

## 第五節 量尺化程序

教育測驗有兩個主要目的，目的一為測量特定學生的知識和技能，學生的表現關乎他或她的未來（職業、入學等），因此降低個體誤差的估計是非常重要的。目的二是評量一個群體的知識和技能，個體的表現將不會影響他們的學校職業或專業生活，這種情況下，降低對目標群體推論的誤差將比降低個體層級的測量誤差更重要，國際的大型測驗是屬於第二個目的（von Davier, Gonzalez, & Mislevy, 2009）。大型測驗的目標是從一群具有代表性的樣本，收集特定內容領域的知識或技巧的相關資料，關注群體的進展情形，因此大型測驗有興趣的統計量數是群體的平均數、標準差、某些層級表現的百分比、百分位數，和關連於上述統計量數的標準誤。

當測驗的內容廣泛，常使用矩陣抽樣設計（matrix-sampling design），使每一位抽樣到的學生僅需作答部份測驗內容，當所有學生的答題反應被收集集合之後，可涵蓋所有的測驗內容。然而在這樣的設計之下，由於學生只使用題庫內的一部分試題測量，個體能力的測量會伴隨著相當程度的測量誤差，傳統的個體能力的估計方法，像是最大概似估計法（maximum likelihood estimation, MLE）、期望後驗法（expected a posterior, EAP）這種對於學生個別能力的估計提供最佳的點估計的方式，並不適用於群體能力的估計，集合個體的能力值估計群體的特性將會產生嚴重的偏誤。（Foy, Galia, & Li, 2008; Lee, Grigg & Dion, 2007 ;OECD, 2005）。

目前 NAEP、TIMSS 和 PISA 等大型測驗是以試題反應理論（item response theory, IRT）為基礎，透過多重插補，也就是「可能值方法」（plausible values methodology）（Allen, Carlson, Johnson, & Mislevy, 1999; Foy, Galia, & Li, 2008; OECD, 2005）進行量尺化程序，可能值方法是以潛在迴歸模式，加入學生背景變項計算後驗分佈，並抽取可能值，以利於次級資料分析者使用，可能值方法中沒有先估計個體的能力再計算群體參數，而是使用可得的資料，包含學生的答題反

應和背景變項資料直接估計母群的參數，可以使參數的估計較準確（Mislevy & Sheehan, 1989）。以下將以 NAEP1998(Allen, Carlson, Johnson, & Mislevy, 1999)，PISA2003（OECD, 2005）和 TIMSS2007（Foy, Galia, & Li, 2008）的技術報告為主，針對這三大測驗與國內目前正在進行之 TASA 所使用的量尺化程序作一整理說明。

## 壹、測驗實施與題本設計

不同測驗之量尺化程序的方式會依據測驗實施目的與測驗題本設計的不同而有所差異，因此，要瞭解測驗之量尺化程序，首先必須知道測驗之實施目的與題本設計的方式。

### 一、測驗實施

NAEP 測驗設計共有三種，包含：（1）主要測驗（main assessment）；（2）長期趨勢測驗（assessment for long-term trend）；（3）特殊測驗（special assessment）。PISA 主要是測量國際間 15 歲學生在閱讀、數學與科學之知識能力。每隔三年定期施測此三項領域，並且從中挑出一項當作主要領域，其餘兩項則為次要領域，2000 主要領域為閱讀，2003 年為數學，2006 年為科學。TIMSS 主要測量學生之數學與科學能力，從 1995 年開始每隔四年定期施測，每位學生皆會接受到數學與科學兩項領域。TASA 施測對象為臺灣 4、6、8 年級與高中職二年級學生，施測科目包含國語文、英語文、數學、自然、社會，以評量臺灣學生學習成就表現。以下就三大測驗與 TASA 之測驗實施年級與時間、測驗科目整理成表 4-5-1，表 4-5-2 作一綜合性比較。

表 4-5-1 測驗實施年級與時間之綜合比較

|       |  |
|-------|--|
| NAEP  | 主要測驗：4、8、12 年級學生<br>長期趨勢測驗：9、13、17 歲學生   |
| PISA  | 15 歲學生，每三年舉辦一次   |
| TIMSS | 4、8 年級學生，每三年舉辦一次   |
| TASA  | 4、6、8 年級與高中職二年級<br>2005 年施測 6 年級<br>2006 年施測 4、6、8 年級與高中職二年級<br>2007 年施測 4、6、8 年級與高中職二年級<br>2009 年以後每年施測順序依序為 4 與 6 年級、8 年級、高中職二年級 |

表 4-5-2 測驗科目之綜合比較

|       |  |
|-------|--|
| NAEP  | 主要測驗：數學、科學、閱讀、寫作、社會學、美國歷史、公民、地理、文學、音樂、美術、電腦技能等<br>長期趨勢測驗：閱讀與寫作、數學與科學等                      |
| PISA  | 每次評量會從數學、科學與閱讀三種領域中選出一種為主要領域、其他兩種為次要領域。<br>2000 年主要領域為閱讀<br>2003 年主要領域為數學<br>2006 年主要領域為科學 |
| TIMSS | 數學與科學  |
| TASA  | 2005 年施測科目為國語文、英語文、數學<br>2006 年以後施測科目為國語文、英語文、數學、社會、自然<br>※社會科小四不施測<br>※2009 年以後英語文小四不施測   |

## 二、題本設計

當題本設計-集中式 BIB 設計 (focused balanced incomplete block) 在 1988 年提出後，BIB 與 PBIB (partially balanced incomplete block) 等化設計之變化類型一直沿用至今。BIB 設計與 PBIB 設計之題本皆由試題組合成之區塊連結組合而成，故學生不需作答太多的試題，僅是接受部份試題區塊，且藉由學生接受部份相同之試題區塊來連結題本。而集中式 BIB 與 PBIB 設計提供特定科目之估計，集中式 (focused) 指的是題本由相同科目之試題區塊組合而成，即每位學生接受相同之受測科目 (Nancy, James & John, 2001)。

### (一) BIB 設計

BIB 設計是由 Yates (1936) 提出，並於 1992 年 Rust & Johnson 應用於測驗領域的題庫設計。此設計是指題庫中所有的試題區塊出現次數是相同的，且成對試題區塊出現於題本中的次數也必須是相同的。所謂的「平衡」是由於成對試題區塊出現於題本中的次數是相同的，因此在成對試題區塊平均數間之比較有相同的精準度。各題本中的試題區塊可能部分相同或完全不同，但是每一個試題區塊在所有題本中出現的次數是一樣的 (Kuehl, 2000；曾玉琳、王暄博、郭伯臣、許天維，2005)，亦即題庫中的每個試題所受測的學生約為相同的。

### (二) PBIB 設計

PBIB 是由 Bose & Nair (1939) 提出，在此設計中各試題區塊出現次數需相等，但是成對試題區塊的出現次數是不完全相同的，亦即某些成對試題區塊的出現次數是多過於其餘成對試題區塊的出現次數。平衡設計在需要的不完整區塊無法建構出每種實驗情境，使得各區塊需要出現的重複數可能變成過高，題本也隨之增加，但是部分平衡設計在較少的重複數下便可架構出來。

為了清楚呈現大型測驗之題本設計，以下將舉例說大型測驗之題本設計

#### 1.NAEP

以 NAEP1998 年技術報告提到之 4 年級公民為例。公民使用之題本設計為 BIB 設計，而 1998 年 4 年級公民使用的是 6 個試題區塊組合成 18 個題本之 BIB 設計變化類型，其變化的原因是為了讓試題區塊在題本前後出現次數一致，故將第 16 到 18 個題本設計為將 13 到 15 個題本的兩個試題區塊作交換後組成 (Andrew & Terry, 2001)。以表 4-5-3 作說明。

表 4-5-3 NAEP1998 年 4 年級公民題本區塊設計表

| 題本 | 區塊 I | 區塊 II | 題本  | 區塊 I | 區塊 II |
|----|------|-------|-----|------|-------|
| B1 | M1   | M2    | B10 | M4   | M6    |
| B2 | M2   | M3    | B11 | M5   | M1    |
| B3 | M3   | M4    | B12 | M6   | M2    |
| B4 | M4   | M5    | B13 | M1   | M4    |
| B5 | M5   | M6    | B14 | M2   | M5    |
| B6 | M6   | M1    | B15 | M3   | M6    |
| B7 | M1   | M3    | B16 | M4   | M1    |
| B8 | M2   | M4    | B17 | M5   | M2    |
| B9 | M3   | M5    | B18 | M6   | M3    |

資料來源：NAEP1998 Technical Report (p.408)

## 2.PISA

PISA2006 年使用每個題本包含 4 個試題區塊，每個試題區塊在題本中出現次數 4 次，成對試題區塊在各題本中出現次數 1 次之 BIB 設計 (OECD, 2009)，表 4-5-4 為 PISA2006 年之題本區塊設計。

表 4-5-4 PISA2006 年題本區塊設計表

| 題本  | 區塊 I | 區塊 II | 區塊 III | 區塊 IV |
|-----|------|-------|--------|-------|
| B1  | S1   | S2    | S4     | S7    |
| B2  | S2   | S3    | M3     | R1    |
| B3  | S3   | S4    | M4     | M1    |
| B4  | S4   | M3    | S5     | M2    |
| B5  | S5   | S6    | S7     | S3    |
| B6  | S6   | R2    | R1     | S4    |
| B7  | S7   | R1    | M2     | M4    |
| B8  | M1   | M2    | S2     | S6    |
| B9  | M2   | S1    | S3     | R2    |
| B10 | M3   | M4    | S6     | S1    |
| B11 | M4   | S5    | R2     | S2    |
| B12 | R1   | M1    | S1     | S5    |
| B13 | R2   | S7    | M1     | M3    |

註：M 代表數學，S 代表科學，R 代表閱讀

資料來源：PISA2006 Technical Report (p.29)

### 3.TIMSS

每個題本由四個試題區塊組合而成(每個題本均包含數學與科學各兩個試題區塊),而為了連結不同題本,每個試題區塊在題本中出現2次(Graham, Christine, Alka, & Ebru, 2008)。表 4-5-5 為 TIMSS2007 年之題本區塊設計。

4-5-5 TIMSS2007 年題本區塊設計表

| 題本  | 區塊 (Part I) |     | 區塊 (Part II) |     |
|-----|-------------|-----|--------------|-----|
| B1  | M01         | M02 | S01          | S02 |
| B2  | S02         | S03 | M02          | M03 |
| B3  | M03         | M04 | S03          | S04 |
| B4  | S04         | S05 | M04          | M05 |
| B5  | M05         | M06 | S05          | S06 |
| B6  | S06         | S07 | M06          | M07 |
| B7  | M07         | M08 | S07          | S08 |
| B8  | S08         | S09 | M08          | M09 |
| B9  | M09         | M10 | S09          | S10 |
| B10 | S10         | S11 | M10          | M11 |
| B11 | M11         | M12 | S11          | S12 |
| B12 | S12         | S13 | M12          | M13 |
| B13 | M13         | M14 | S13          | S14 |
| B14 | S14         | S01 | M14          | M01 |

註：M 代表數學，S 代表科學

資料來源：TIMSS2007 Technical Report (p.34)

### 4.TASA

題本設計方面,由於英語文包含聽讀與寫說,故採取每個題本由相同試題區塊組成之題本排列設計,為了連結各題本,試題區塊在不同題本出現不只一次;數學在 2005、2006 年使用 PBIB 設計;2007 年施測科目除了英語文外,其餘科目皆使用 NEAT 設計;除此之外,其餘年段與科目皆使用 BIB 設計。表 4-5-6 為 TASA2009 年數學科 4 年級之題本區塊設計(每個題本由 3 個試題區塊組合而成,共計 13 個試題區塊組合成 26 個題本之 BIB 設計)。

表 4-5-6 TASA2009 年數學科 4 年級題本區塊設計表

| 題本序號 | 區塊 I | 區塊 II | 區塊 III | 題本序號 | 區塊 I | 區塊 II | 區塊 III |
|------|------|-------|--------|------|------|-------|--------|
| S1   | M11  | M10   | M1     | S14  | M4   | M1    | M12    |
| S2   | M6   | M11   | M8     | S15  | M6   | M10   | M13    |
| S3   | M2   | M12   | M6     | S16  | M13  | M3    | M7     |
| S4   | M10  | M9    | M7     | S17  | M8   | M9    | M12    |
| S5   | M7   | M11   | M2     | S18  | M4   | M2    | M10    |
| S6   | M7   | M6    | M4     | S19  | M3   | M5    | M6     |
| S7   | M8   | M7    | M1     | S20  | M5   | M8    | M10    |
| S8   | M1   | M6    | M9     | S21  | M2   | M5    | M9     |
| S9   | M12  | M13   | M11    | S22  | M9   | M13   | M4     |
| S10  | M12  | M7    | M5     | S23  | M5   | M1    | M13    |
| S11  | M9   | M3    | M11    | S24  | M13  | M8    | M2     |
| S12  | M10  | M12   | M3     | S25  | M11  | M4    | M5     |
| S13  | M1   | M2    | M3     | S26  | M3   | M4    | M8     |

資料來源：TASA2009 數學科成果報告（頁 12，未出版）

從上述之綜合分析可以發現，不同的大型測驗使用不同的題本設計，整理如表 4-5-7。

表 4-5-7 題本區塊設計之綜合比較

|       |  |
|-------|--|
| NAEP  | 1996 年技術報告中指出各科目之題本區塊設計為<br>數學：BIB 設計<br>科學：BIB 設計<br>閱讀：PBIB 設計<br>寫作：PBIB 設計   |
| PISA  | BIB 設計   |
| TIMSS | 每個題本由四個試題區塊組合而成（每個題本均包含數學與科學各兩個試題區塊），而為了連結不同題本，每個試題區塊在題本中出現 2 次  |
| TASA  | 除英語文與 2005、2006 年數學外，2005 年、2006 年、2009 年其餘科目皆為 BIB 設計<br>2005 年數學為 PBIB 設計，2006 年數學為 BIB 設計<br>2007 年除英語文外為 NEAT 設計<br>英語文因為包含聽讀與寫說，故採取每個題本由相同試題區塊組成之題本設計 |

NAEP 的主要測驗和 TASA 一個題本僅涵蓋一種領域，NAEP1996 年的長期追蹤測驗（NAEP1998 沒有實施長期追蹤測驗）、PISA 和 TIMSS 一個題本可能涵蓋不同領域，整理如表 4-5-8。

表 4-5-8 題本領域之綜合比較

|       |  |
|-------|--|
| NAEP  | 主要測驗：一個題本僅一種領域<br>長期趨勢測驗：一個題本涵蓋閱讀與寫作、數學與科學（1996） |
| PISA  | 題本由閱讀、數學與科學三種不同領域組成，題本可能僅只有一個領域或是三種領域皆有涵蓋        |
| TIMSS | 題本均涵蓋數學與科學兩種領域                                   |
| TASA  | 一個題本僅一種領域  |

### 三、綜合討論與建議

測驗實施對象與施測科目而言，TASA 測驗實施目的與 NAEP 較為相似，再加上 TASA 測驗實施目的為已確定的既有目標，因此本研究並無針對 TASA 測驗實施目的進行建議之處。

而題本設計方面，NAEP 使用 BIB 或 PBIB 設計，PISA 使用之題本設計為 BIB 設計，TIMSS 考量到學生需接受數學與科學兩種領域之測驗，故題本排列採取數學與科學之試題塊分別放置在題本前、後位置順序。就施測目的與 BIB 設計之優點，也考量到 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。

### 貳、可能值方法

目前在 NAEP1998、PISA2003 和 TIMSS2007 的技術報告中，學生的成就資料是以「可能值」的資料型態提供給次級資料的分析者。在 NAEP1998、PISA2003 和 TIMSS2007 的技術報告中說明，試題反應模式中，個體的能力值是觀察不到，即個體能力的測量含有不確定性，這些不確定性在計算群體統計量和相關連的標準誤時，應被考量。可能值是每一位學生的不可觀察特質（能力）的多重差補值，



這些多重差補值能反應個體能力估計的不確定性，可能值最早是在 1986-1984 年 NAEP 的資料分析中被使用。表 4-5-9 整理列出各技術報告中所描述可能值的使用時機。

表 4-5-9 可能值的使用時機

| NAEP  | PISA   | TIMSS  |
|---|--|--|
| IRT 模式是用來測量個體的能力，當個題被施測的題數夠多時（60 題以上），個體能力測量的不確定性可以被忽略。但這種方式在大型測驗中卻不可行，原因是：受試者只被施測相對較少的試題、測驗的形式（題數、題型、試題的內容）不一樣，可能值可用來解決這些困境。 | 所有的試題反應模式中，學生的能力值是觀察不到的，它們是屬於遺失資料，需要從觀察得到的試題反應推論而得。有許多方法都可以推論能力值 PISA 是使用多重插補的方式，也就是可能值。 | 矩陣抽樣設計，學生只使用題庫內的一部分試題測量，個體能力的測量會伴隨著相當程度的測量誤差，透過個體能力估計母群的參數會有偏誤，可能值是解決此一問題之一方法。 |

### 一、可能值方法的理論

可能值是在給予學生的答題反應和相關條件變項的條件下，呈現學生可能合理的能力值範圍，不是直接估計每一位學生的能力值，而是估計一位學生能力值的機率分佈，即後驗分佈，從後驗分佈中隨機抽取學生的可能值。在 NAEP1998、TIMSS2007 和 PISA2003 的技術報告中，可能值理論推導的公式大致是相同的，只是數學符號定義和條件變數的定義不大一樣，茲以 PISA2003 的可能值理論推導為例說明。

試題反應模式為條件機率的模式，它描述了以能力值  $\theta$  為條件而產生試題反應的過程。此模式完整的定義需要界定能力值  $\theta$  的密度函數  $f_{\theta}(\theta; \alpha)$ 。令  $\alpha$  為  $\theta$  分佈的參數集。當定義單向度邊際試題反應模式 (uni-dimensional marginal item response models)，常假設抽樣的學生是來自於一個常態分佈的母體，其平均數為  $\mu$ ，變異數為  $\sigma^2$ 。也就是：

$$f_{\theta}(\theta; \alpha) \equiv f_{\theta}(\theta; \mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{(\theta - \mu)^2}{2\sigma^2}\right] \quad (4.4.1)$$

或者同義的式子，

$$\theta = \mu + E \quad (4.4.2)$$

其中， $E \sim N(0, \sigma^2)$ 。

Adams (1997) 等人使用迴歸模式  $Y_n^T \beta$  取代平均數  $\mu$ ，其中  $Y_n$  是一個  $u$  的向量，對於學生  $n$ ， $Y_n$  是條件變數，是固定且是已知， $\beta$  是一個相對應的迴歸係數向量。例如， $Y_n$  可以由性別或社經地位等學生變項所構成。則學生  $n$  的母群模式可表示為

$$\theta_n = Y_n^T \beta + E_n \quad (4.4.3)$$

其中，假設  $E_n \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ 。 $E_n$  的分佈應該會和  $\theta_n$  相同，只是將其轉換為平均數為 0，利用迴歸模式  $Y_n^T \beta$  取代平均數  $\mu$ ，其中  $Y_n$  為  $u$  的矩陣， $\beta$  為迴歸係數。例如  $Y_n$  可以被視為學生的性別、社經地位或者主修的科目，則母體的模式可以被替換為如下

$$f_n(\theta_n; Y_n, \beta, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2} (\theta_n - Y_n^T \beta)^T (\theta_n - Y_n^T \beta)\right] \quad (4.4.4)$$

為一常態分配，平均數為  $Y_n^T \beta$ ，及變異數為  $\sigma^2$ ，若使用公式 (4.4.4) 估算母體分配，則需要估算的參數為  $\beta$ ， $\sigma^2$  和  $\xi$  (試題參數)，其邊際後驗機率可以被表示為

$$h_{\theta}(\theta_n; Y_n, \xi, \beta, \sigma^2 | x_n) = \frac{f_n(x_n; \xi | \theta_n) f_{\theta}(\theta_n; Y_n, \beta, \sigma^2)}{f_x(x_n; Y_n, \xi, \beta, \sigma^2)} \quad (4.4.5)$$

如果是多維度變量母群模式，模式如下：

$$f_{\theta}(\theta_n; \gamma, \Sigma) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left[-\frac{1}{2} (\theta_n - \gamma)^T \Sigma^{-1} (\theta_n - \gamma)\right] \quad (4.4.6)$$

其中， $\gamma$  是一個  $u \times d$  的迴歸係數矩陣， $\Sigma$  是一個  $d \times d$  的變異數共變數矩陣，

每一位受試者的能力值之後驗分佈，如下所示 (Adams, Wilson & Wang, 1997)：

$$\begin{aligned}
 h_{\theta}(\theta_n; \xi, \gamma, \Sigma | x_n) &= \frac{f_n(x_n; \xi | \theta_n) f_{\theta}(\theta_n; \gamma, \Sigma)}{f_x(x_n; \xi, \gamma, \Sigma)} \\
 &= \frac{f_n(x_n; \xi | \theta_n) f_{\theta}(\theta_n; \gamma, \Sigma)}{\int_{\theta} f_n(x_n; \xi | \theta_n) f_{\theta}(\theta_n; \gamma, \Sigma)} \quad (4.4.7)
 \end{aligned}$$

表 4-5-10 可能值的理論公式

| NAEP   | PISA   | TIMSS   |
|--|--|---|
| 預期條件分佈<br>(predictive<br>conditional :<br>distribution)<br>$p(\underline{\theta}_r   \underline{x}_r, \underline{y}_r, \Gamma, \Sigma)$<br>$\propto p(\underline{x}_r   \underline{\theta}_r, \underline{y}_r, \Gamma, \Sigma)$<br>$\times p(\underline{\theta}_r   \underline{y}_r, \Gamma, \Sigma)$<br>$= p(\underline{x}_r   \underline{\theta}_r) \times p(\underline{\theta}_r   \underline{y}_r, \Gamma, \Sigma)$<br>(4.4.8)<br>$\underline{\theta}_r$ : 學生 $r$ 的能力向量<br>$\underline{x}_r$ : 學生 $r$ 的作答反應<br>向量<br>$\underline{y}_r$ : 學生 $r$ 的條件變數<br>向量<br>$\Gamma, \Sigma$ : 迴歸模式的<br>參數<br>$p(\underline{\theta}_r   \underline{y}_r, \Gamma, \Sigma)$<br>服從常態<br>分佈 | $h_{\theta}(\theta_n; Y_n, \xi, \gamma, \Sigma   x_n)$<br>$= \frac{f_n(x_n; \xi   \theta_n) f_{\theta}(\theta_n; Y_n, \gamma, \Sigma)}{f_x(x_n; Y_n, \xi, \gamma, \Sigma)}$<br>$= \frac{f_n(x_n; \xi   \theta_n) f_{\theta}(\theta_n; Y_n, \gamma, \Sigma)}{\int_{\theta} f_n(x_n; \xi   \theta_n) f_{\theta}(\theta_n; Y_n, \gamma, \Sigma)}$<br>$\theta_n$ : 學生 $n$ 的能力向量<br>$Y_n$ : 學生 $n$ 的條件變數向量<br>$x_n$ : 學生 $n$ 的作答反應向量<br>$\xi$ : 試題參數<br>$\gamma, \Sigma$ : 迴歸模式參數 | $P(\theta_j   x_j, y_j, \Gamma, \Sigma)$<br>$\propto P(x_j   \theta_j, y_j, \Gamma, \Sigma)$<br>$P(\theta_j   y_j, \Gamma, \Sigma)$<br>$= P(x_j   \theta_j)$<br>$P(\theta_j   y_j, \Gamma, \Sigma)$<br>$P(x_j   \theta_j)$ : 試題反<br>應模式<br>$P(\theta_j   y_j, \Gamma, \Sigma)$ : 在<br>背景變項 $y_j$ 、迴歸<br>參數 $\Gamma$ 和 $\Sigma$ 的條件<br>下, 學生 $j$ 的能力值<br>的多變量聯合密度<br>函數。在計算的過<br>程中, 試題參數是<br>固定的並且被視為<br>是母群的值。 |

在 PISA 中，使用公式 4.4.7 產生每一位學生的可能值。表 4-5-10 整理列出各技術報告中可能值的理論公式。

大型測驗中，使用背景問卷的資料作為母群回歸模式中的條件變數，由於背景問卷資料變數很多，故在大型測驗中常使用主成分分析的方式，縮減問卷變項再納入條件變數，PISA2003 中指出，主成分分析的變數需能解釋原始資料 95% 的變異，而 NAEP1998 和 TIMSS2007 則是要能解釋原始資料 90% 的變異。在 PISA2003 和 TIMSS2007 中設定主要條件變數，主要條件變數是指沒有透過主成分分析，一定會被納入納入迴歸模式中的條件變數，如性別。表 4-5-11 整理列出各技術報告中條件變數的定義。

表 4-5-11 條件變數的設定

| NAEP  | PISA  | TIMSS   |
|---|---|---|
| <p>1.從背景變數中，取出約 200 個主成分，包含背景變項的主要效果和交互作用效果，能解釋約 90% 的變異。</p> <p>2.對於長期追蹤研究，直接使用個數較少的背景變項的主要效果和交互作用，而不使用主成分分析</p> | <p>1.五個變數（題本 ID(booklet ID)、性別、母親的職業、父親的職業和學校的數學平均分數）直接視為是主要條件變數。</p> <p>2.將學生問卷中的變數虛擬編碼（dummy coded）。</p> <p>3.對於每一個國家，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 95% 的變異。</p> | <p>1.性別（虛擬編碼）、試卷的語言（虛擬編碼）、學生所隸屬的學校班級（criterion-scales）、特定選擇的國家變數（虛擬編碼）是主要條件變數</p> <p>2.類別變數，使用虛擬編碼。</p> <p>3.連續變數（出生年、家裡人口）使用 criterion scaling，計算 interim achievement score，計算的方式請參閱 Beaton（1969）。</p> <p>4.使用 PCA 分析虛擬編碼和 criterion-scales 變數，取可解釋 90% 變異的成分。</p> |

## 二、可能值的抽取步驟

可能值是從學生的能力後驗分佈中隨機抽取，目前在 NAEP1998、TIMSS2007 和 PISA2003 的技術報告中每一位學生抽取 5 個可能值，以 PISA 為例，說明可能值抽取的步驟。NAEP1998 和 TIMSS2007 的可能值抽取步驟列於表 4-5-11。

NAEP1998 和 TIMSS2007 是使用 EM 演算法則估計後驗分佈，PISA2003 是以對於每一個國家的資料集是合適的試題反應模式且使用國際間校正的定錨試題的參數和經由主成分分析得到的條件變數估計，使用蒙地卡羅積分法，對於每一位學生，從能力值的邊際後驗機率 (4.4.7) 隨機抽取可能值，表 4-5-12 整理列出各技術報告中可能值之抽取步驟。

表 4-5-12 可能值之抽取步驟

| NAEP   | PISA  | TIMSS   |
|--|---|---|
| 使用 EM 演算法則<br>估計得到 $\hat{\Gamma}, \hat{\Sigma}$<br>步驟一：<br>從平均數是 $\hat{\Gamma}$ ，<br>變異數是 $\hat{\Sigma}$ 的分佈中<br>抽取一個 $\Gamma$ 。<br>步驟二：<br>將抽取到的 $\Gamma$<br>和固定的 $\hat{\Sigma}$ 帶入<br>公式 (4.4.8)，使用<br>EM 演算法<br>可以計算受試者 r<br>的預期條件分佈<br>$\bar{\theta}_r$ 和 $\Sigma_r$<br>步驟三：<br>從近似多變量<br>常態分佈 | 步驟一：<br>對於每一個受試者 n，<br>從多變量常態分佈，<br>$f_{\theta}(\theta_n; Y_n, \gamma, \Sigma)$<br>產生 M vector-valued<br>random deviates,<br>$\{\varphi_{mn}\}_{m=1}^M$ 。<br>步驟二：<br>使用蒙地卡羅積分法逼<br>近式子 (4.4.7) 的分母<br>$\int_{\theta} f_x(x; \xi   \theta) f_{\theta}(x, \gamma, \Sigma) d\theta$<br>$\approx \frac{1}{M} \sum_{m=1}^M f_x(x; \xi   \varphi_{mn}) \equiv \mathfrak{S}$<br>同時，計算<br>$P_{mn} = f_x(x_n; \xi   \varphi_{mn})$<br>$f_{\theta}(\varphi_{mn}; W_n, \gamma, \Sigma)$<br>$\{\varphi_{mn}, P_{mn} / \mathfrak{S}\}_{m=1}^M$ | 步驟一：<br>從一個近似常態的分配<br>$P(\Gamma, \Sigma   x_j, y_j)$ ，<br>固定 $\Sigma$ 為 $\hat{\Sigma}$ ，<br>抽取一個 $\Gamma$ 。<br>步驟二：<br>在 $\Gamma$ 的條件下，<br>(且固定 $\Sigma = \hat{\Sigma}$ )<br>學生 j 後驗分佈<br>的平均 $\theta_j$ 和變異數 $\sum_j^p$<br>使用 EM 的演算法則計<br>算。<br>步驟三：<br>能力值從一個<br>多變量常態分佈<br>(平均 $\theta_j$ 、變異數 $\sum_j^p$ )<br>獨立抽取。<br>這三個步驟重複五次 |

|  |  |  |
|--|--|--|
| <p>(<math>\bar{\theta}_r</math> 和 <math>\Sigma_r</math>)</p> <p>的分佈抽取 <math>\theta_r</math></p> <p>此三個步驟重複五次，產生五個集合的能力差補值。</p> | <p>的集合可視為式子 (4.4.7) 的後驗機率函數之近似。</p> <p>步驟三：</p> <p>機率值 <math>\varphi_{nj}</math> 可藉由以下公式求得：</p> $q_{nj} = \frac{P_{mn}}{\sum_{m=1}^M P_{mn}}$ <p>步驟四：</p> <p>隨機產生 L 個服從均勻分佈的值 <math>\{\eta_i\}_{i=1}^L</math>；對於每一次隨機抽取，若 <math>\varphi_{ni_0}</math> 滿足下列條件則選取當作可能值：</p> $\sum_{s=1}^{i_0-1} q_{sn} < \eta_i < \sum_{s=1}^{i_0} q_{sn}$ | <p>每一位學生產生 5 個 <math>\theta_j</math> 的差補值。</p> |
|--|--|--|

### 三、計算可能值之軟體

TIMSS2007 使用 ETS 的 MGROUP 的軟體產生 IRT 的能力值，輸入的值是學生的作答反應組型、試題參數和條件變數，輸出值是可能值。PISA2003 是使用 ACER 所發展的 ConQuest 軟體。NAEP1998 是使用 EST 所發展的 MGROUP、CGROUP、BGROUPM 和 GROUP 軟體。MGROUP 適用一主題內測量 p 個量尺，CGROUP: 是不同主題測量多個量尺，BGROUP 只能測量單一量尺，計算較準確，且因為在 E 步驟使用 numeric quadrature，只能應用於一維或二維向度的能力。如果量尺只有單一向度，MGROUP, CGROUP, BGROUP 準確度差不多；如果量尺是多向度，CGROUP 和 BGROUP 對於能力間的相關估計較準確，但 BGROUP 只能用於雙變數。表 4-5-13 列出三大測驗所使用之可能值分析軟體。

表 4-5-13 可能值分析軟體

| NAEP   | PISA     | TIMSS  |
|--------|----------|--------|
| MGROUP | ConQuest | MGROUP |
| CGROUP |          |        |
| BGROUP |          |        |

#### 四、使用可能值的資料分析

假如所有抽樣的學生  $\theta$  是知道的，則可以計算統計量  $t(\theta, y)$ ，如樣本平均數，而後推論相對應的母群參數  $T$ ，可惜的是， $\theta$  是未知的。大型測驗中，將  $\theta$  視為遺失資料並且用條件期望值近似  $t(\theta, y)$ 。

$$\begin{aligned}
 t^*(x, y) &= E[t(\underline{\theta}, \underline{y}) | \underline{x}, \underline{y}] \\
 &= \int t(\underline{\theta}, \underline{y}) p(\underline{\theta} | \underline{x}, \underline{y}) d\underline{\theta}
 \end{aligned}
 \tag{4.4.9}$$

其中  $(\underline{\theta}, \underline{y}) = (\theta_1, y_1, \theta_2, y_2, \theta_3, y_3, \dots, \theta_J, y_J)$ ， $(\theta_j, y_j)$  是學生  $j$  的能力向量和條件變數。給予學生  $j$  的答題反應  $x_j$ ，學生背景變數  $y_j$ ，試題參數，從能力值的條件分布中隨機抽樣（可能值）可以近似  $t^*$ ，計算  $t$  的  $\theta$  值是從學生的條件分布中重複隨機抽取，Rubin (1987) 指出這種重複的歷程可以將插補的不確定性量化，如透過不同的可能值集合，可以計算不同的  $t$ ，這些  $t$  的平均，就是  $t^*$  的數值近似，他們所呈現的變異，反應無法直接觀察  $\theta$  的不確定性。需注意的是，這種變異並未包含抽樣的變異，抽樣的變異在 TIMSS2007 和 NAEP1998 中是藉由 jackknife variance estimation procedure 估計而得，在 PISA2003 是使用 Fay's variant of the Balanced Repeated Replication 估計。

可能值並非估計學生的個別分數，而是對相似的學生（學生有相似的答題反應和背景變項）插補分數，這樣估計群體時會較準確。當模式被正確界定时，可能值可以提供群體參數的一致性估計，但他們並非個體能力的不偏估計，使用可能值的平均並不能代表個別學生的能力（Mislevey, Beaton, Kaplan, & Sheehan, 1992）。

可能值可以被用來計算公式 (4.4.9) 進而得到  $T$ ，計算方式如下：

步驟一 使用每一位學生的第一組可能值向量計算  $T$ ，就像可能值是  $\theta$  的真值，將結果記為  $T_1$

步驟二 計算  $T_1$  的抽樣變異，也就是第一組可能值的抽樣變異  $Var_1$

步驟三 重複步驟一、步驟二分別計算第二組～第五組的可能值得到  $T_u$  和  $Var_u$ ， $u=2, \dots, 5$ 。

步驟四  $T$  的最佳估計值是從不同集合的可能值平均而得。

$$\hat{T} = \frac{\sum_u T_u}{5}$$

$\hat{T}$  的變異數估計包含兩個成分：

$$\bar{U} = \frac{\sum_u Var_u}{M}$$

$$B_M = \frac{\sum_u (T_u - \hat{T})^2}{M - 1}$$

$\hat{T}$  的總變異

$$Var(\hat{T}) = \bar{U} + (1 + M^{-1})B_M$$

在  $Var(\hat{T})$  的第一個變異成分源自於從母群抽樣學生的不確定性(抽樣變異)，第二個成分是抽樣學生的  $\theta$  無法準確知道，只能透過  $x$  和  $y$  間接而得，是屬於測量變異。表 4-5-14 是三大測驗所使用之標準誤計算公式。



表 4-5-14 標準誤計算公式

| NAEP  | PISA   | TIMSS  |
|---|--|--|
| $V = U^* + (1 + M^{-1})B$ <p>where</p> $U^* = \sum_{m=1}^M \frac{U_m}{M}$ $B = \sum_{m=1}^M \frac{(\hat{t}_m - t^*)^2}{(M-1)}$ <p>抽樣變異：Jackknife estimate</p> | $V = U^* + (1 + M^{-1})B_M$ <p>where</p> $U^* = \frac{1}{M} \sum_{m=1}^M U_m$ $B_M = \frac{1}{M-1} \sum_{m=1}^M (\hat{r}_m - r^*)^2$ <p>抽樣變異：Fay's variant of the Balanced Repeated Replication</p> $\sigma_{(\hat{\theta})}^2 = \frac{1}{G(1-K)^2} \sum_{i=1}^G (\hat{\theta}_{(i)} - \hat{\theta})^2$ $\sigma_{(\hat{\theta})}^2 = \frac{1}{80(1-0.5)^2} \sum_{i=1}^{80} (\hat{\theta}_{(i)} - \hat{\theta})^2$ $= \frac{1}{20} \sum_{i=1}^{80} (\hat{\theta}_{(i)} - \hat{\theta})^2$ | $Var(\hat{T}) = \bar{U} + (1 + M^{-1})B_M$ $\bar{U} = \frac{\sum_u Var_u}{M}$ $B_M = \frac{\sum_u (T_u - \hat{T})^2}{M-1}$ <p>抽樣變異：Jackknife variance estimation procedure</p> |

### 五、建立共同量尺

量尺化程序包含同一施測年度中各題本間的量尺化過程，以及不同施測年度間的量尺化過程，即將該年度之施測資料拉到前一年度的量尺化過程，NAEP、PISA、TIMSS 是使用以試題反應理論模式為基礎再加入背景變項之迴歸模式所估計出的「可能值」建立共同量尺，TASA 仍是使用以試題反應理論模式為基礎估計而得的「能力值」建立共同量尺。以下將介紹此兩種過程。

## (一) 同年度題本間量尺化程序

NAEP1998 年閱讀與寫作之題本設計使用 BIB 設計，公民使用 PBIB 設計，PISA2006 年使用 BIB 設計。BIB 與 PBIB 設計的特點在於試題區塊在不同題本間出現次數相同，故不同題本間可以透過相同之試題區塊來作連結。NAEP1998 年技術報告中提到，不同科目間之參數估計採取各別估計，且採用同時估計法將該科所有試題連結到同一量尺上。而 PISA 在技術報告中找不到同年度題本間量尺化相關文獻。

TIMSS2007 年之各題本包含 4 個試題區塊，為了連結各題本，每個試題區塊在題本中出現次數為 2 次。在 TIMSS2007 年技術報告中提到，題本間之連結數學與科學皆採取同時估計法來將試題連結到同一量尺上。

TASA 除英語文外，其餘科目在 2005 年、2006 年、2009 年皆使用 BIB 或 PBIB 設計，2007 年使用 NEAT 設計。雖然題本設計在不同年度間有所改變，但是在同年度題本間量尺化過程皆採取同時估計法將所有試題拉到同一量尺上。

## (二) 不同年度間量尺化程序

### 1. NAEP

NAEP 在不同年度間量尺化程序是經由共同校準連結到之前的測驗，藉由共同群體 (common population) 與線性轉換 (linear transformation) 來將兩次測驗連結在一起 (NAEP Technical Documentation, 2009)。連結步驟如下：

- (1) 得到前一次測驗之報告測度 (reporting metric)；
- (2) 將兩次測驗資料結合在一起作同時估計以得到一個臨時性量尺 (provisional scale)；
- (3) 由臨時性量尺中挑選出前一次測驗之 PVs；
- (4) 透過線性轉換法將前一次測驗之臨時性量尺轉換至前一次測驗之報告測度上；
- (5) 使用相同之線性轉換法將該測驗之臨時性量尺轉換成報告測度。

連結之後之量尺分數平均數接近 0，再轉換成平均數 150、標準差 35 或是平均數 250、標準差 50 之報告測度。

線性轉換公式如下：

$$Z_{reporting} = Z_{provisional}$$

$$\frac{PV_r - \mu_r}{\sigma_r} = \frac{PV_p - \mu_p}{\sigma_p}$$

$$PV_r - \mu_r = \frac{\sigma_r}{\sigma_p} PV_p - \frac{\sigma_r}{\sigma_p} \mu_p$$

$$PV_r = \frac{\sigma_r}{\sigma_p} PV_p + \mu_r - \frac{\sigma_r}{\sigma_p} \mu_p$$

令  $A = \frac{\sigma_r}{\sigma_p}$ ,  $B = \mu_r - A\mu_p$

$$PV_r = A PV_p + B$$

## 2. PISA (TIMSS 在不同年度間量尺化方法與 PISA 相同)

連結 PISA 2000 年與 PISA 2003 年閱讀與科學之步驟 (OECD, 2009)：

- (1) 將 PISA 2000 年 OECD 會員國施測資料在固定 2003 年之定錨試題參數後重新估計；
- (2) 將 25 個會員國資料結合在一起來計算每個領域之平均數與標準差，此時各國的權重是相同的；
- (3) 將步驟 2 之平均數與標準差與 PISA 2000 年報告測度之平均數與標準差作比較，並且將 PISA 2003 年之量尺分數透過線性轉換到 PISA 2000 年。

## 3. TASA

TASA 在不同年度間量尺化程序是採取固定試題參數法，並透過同時估計法將所有試題一起估計得到新的年段之量尺分數。步驟如下：

- (1) 取得該年度與前一次測驗之定錨試題試題參數；
- (2) 藉由固定前一次測驗估計而得之定錨試題參數來估計該年度測驗；
- (3) 由步驟 2 獲得連結之後之參數值轉換成平均數 250、標準差 50 之量尺分數。

TASA 在不同年度間量尺化程序，如果以實證資料 2006 小四數學、2007 小四數學為例來說明，其流程如下圖 4-5-1：

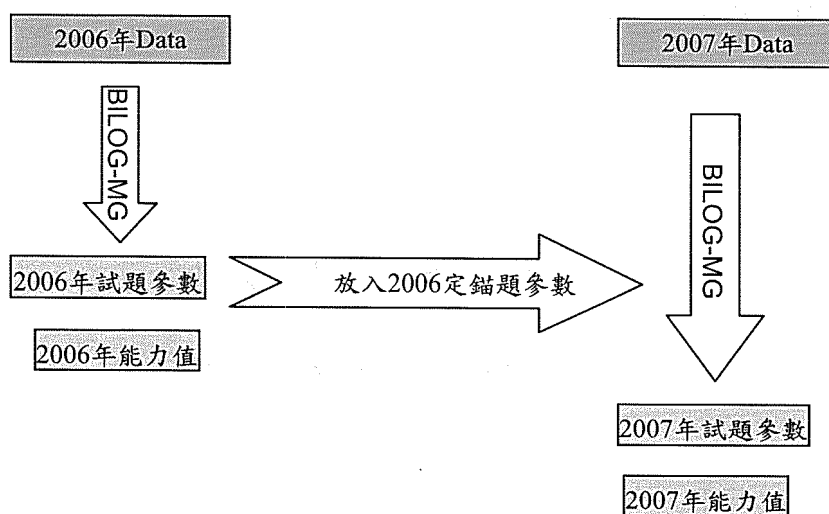


圖 4-5-1 TASA 不同年度間量尺化過程

各大型測驗所使用的量尺分數範圍不大一樣，整理如表 4-5-15

表 4-5-15 量尺分數範圍之綜合比較

|       |  |
|-------|--|
| NAEP  | 數學：平均數 250，標準差 50<br>科學、閱讀、寫作：平均數 150，標準差 35 |
| PISA  | 平均數 500，標準差 100                              |
| TIMSS | 平均數 500，標準差 100                              |
| TASA  | 平均數 250，標準差 50                               |

## 六、綜合討論與建議

在建立量尺已經探討了NAEP、PISA、TIMSS與TASA之同年度間量尺化程序及不同年度間量尺化程序，目前TASA的資料分析並未使用可能值方法建立量尺，透過上述之文獻探討，建議可能值分析流程與建立共同量尺的過程。

## (一) 可能值方法

TASA的施測採BIB等化設計，受試者只需接受若干試題區塊的試題，且不同受試者可能接受部分相同、完全相同、或完全不同的試題區塊。其優點在於能進行大量的施測試題（如可高達195題），可含括較廣的內容領域，但又不造成受試者之精神負荷（因為每位受試者僅接受約45題），除學科能力之評量外，納入學校、學生背景問卷，能探討如社經地位高低、子女數（如探討少子化影響）、單親家庭、動機、情緒等心理狀態對學業成就之影響等，測驗的資料可對個人或團體進行追蹤比較（臺灣學生學習成就評量資料庫電子報，2009）。由於學生只接受某些試題區塊測量，個體能力的測量會伴隨著相當程度的測量誤差，且TASA也有背景變項之問卷，建議使用可能值方法計算可能值，除了能得到群體統計量的良好估計值外，藉由可能值資料的釋出，能提供給次級資料分析者進行學生學習成效相關因素之探討，促進相關教育議題的討論與連結。

### 1. 可能值方法的理論

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，理論的模式可參閱PISA2003的技術報告，背景變項的建立方式為參考NAEP（Beaton, 1987）、TIMSS（Macaskill, Adams & Wu, 1998）和PISA（OECD, 2005）。建立步驟如下所示：

步驟一：將性別、社經背景直接視為是條件變數。

步驟二：將學生問卷中的變數虛擬編碼。

步驟三：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。

步驟四：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。

步驟五：使用上述的方法抽取五個可能值向量。

### 2. 可能值得抽取步驟

可能值抽取方式如文獻中說明。

### 3. 產生可能值的軟體

由於ConQuest軟體，具有視窗化界面且有詳細的操作手冊可供參考，建議以ConQuest軟體作為分析軟體。

### 4. 可能值的資料分析

標準誤的計算須包含測量變異和抽樣變異。透過五個可能值可計算測量變異，在抽樣變異建議以Jackknife variance estimation procedure為主。

### 5. 建立共同量尺

同年度間量尺化程序方面，由於目前僅探討了同年度題本間量尺化程序以及NAEP在不同年度之量尺化程序、PISA閱讀與科學連結2000年與2003年之量尺化程序。研究發現NAEP與TIMSS在同年度題本間量尺化程序皆採取同時估計法，此方法與目前TASA之方法一致。因此建議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。

在不同年度間量尺化程序方面，由於NAEP是將兩年度資料放在一起進行同時估計，再經由線性轉換進行等化，如此考量到TASA為每年施測，如果採行此方法進行不同年度間量尺化，會造成資料量過大，較不適用於TASA上，而PISA、TIMSS使用分開估計法比較適用於TASA資料上，因此建議TASA使用PISA在不同年度間量尺化方法進行等化，其中以2006、2007年小四數學為例，如下圖4-5-2。

- (1) 將TASA2006年與2007年小四數學分開估計，取得各年度學生能力PV值、試題參數。
- (2) 將2007年定錨題試題參數放入2006年資料中，將2006年資料重新估計出新的2006年學生能力PV值。
- (3) 將2007年學生能力PV值利用線性轉換方法轉換成新的2007年學生能力PV值。

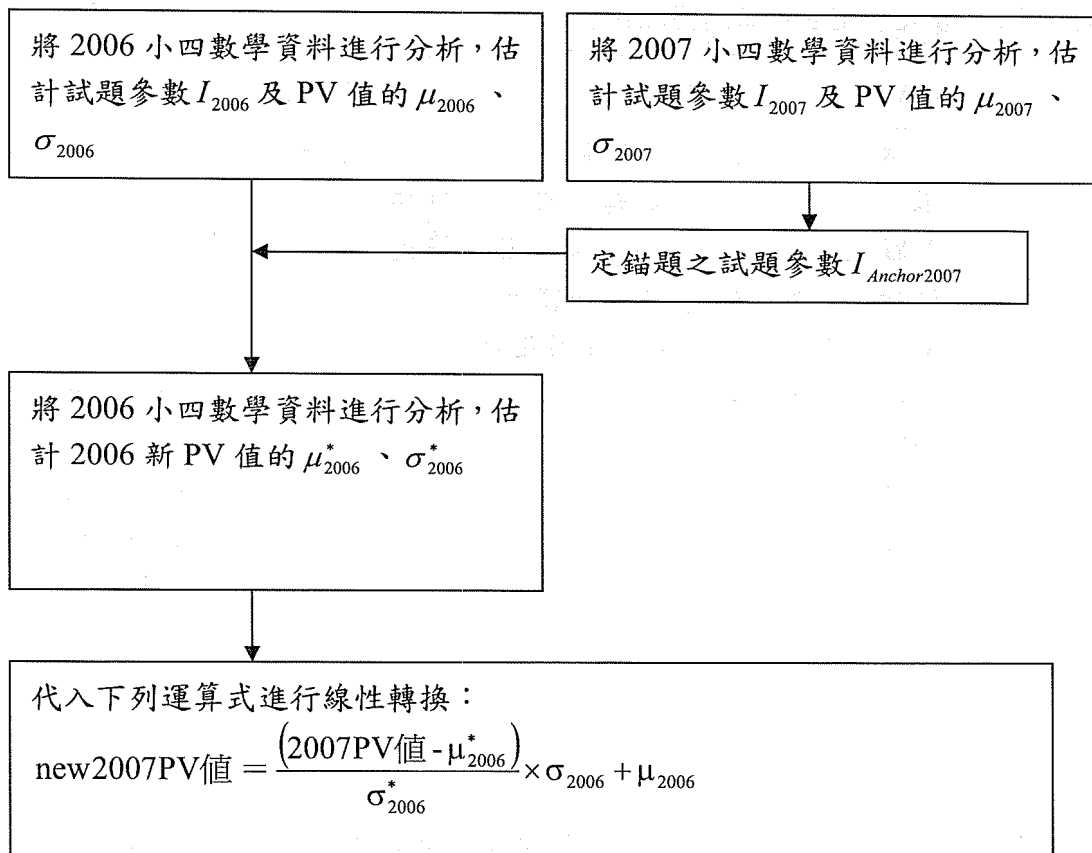


圖 4-5-2 TASA 實徵資料中不同年度間量尺化之過程

### 參、建立試題圖 (item map)

試題圖主要是顯示各試題在量尺上的分佈，讓讀者可以清楚的看到各題在難度上的排序與差異，因此本研究針對 NAEP、PISA、TIMSS、TASA 探討是否有建置試題圖 (item map)，並探討試題圖中所含有的訊息。表 4-5-16 為各大型測驗試題圖之比較。以下針對 NAEP、PISA、TASA 試題圖內容進行描述如下：

表 4-5-16 各大型測驗中試題圖之比較

|     | NAEP  | PISA   | TIMSS | TASA   |
|-----|---|--|-------|--|
| 試題圖 | 1. 結合量尺分數<br>2. 將量尺分數區分為進階、精熟、基礎<br>3. 分為 MC 與 CR 試題<br>4. 針對題目進行描述 | 1. 結合量尺分數<br>2. 如為部份給分試題則將各難度類別皆列出在試題圖中<br>3. 針對題目進行描述 | 無此資料  | 1. 結合量尺分數<br>2. 將量尺分數區分為進階、精熟、基礎<br>3. 皆為 MC 試題（因計算量尺分數時並無納入 CR 試題）<br>4. 針對題目進行描述 |

#### 一、NAEP

試題圖主要是藉由量尺上各試題的位置來說明每一年級量尺學生在各科目領域的認知與完成程度；而試題在量尺上的位置代表學生可能答對該題。

NAEP 試題圖主要在重點在答對試題所需的知識與技能 (knowledge and skills)。以 MC 試題來說，即為答對試題的機率；CR 試題則為試題的分數層級。NAEP 所使用之 4 選項 MC 試題為 74% 的答對機率，5 選項 MC 試題為 72% 的答對機率，CR 試題則為 65% 獲得分數層級的機率。由於 CR 試題有不同的分數層級，所以 CR 試題在 item map 上不只有一個位置 (NAEP Technical Documentation, 2009)。

圖 4-5-3 為 NAEP2009 四年級數學的試題圖，其中內容上分類為數的特性與運算、測量、幾何、資料分析統計與機率、代數五個內容。在這個試題圖中，有標示出進階層級的量尺分數 282、精熟層級的量尺分數 249、基礎層級的量尺分數 214，並且標示出各個技能相對應試題的類型是屬於 MC 試題或者是 CR 試題。而在 MC 試題中的「在圖表中找到中位數」這個技能需要量尺分數 300 才有較高的機率可以答對，而這個技能是屬於在進階的量尺分數以上。在每個 CR 試題中



皆會有對於此題的分數層級描述，如「正確」、「部分正確」、「不正確」這些層級上的描述，如下圖 4-5-3 中，在「格子上畫點以滿足給定的條件」這個技能有四種分數層級，在量尺分數為 277 時分數層級為具延伸性的 (Extended)、量尺分數為 263 時分數層級為滿意的 (Satisfactory)、量尺分數為 260 時分數層級為部份的 (Partial)、量尺分數為 230 時分數層級為最小的 (Minimal)。

## 二、PISA

從 PISA 科學評量中的資料去估計出試題難度，其中某些試題為部份給分試題，而在建立試題圖的過程必須將部份給分試題的各個難度類別視為不同的難度，以 PISA 科學評量為例，題目共有 103 題，但難度類別有 109 個，因為其中有部份給分的試題。(OECD, 2009)

在試題發展的過程中，專家學者試圖對每個試題進行質性分析並且針對每個試題所需具備的認知需求觀點進行描述，每個題目被決定需要具備哪些能力與知識類型才能夠正確回答。其中也定義了各題對於科學的相關是個人的、社會的或全球性的。(OECD, 2009)

圖 4-5-4 是 PISA2006 中部份科學試題的試題圖，其中每一列為不同試題，而在這些被選出的部份試題依據了試題難度在試題圖上被排序，其中難度最難的被排序在試題圖的最上面，難度最低的被排序在最下面，而在試題圖中提供了相對應的量尺分數以及試題需求描述。

| Content Classifications:           |               |            |  |           |
|------------------------------------|---------------|------------|--|-----------|
| ◆ Number Properties and Operations | ■ Measurement | ▲ Geometry | ▼ Data Analysis, Statistics, and Probability | ◆ Algebra |

- 500  
↓
- 310  
▼ 300 Find the median price from a table (MC)
- 300  
◆ 299 Identify the expression that models a scenario (MC)  
▲ 295 Identify parallel and perpendicular lines (MC)  
◆ 291 Solve a story problem involving remainders (MC)
- 290  
■ 288 Indicate measurements on a ruler—Correct (CR)  
◆ 288 Identify the fraction closest to the given value (MC)  
◆ 285 Reason using equivalences to make and explain a conclusion (calculator available)—Satisfactory (CR)  
**287 Advanced**  
◆ 281 Identify a pictorial representation of equivalent fractions (MC)
- 280  
▲ 277 Plot points on a grid to satisfy the given conditions—Extended (CR)  
◆ 276 Reason using equivalences to make and explain a conclusion (calculator available)—Partial (CR)  
◆ 273 Reason about odd and even numbers—Correct (CR)  
▼ 270 Read and interpret a line-graph line graph (MC)
- 270  
◆ 267 Reason using equivalences to make and explain a conclusion (calculator available)—Minimal (CR)  
◆ 265 Divide a three-digit number by a one-digit number (MC)  
▲ 263 Plot points on a grid to satisfy the given conditions—Satisfactory (CR)  
▲ 260 Plot points on a grid to satisfy the given conditions—Partial (CR)
- 260  
■ 257 Identify the figure with the greatest area on a grid (MC)  
▲ 252 Identify the shape of a shaded region (MC)  
▼ 250 Determine the probability of a particular event (MC)
- 250  
**249 Proficient**  
■ 246 Solve a story problem involving quarts and cups (MC)  
◆ 243 Subtract a two-digit number from a three-digit number (MC)  
◆ 241 Determine the missing shapes in a pattern (MC)  
■ 240 Indicate measurements on a ruler—Partial (CR)
- 240  
◆ 237 Determine a ratio from a diagram (MC)  
◆ 233 Determine the value of an unknown in a number sentence—Correct (CR)  
▲ 230 Plot points on a grid to satisfy the given conditions—Minimal (CR)  
◆ 230 Use place value to write a number—Correct (CR)
- 230  
▲ 228 Determine how many given pieces cover a shape (MC)  
◆ 222 Represent the same whole number in different ways—Correct (CR)  
▼ 222 Make a pictograph of the given information—Correct (CR)
- 220  
**221 Basic**
- 210  
◆ 207 Recognize the result of multiplying by 10 (MC)  
◆ 205 Compute the product of a 2-digit number and a 1-digit number (MC)  
■ 202 Identify an appropriate unit for measuring length (calculator available) (MC)
- 200  
◆ 199 Find the unknown in a whole number sentence (MC)
- 190  
◆ 189 Represent the same whole number in different ways—Partial (CR)  
◆ 188 Compute a value using multiplication and division (calculator available) (MC)  
▲ 183 Identify the figure that is not symmetric (calculator available) (MC)
- 180  
■ 176 Identify the appropriate measuring device (MC)
- 170  
↓
- 0

圖 4-5-3 2009 年 NAEP 數學試題圖

資料來源：NAEP Item Map: Mathematics, Grade 4, 2009

( <http://nces.ed.gov/nationsreportcard/itemmaps/> )

| Code       | Item name                  | Item difficulty on PISA scale | Item demands   | Competency                    |                                     | Knowledge        |                |                         |                    |                    |                         | Focus    |        |        |
|------------|----------------------------|-------------------------------|--|-------------------------------|-------------------------------------|------------------|----------------|-------------------------|--------------------|--------------------|-------------------------|----------|--------|--------|
|            |                            |                               |  | Identifying scientific issues | Explaining phenomena scientifically | of               |                |                         | about              |                    |                         | Personal | Social | Global |
|            |                            |                               |  |                               |                                     | Physical systems | Living systems | Earth and space systems | Technology systems | Scientific enquiry | Scientific explanations |          |        |        |
| S485Q05(2) | ACID RAIN                  | 717                           | The reason for a control in an investigation is understood and explicitly recognised. An ability to understand the modelling in the investigation is a pre-requisite.                        | •                             |                                     |                  |                |                         |                    |                    | •                       | •        |        |        |
| S114Q05    | GREENHOUSE                 | 709                           | There is a pre-requisite to understand the need to control variables. Knowledge of factors contributing to the greenhouse effect is then applied in determining a variable to be controlled. |                               | •                                   |                  |                | •                       |                    |                    |                         |          | •      |        |
| S114Q04(2) | GREENHOUSE                 | 659                           | Given a conclusion can compare two graphs and locate corresponding areas that are at odds with that conclusion and accurately describe that difference.                                      |                               |                                     | •                |                |                         |                    |                    | •                       |          | •      |        |
| S447Q05    | SUNSCREENS                 | 616                           | Correctly interprets a dataset expressed diagrammatically and provides an explanation that summarises the data.  |                               |                                     | •                |                |                         |                    |                    | •                       | •        |        |        |
| S447Q02    | SUNSCREENS                 | 588                           | The control 'aspects' of an investigation are recognised.  | •                             |                                     |                  |                |                         |                    |                    | •                       | •        |        |        |
| S193Q05    | PHYSICAL EXERCISE          | 583                           | Recognition that increased exercise results in increased respiration and thus the need for more oxygen and/or removal of more carbon dioxide.  |                               | •                                   |                  | •              |                         |                    |                    |                         | •        |        |        |
| S114Q04(1) | GREENHOUSE                 | 568                           | Recognises differences in two graphs relating to a phenomenon but cannot provide a clear explanation as to why the differences are at odds with a given conclusion.                          |                               |                                     | •                |                |                         |                    |                    | •                       |          | •      |        |
| S213Q01    | CLOTHES                    | 567                           | Can apply knowledge of the features of a scientific investigation to decisions about whether specific issues are scientifically investigatable.  | •                             |                                     |                  |                |                         |                    |                    | •                       | •        |        |        |
| S493Q01    | PHYSICAL EXERCISE          | 545                           | Can identify some features of physical exercise that are advantageous to health – cardiovascular system, bodyweight.   |                               | •                                   |                  | •              |                         |                    |                    |                         | •        |        |        |
| S114Q03    | GREENHOUSE                 | 529                           | Shows an understanding of what two graphs relating to a phenomenon are depicting and can compare them for similarities.  |                               |                                     | •                |                |                         |                    |                    | •                       |          | •      |        |
| S485Q05(1) | ACID RAIN                  | 513                           | Recognises that a comparison is being made between two tests but is unable to articulate the purpose of the control.   | •                             |                                     |                  |                |                         |                    |                    | •                       | •        |        |        |
| S477Q04    | MARY MONTAGU               | 507                           | Recognises that the immune systems of young and old people are less resistant to viruses than those of the general population.   |                               | •                                   |                  | •              |                         |                    |                    |                         | •        |        |        |
| S447Q03    | SUNSCREENS                 | 499                           | Can recognise the change and measured variables from a description of an investigation and as a consequence identify the question motivating the investigation.                              | •                             |                                     |                  |                |                         |                    |                    | •                       | •        |        |        |
| S426Q07    | GRAND CANYON               | 485                           | Can recognise issues in which scientific measurement can be applied to answering a question.   | •                             |                                     |                  |                |                         |                    |                    | •                       | •        |        |        |
| S485Q03    | ACID RAIN                  | 460                           | Recognises that the loss of gas in a chemical reaction results in a reduction of mass for the products left behind.  |                               |                                     | •                | •              |                         |                    |                    |                         | •        |        |        |
| S426Q03    | GRAND CANYON               | 451                           | Applies knowledge that water increases in volume as it changes from liquid to solid.   |                               | •                                   |                  |                | •                       |                    |                    |                         | •        |        |        |
| S477Q03    | MARY MONTAGU               | 431                           | Recalls knowledge of the role of antibodies in immunity.   |                               | •                                   |                  | •              |                         |                    |                    |                         | •        |        |        |
| S508Q03    | GENETICALLY MODIFIED CROPS | 421                           | Understands that a fair test involves finding out if an outcome is affected by a range of extraneous conditions.   | •                             |                                     |                  |                |                         |                    |                    | •                       | •        |        |        |
| S213Q02    | CLOTHES                    | 399                           | Can select the correct apparatus to measure an electric current.   |                               | •                                   |                  |                | •                       |                    |                    |                         | •        |        |        |
| S493Q03    | PHYSICAL EXERCISE          | 386                           | Rejects the notion that fats are formed in the muscles and knows that the rate of flow of blood increases during exercise.   | •                             |                                     |                  | •              |                         |                    |                    |                         | •        |        |        |

圖 4-5-4 PISA2006 中部份科學試題的試題圖

資料來源：PISA2006 技術報告（頁 291）

### 三、TASA

在 TASA2006 年小四數學成果報告中，將所公佈的樣本測驗包括 24 題選擇題。為了具體描述各水準學生數學表現概述，以下將依上述試題難度分類規則及通過標準，輔以公佈試題內容，說明各成就水準四年級學生的實際表現概況。同時呈現這些公佈試題的試題圖，如表 4-5-17。試題圖應用的功能是當學生面對其能力分數鄰近的試題時，答對率在選擇題約為 74%，應用題約為 65%。(2006 年國小四年級學生『數學』學習成就之研究調查期末報告，2006)

### 四、綜合討論與建議

在 NAEP、PISA、TASA 試題圖中，皆以量尺分數為基準，針對測驗試題進行難度上的排序並將學生相對應的量尺分數對應出來在試題圖上，並且在相對應的題目詳細敘述該题目的知識內容，在 NAEP 與 TASA 中更在試題圖上將量尺分數分為進階、精熟、基礎，讓讀者可以清楚的了解試題是屬於在何種層級，NAEP 與 TASA 試題圖中將試題清楚註明是 MC 試題或 CR 試題。在 PISA 的試題圖中，針對各題進行質性的探究，歸納出各題所對應的知識層面。建議 TASA 可以參考 2006 年試題圖並納入部分給分模式的 CR 試題到試題圖中，並且可以加入 PISA 在製作試題圖時，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。

表 4-5-17 TASA2006 小四數學之試題圖

| Grade 4                  |             | 2006 TASA—Mathematics Scale  |                                      |
|--------------------------|-------------|--|--------------------------------------|
| 題型代碼對照：                  | Score Scale | 試題說明   | 題型                                   |
| MC=Mutiple Choice 選擇題    | 500         |  |                                      |
|                          | 410         | 414 估計日常容器（如冰箱）的容積   | (MC)                                 |
|                          | 400         |  |                                      |
|                          | 390         |  |                                      |
|                          | 370         |  |                                      |
|                          | 360         | 365 運用長條圖的多項資訊進行估算<br>363 比較小數、帶分數、和假分數的大小   | (MC)<br>(MC)                         |
|                          | 350         | 356 在不同區段的計程車計費方式下，由所付費用估計里程   | (MC)                                 |
| <b>345</b> 進階 Advanced   | 340         | 344 比較三位小數的位值<br>341 在長方形的面積情境中，找出特定數量範圍中 15 的倍數   | (MC)<br>(MC)                         |
|                          | 330         |  |                                      |
|                          | 320         |  |                                      |
|                          | 310         | 313 依據已知長、寬、高的長方體容器，計算可含納小長方體的最大數量   | (MC)                                 |
|                          | 300         |  |                                      |
|                          | 290         | 299 比較不同圖形的周長<br>299 判斷二位小數位值與分數的對應  | (MC)<br>(MC)                         |
|                          | 280         |  |                                      |
|                          | 270         | 279 比較小數和帶分數的大小<br>273 辨識三角形角度及性質的正確描述   | (MC)<br>(MC)                         |
| <b>268</b> 精熟 Proficient | 260         | 262 解決單位（真）分數的簡單整數倍問題  | (MC)                                 |
|                          | 250         | 258 依據四則問題情境，辨識非正確的解題列式<br>257 在非整點、整點半的情形下，計算兩時刻間的二階單位時間問題<br>252 解決多位數的整數多步驟問題<br>250 在非整點、整點半的情形下，計算兩時刻間的二階單位時間問題<br>250 依據四則問題情境，辨識正確的解題列式 | (MC)<br>(MC)<br>(MC)<br>(MC)<br>(MC) |
|                          | 240         | 247 解決有餘數的整數除法問題   | (MC)                                 |
|                          | 230         |  |                                      |
|                          | 220         | 225 在邊長相同之情況下，比較不同圖形的周長<br>221 辨識全等圖形  | (MC)<br>(MC)                         |
|                          | 210         |  |                                      |
| <b>200</b> 基礎 Basic      | 200         |  |                                      |
|                          | 0           | 193 運用二維表格中的多項資訊解決問題<br>170 運用二維表格中的兩項資訊解決問題<br>161 解決習見整數多步驟的問題   | (MC)<br>(MC)<br>(MC)                 |