

第五章 結論與未來研究方向

第一節 結論

一、抽樣設計與抽樣權重

本研究藉由國外大型測驗抽樣設計之探討，可知大型測驗主要皆是透過多階段的抽樣方法抽取受試樣本，主要分成兩個階段，包含受試學校與受試學生的選取。因此，檢視現行TASA的抽樣設計是否有需改善的部分，經研究發現TASA抽樣設計推估抽樣權重時，若不考慮城鄉層級的計算，最終權重總合與母群體受試人數相比有低估的情況發生（以TASA 2006年數學科為例）。由於TASA樣本學校被抽取的機率並沒有依據PPS的抽樣方式，才造成學生最終權重有高估或低估的情況發生，因此，本研究建議應調整樣本學校的權重（即調整樣本學校被抽取的機率），以正確估計受試者權重。

是故，本研究建議 TASA 使用二階段分層抽樣設計（國中與國小的部分），第一階為分層叢集隨機抽樣，根據縣市與班級數兩個變項進行分層；第二階段再根據所抽取的樣本學校，以學生個人為抽樣單位進行簡單隨機抽樣。且因應 2010 年台灣行政地區之改變，以及為了進行縣市或地區性受試樣本成就表現之比較，將依據五都、四個地理區、離島地區劃分 PSU。透過此抽樣設計，並採用 PISA 的 Fay 複製方法估計抽樣變異。

二、測量模式

（一）選擇題部分，建議 TASA 繼續使用三參數對數模式，題組題的部分，使用題組模式進行分析；開放性試題部分，目前 TASA 針對開放性試題僅進行敘述統計分析，未來建議使用一般化部分給分模式進行試題分析。

（二）模式適合度評估方法，TASA 目前尚未使用任何方法進行模式適合度評估，建議參考 NAEP、TIMSS 使用圖形化判斷方式進行模式適合度評估。

三、試題資料

選擇題部分，NAEP、TIMSS 及 TASA 皆採用三參數對數模式，因此 TASA 目前之分析方法可持續沿用，但是對於開放性試題僅提供敘述統計資料，數據略嫌不足，需進一步進行更深入的試題參數估計，以便獲得更多試題訊息。另外，針對開放性試題進行二次評分之評分者間一致性檢定，為目前 TASA 所缺乏的，因此 TASA 於之後之正式施測中，對每一開放性試題應要求至少經過兩位閱卷者評分，如此有助於進行二次評分之評分者間一致性檢測，提高 TASA 測驗評量之信度。DIF 分析部份，TASA 問卷背景調查中包含性別、國籍原生性及慣用語言等資料，因此未來有必要挑選部分變項進行全面性之差異試題功能分析探討，以降低試題在族群間產生作答反應差異之情況發生。

四、問卷背景變項資料

TASA 現行分析以三參數對數模式為主，學生能力值為 EAP 期望後驗估計所得，加上在抽樣上並無抽樣權重之設計，因此分析上並無考慮抽樣誤差及測量誤差，此與 NAEP、PISA 及 TIMSS 有極大之差異。因此，增加抽樣權重與可能值之計算，藉以降低抽樣誤差及測量誤差，為 TASA 目前需改進之方向。另外，問卷題項分析前之信、效度分析乃是基本原則，對各國際大型資料庫而言，皆為不可或缺之一環，因此，TASA 針對問卷背景變項分析，可考慮增加 Cronbach's alpha 信度分析與 CFA 建構效度分析。

五、量尺化程序

(一) 測驗實施與題本設計

在題本設計，考量到 TASA 目前使用題本設計之方法，其實與 NAEP、PISA 同為 BIB 設計，所以建議 TASA 在題本設計繼續使用 BIB 設計。

(二) 可能值方法

可能值方法是透過潛在迴歸模式，加入學生背景變項計算後驗分佈，從後驗分佈中隨機抽取可能值，建議 TASA 應以可能值方法進行量尺化程序，步驟如下所示：(1)將性別、社經背景直接視為是條件變數。(2)將學生問卷中的變數虛擬編碼。(3)使用主成分分析分析虛擬編碼的變數，並且計算每一位學生的主成分分數，主成份的數量必須要能解釋原始資料90%的變異。(4)使用已校正的試題參數，和經由主成分分析得到的條件變數估計群體參數分佈。(5)使用上述的方法抽取五個可能值向量。

建議 TASA 在同年度間量尺化方法一樣繼續採用同時估計法來進行量尺化，而 TASA 在不同年度間量尺化過程上，建議採用 PISA 在不同年度間量尺化方法進行等化。

(三) 建立試題圖

TASA2006 年小四數學有針對樣本試題進行試題圖的繪製，但在 2009 年並無相關的試題圖內容，因此建議 TASA 可以在成果報告中加入試題圖的分析，使得 TASA 更加完善，並在試題圖中加入 MC 與 CR 試題的區別，讓部份給分的試題也能夠加入分析中，而且可以針對每個試題進行題目本身的質性探究，說明試題所對應的知識及技能。

六、出版報告

(一) 增加大眾版成果報告

TASA 目前尚無以非測統專業人員為閱讀者之成果報告，因此提出增加大眾版成果報告之建議，並朝「呈現主題明確，使讀者能於短時間內讀到重點」、「圖文並重，以淺顯易懂的圖表取代學術性的文字」兩特色做規劃，以期透過大眾版成果報告之呈現，提供教育部長官與台灣民眾一個簡要瞭解 TASA 重點成果之報告。

TASA 大眾版成果報告之內容包括關鍵報告（當年成就表現重點歸納）、測驗簡介（包括當年度測驗內容的改變）、評量架構與評量設計、成就報告、背景變項、技術附註、附錄表格、其他資訊（機關與作者資訊）。

(二) 學術版成果報告內容調整

TASA 現有之學術版成果報告，將「命題架構」、「正式測驗題本試題分配」、「表現水準之界定」、「常模樣本整體表現情形之現況描述」、「不同成就水準的分佈」、「不同地理區域、縣市之學生在學科學習成就的差異分析」、「不同學生背景變項在學科學習成就表現之差異分析」與「不同學校背景變項對學生在學科學習成就表現之差異分析」，接放入「各年段評量結果分析」一章中探討。

為考量成果探討之嚴整性，建議學術版成果報告將「成就報告」、「評價標準」、「學科內容與認知結構成就百分比」與「背景變項」分開章節做探討，並於緒論前增加「執行摘要」一章，對學術版成果報告做一重點歸納。

(三) 技術報告內容調整

綜合 NAEP、PISA、TIMSS、TASA 之技術報告內容分析發現，NAEP 1998、PISA 2006 與 TIMSS 2007 之技術報告中，對於「品質管理」、「檢視權重和抽樣變異」皆有獨立章節探討，因此建議 TASA 技術報告增加「品質管理」、「檢視權重和抽樣變異」兩章節以及「附錄」。

(四) 資料分析手冊

PISA 提供了 SAS 與 SPSS 的資料分析手冊，使研究者了解 PISA 資料庫和使用 SAS、SPSS 來完成分析，建議未來 TASA 提供資料分析手冊給次集資料分析者，提供正確的資料分析程序與方法，避免次級資料分析者誤用。

第二節 未來研究方向

- 一、TASA 自 2005 年始進行施測，即未探討目前抽樣設計將受試樣本權重值視為相同與忽略抽樣變異的估計等可能產生等問題，針對本研究建議之 TASA 抽樣設計權重值的計算，以及原先 TASA 抽樣設計權重值的修正等問題，將是本研究後續應該繼續探討的部分。
- 二、模式適合度之評估方面，大多數的研究均是以單一試題檢驗模式適合度，未來可探討針對整份試卷之模式適合評估。
- 三、以 TASA 的評量架構而言，應是屬於多向度能力之評量架構，未來可探討多向度試題反應理論模式應用於 TASA 資料分析之可行性。
- 四、題本設計上，可將 TASA2006 年 BIB 設計、TASA2007 年 NEAT 設計與 TASA2009 年 BIB 設計量尺化後，探討目前 TASA 現有資料上題本設計在可能值量尺化方法上之影響。
- 五、可能值方法主要以單向度試題反應理論或多向度單參數試題反應理論為基礎，考量 TASA 之評量架構與測驗題型，未來可探究以多向度三參數試題反應理論為基礎之可能值方法。
- 六、可能值方法主要為納入背景變項的考慮，使得回復群體參數更為準確，未來期望進行模擬研究，探討有無納入背景變項，對回復群體參數的影響，並佐以 TASA 實徵資料進行探討。
- 七、目前 TASA 的統計考驗並未納入標準誤的計算，未來建議計算 TASA 的標準

誤，並利用其標準誤進行統計考驗分析。

- 七、針對開放性試題，TASA 未來研究可探討 NAEP、TIMSS 所使用二參數對數模式、一般化部份給分模式與 PISA 所使用的多向度隨機係數多項洛基模式進行參數估計與試題特性分析上之差異與適用性，做為自身參考採用依據。
- 八、問卷背景變項分析部份，TASA 未來研究方向可增加探討在單向度 IRT 與多向度 IRT 理論架構下，DIF 分析、Cronbach's alpha 信度分析與 CFA 建構效度分析是否存在差異性。
- 九、在建立共同量尺上，考量到 PISA 與 TIMSS 使用相同的等化方法，建議 TASA 也採用相同等化方法，未來研究可進行 TASA 實證資料不同年度間量尺分數的估算，並探討運用可能值進行等化後之差異。
- 十、在 TASA2006 年並非每科都有針對試題進行試題圖的分析，所以未來研究可以針對 TASA2006、2007 年、2009 年、2010 年的資料進行試題圖的分析，讓 TASA 資料更加完善。
- 十一、TASA 目前出版報告類別有：技術報告、成果報告、資料釋出光碟以及 TASA 電子報等四種，加上本研究所建議增加的大眾版成果報告，TASA 出版報告共有：技術報告、成果報告(大眾版)、成果報告(學術版)、資料釋出光碟以及 TASA 電子報等五種，以國內需求而言尚屬完善。未來之研究方向，建議整理國際上使用 NAEP、PISA、TIMSS 資料完成之研究報告，以及其研究題目與成果，以供國內學術研究發展之參考。