. J. (2000). *Comparison of telephone, mail, web, and IVR surveys of drug and alcohol use among University of Georgia students*. Paper presented at the American Association of Public Opinion Research, Portland, Oregon.
- Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology, 55*, 803-822.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum

- Associates.
- Buchanan, T., & Smith, J. L. (1999). Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology, 90*, 125-144.
- Couper, M. P. (2000). Web survey: A review of issue and approaches. *Public Opinion Quarterly, 64*, 464-495.
- Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly, 64*, 413-428.
- Denscombe, M. (2006). Web-based questionnaires and the mode effect: An evaluation based on completion rates and data contents of near-identical questionnaires delivered in different modes. *Social Science Computer Review, 24*(2), 246-254.
- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method*. New York: John Wiley.
- Dillman, D. A., & Bowker, D. K. (2001). *The Web questionnaire challenge to survey methodologists*. Retrieved December 18, 2006, from http://survey.sesrc.wsu.edu/dillman/zuma_paper_dillman_bowker.pdf
- Goho, J. (2002). *Mixed mode effects in a community college graduate survey*. Paper presented at the 42nd Annual Forum of the Association for Institutional Research, Toronto, Canada.
- Grandcolas, U., Rettie, R., & Marusenko, K. (2003). Web survey bias: Sample or mode effect? *Journal of Marketing Management, 19*, 541-561.
- Groves, R. M. (1989). *Survey errors and survey costs*. New York: John Wiley and Sons.
- Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*, 2-31.

- Groves, R. M., Singer, E., & Corning, A. (2000). Leverage-Salience theory of survey participation: Description and an illustration. *Public Opinion Quarterly*, 64, 299-308.
- Hayslett, M. M., & Wildemuth, B. M. (2004). Pixels or pencils? The relative effectiveness of Web-based versus paper surveys. *Library & Information Science Research*, 26, 73-93.
- Huang, H. M. (2006). Do print and Web surveys provide the same results? *Computers in Human Behavior*, 22, 334-350.
- Katja, L. M., & Vasja, V. (2004). *A meta-analysis of response rates in Web surveys compared to other survey modes*. Paper presented at American Association for Public Opinion Research Annual Meeting, Phoenix, AZ.
- Keeter, S., Miller, C., Kohut, A., Groves, R. M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, 64, 125-148.
- Leung, D. Y. P., & Kember, D. (2005). Comparability of data gathered from evaluation questionnaires on paper and through the internet. *Research in Higher Education*, 46(5), 571-591.
- Linacre, J. M. (1989). *Many-facet Rasch measurement*. Chicago: MESA.
- Linacre, J. M. (2006). *A user's guide to FACETS*. Retrieved October 18, 2007, from <http://www.winsteps.com/facets.htm>
- Linacre, J. M., & Wright, B. D. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.
- McCabe, S. E. (2004). Comparison of web and mail surveys in collecting illicit drug use data: A randomized experiment. *Journal of Drug Education*, 34(1), 61-72.
- Morrel-Samuels, P. (2003). Web surveys' hidden hazards. *Harvard Business Review*, 81(7), 16-18.

- Newell, C. E., Rosenfeld, P., Harris, R. N., & Hindelang, R. L. (2004). Reasons for nonresponse on U.S. Navy Surveys: A closer look. *Military Psychology, 16*(4), 265-276.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Riva, G., Teruzzi, T., & Anolli, L. (2003). The use of the internet in psychological research: Comparison of online and offline questionnaires. *Cyberpsychology and Behavior, 6*, 73-80.
- Shaw, D., & Davis, C. H. (1996). The Modern Language Association: Electronic and paper surveys of computer-based tool use. *Journal of the American Society for Information Science, 47*, 932-940.
- Synodinos, N. E., Papacostas, C. S., & Okimoto, G. M. (1994). Computer administered versus paper-and-pencil surveys and the effect of sample selection. *Behavior Research Methods, Instruments and Computers, 26*, 395-401.
- Tse, A. C. B. (1998). Comparing the response rate, response speed and response quality of two methods of sending questionnaires: E-mail vs. mail. *Journal of the Market Research Society, 40*(4), 353-361.



附錄 教育經費問題調查問卷

親愛的老師，您好，您辛苦了：

本中心接受國科會委託，為瞭解教師對國內教育經費支出問題的看法，特別設計發展本問卷，以進行學術調查研究之用，並期盼研究成果能對政府編列教育經費時提供實徵性的建言。您是被抽樣邀請來填寫此份調查問卷的教師代表之一，首先感謝您在百忙之中，能夠抽空填寫這份調查問卷。

這份問卷的填寫時間不超過5分鐘。請您針對每題所述狀況，勾選一項最能夠反應您對教育經費問題看法的選項即可。您所填寫的任何資料，本中心皆會予以保密，未經您書面同意之前，本中心不會公開任何個人資料或做為其他用途，敬請您放心填寫。填寫完畢，請將本問卷對摺，裝入所附回郵信封寄回即可。

再次感謝您的協助！ 敬祝 身體健康！順心如意！

國立政治大學教育學院

華文測驗與教育評鑑研究中心

余民寧 教授 敬上

2006年5月15日

壹、填寫者基本資料：

- 1.性別：1. 男 2. 女
- 2.學歷：1. 專科 2. 大學 3. 研究所（含以上）
- 3.年齡：1. 25歲以下 2. 26-35歲 3. 36-45歲 4. 46-55歲
5. 56歲以上



【背景說明】：以下是2006/3/8聯合報A11版【話題】所引述商業週刊的專題報導

我教育經費 低於港韓星 扣掉93%人事費用後 更是低得難看！

「苦不能苦孩子，窮不能窮教育」，是大陸農村隨處可見的標語。最近，行政院長蘇貞昌也有類似的宣示，但是蘇院長或許不清楚：臺灣的教育投資，真正用在學生身上的費用，不到南韓的十分之一，更將逐漸被上海超越。

明天（2006/3/8）上市的商業周刊報導，據教育部統計，2004年我國每名小學生分到的教育經費約2,089美元（約合台幣七萬元），比香港的近3,800美元、南韓的3,700多美元、新加坡的2,400美元都低。

更糟的是，人事費佔了臺灣教育經費的九成三，再扣除徵收土地、蓋校舍的3%後，我國用在中小學生身上的「活錢」，只有4%不到、約80美元，是南韓小學生（850多美元）的十分之一不到。

臺中縣梧南國小是所十個班、200多名學生的小學。校長陳水順說，94年度的預算中，老師的「教育訓練費」掛零，水電費被縣府砍成去年的一半，房屋修繕費只有2萬6,000元，連粉刷牆壁都不夠。中午孩子們的營養午餐只有兩個素菜和豆腐清湯。

「我女兒學校在周三下午舉辦班親會，因為學校要省電費，不能排在晚上。」一位臺北縣國中老師無奈表示，「這會讓多少家長無法參加？」

對於臺灣在亞洲國家投資教育競爭力的落後，行政院主計處副主計長陳慶財表示，每年各級政府編列的教育經費，其實都超過「教育經費編列與管理法」規定的21.5%下限。「對教育的重視度，我們絕對不會亞於外面。你看，政府財政那麼困難，我們還是編到24.2%。」

關心教育的前中央大學教授王顯達不客氣指出「經費數字都是假的。」前全國教改協會理事長丁志仁也說：「膨風！吹牛！那是預算數字，決算（真正執行）哪有這麼高！」

高雄市教師會政策中心執行秘書任懷鳴所做的「94年度高雄市教育預算問題」分析報告指出：「市立體育場經費」、「巨蛋體育館地價稅」、「補助綜合體育館興建經費」、「2009世運會籌辦費」等8億元，全都灌入了教育項目。

全國各級公立學校退撫費現在一年要720億元，主計處官員透露，地方政府負擔611億元。未來十年內，估計一年要1,400億元。「教育總經費一年才4,200億元，未來退撫費佔三分之一，教育經費豈不爆掉？」前全國教師會秘書長、南投縣虎山國小教師劉欽旭擔憂的說。

因為教育經費不足，政府想出來另一個高招，是減班減校，偏遠小校被迫裁併。偏遠地區弱勢兒童受教權受影響，未來家庭經濟能力較差的孩子，想靠教育翻身的機會愈來愈低。

根據上海市政府所發表的數字，2003年上海中小學每生實際經費分別達到約台幣2萬5,000元，比前一年增加二成。若論真正用到學生身上的費用，再以當地物價換算，已和臺灣相去不遠。

貳、正式問卷：（下列問題是基於上述真實性報導而擬定，請您依據您個人的看法來回答！）

1. 您針對上述一則有關教育經費問題的新聞報導，是否熟悉？
 1. 之前沒看過
 2. 並未詳看
 3. 看過，有點瞭解
 4. 看過，非常熟悉
2. 您對報導中「我國每位小學生平均分配到的教育經費是居於亞洲四

- 小龍之末」的看法？
1. 非常不擔心 2. 不擔心 3. 擔心 4. 非常擔心
3. 您對報導中「扣除人事費後，用到學生身上的『活錢』僅為南韓的十分之一」的看法？
1. 非常不擔心 2. 不擔心 3. 擔心 4. 非常擔心
4. 您對報導中「官員認為教育經費編列仍超過法定預算，足見政府對教育重視」的看法？
1. 非常不同意 2. 不同意 3. 同意 4. 非常同意
5. 您對報導中「學者認為政府所編列教育經費的數字都是假的！膨風！吹牛！」的看法？
1. 非常不同意 2. 不同意 3. 同意 4. 非常同意
6. 您對報導中「政府把興建體育館、籌辦世運會等費用也算入教育經費項目」的看法？
1. 非常不恰當 2. 不恰當 3. 恰當 4. 非常恰當
7. 您對報導中「未來十年內，『退撫費』即佔三分之一，將會擠爆教育經費？」的看法？
1. 非常不擔心 2. 不擔心 3. 擔心 4. 非常擔心
8. 您對報導中「政府以減班減校作為應付教育經費短絀的作法」的看法？
1. 非常不可行 2. 不可行 3. 可行 4. 非常可行
9. 您對報導中「未來家庭經濟能力較差的孩子，想靠教育翻身的可能性」的看法？
1. 非常不可能 2. 不可能 3. 有可能 4. 非常有可能
10. 整體來說，您對我國「政府財政困難而導致教育經費不足」的看法？
1. 非常不擔心 2. 不擔心 3. 擔心 4. 非常擔心

11. 整體來說，您對我國「因為教育經費不足而導致投資教育競爭力落後」的看法？
1. 非常不擔心 2. 不擔心 3. 擔心 4. 非常擔心
12. 整體來說，您對我國「教育經費編列項目所涵蓋內容」的看法？
1. 非常不合理 2. 不合理 3. 合理 4. 非常合理
13. 整體來說，雖然教育經費不足，但您對「未來教育品質」的看法？
1. 非常沒信心 2. 沒信心 3. 有信心 4. 非常有信心



The Effects of Survey Modes and Issue-Familiarity of Questionnaire on the Response Rates and the Surveyed Contents

Min-Ning Yu* Ren-Hau Li**

Abstract

The main purpose of this study was to investigate the effects of survey modes and issue-familiarity of questionnaires on the response rates and the surveyed contents. The questionnaires were administered to a stratified sample of 1,200 teachers who were randomly selected from elementary and secondary schools in Taiwan. The overall returned samples were 425 teachers and the response rate was 35.4%. Of the returned samples, the Web-based questionnaire group had 195 teachers (45.9%), and the paper-based questionnaire group had 230 teachers (54.1%). This study not only used the chi-square test on the comparison of response rates, but also used the Rasch model converting Likert-type scale into continuous scale data and followed by analyses on surveyed content. This study found that: (1) there were neither interaction effects nor main effects on the response rates between the survey modes and

* Min-Ning Yu: Professor, Department of Education, National Chengchi University

** Ren-Hau Li: Assistant Professor, Department of Psychology, Chung Shan Medical University

issue-familiarity of questionnaires; (2) there were interactions on surveyed content between survey modes and issue-familiarity of questionnaire; yet, by simple main effect tests of differential item functioning (DIF), no matter in high or low issue-familiarity group, there was almost no DIF item appearing between Web-based group and paper-based group, but items that appeared differences between high issue-familiarity group and low issue-familiarity group in Web-based group were much more different from those in paper-based group; (3) main effect test of DIF for survey modes showed that there was no DIF between Web-based group and paper-based group. However, there was serious DIF phenomenon between high issue-familiarity group and low issue-familiarity group, and responses in high issue-familiarity group were more extreme than in low issue-familiarity group; (4) in sum, Web-based questionnaires could be suggested for substitution for paper-based questionnaires or for using them both together on the future researches, because they almost have the same effects. Finally, some suggestions for future studies were also proposed.

Keywords: web-based questionnaire, survey mode effects, issue-familiarity, Rasch model

