

2011 – 03

**A Case Study on the TIMSS Scaling
Procedures**

林妙香、程毅豪、胡膺期、謝育泰、歐陽智聞

Jul. 20, 2011

摘要

台灣自1999年度起參加每四年一度的TIMSS計畫：評鑑四年級與八年級學生在數學與科學教育學習成就國際評比計畫。TIMSS技術報告第十一章簡化描述量尺分數的統計模式及步驟，使得國內研究TIMSS團隊對TIMSS量尺分數(Plausible value)仍缺乏實質的了解。本研究詳盡地解析 TIMSS-能力可能值的量化程序，包括說明變數 $(\theta, \mathbf{X}, \mathbf{Y})$ 在IRT架構下的能力遺失值補差法的貝氏統計原理及其相關分佈參數與模式關連性，並引進 normal quadrature逼近方法估計試題反應概率函數的參數，及呈現EM估計步驟和補差法最佳估計量的算法，以補充TIMSS -11章未說明或不清處之處，使TIMSS研究團隊能瞭解 TIMSS量尺分數的意義，進而導引以較正確方法進行次級分析TIMSS的資料。本文以TIMSS 2003 International Database為研究資料作為具體回答下列問題根據。

1. TIMSS-Report-Chapter 11 所呈現統計模式/方法的意義為何？

$$\begin{cases} t^*(\mathbf{X}, \mathbf{Y}) = E[t(\theta, \mathbf{Y}) | \mathbf{X}, \mathbf{Y}] = \int t(\theta, \mathbf{Y}) p(\theta | \mathbf{X}, \mathbf{Y}) d\theta \\ \theta = \Gamma' \mathbf{y}^c + \varepsilon \end{cases}$$

(\mathbf{X} ：試卷試題反應 (item responses), \mathbf{Y} ：背景資料, θ ：能力)

2. 基本上，TIMSS 是延用 NAEP 所採用的可能值(plausible values)的計算方法。換言之，Mislevy, Johnson, Muraki (1992) 的文獻公式 (5)、(7)、(8) 所呈現的內涵為何？

$$p(\theta | \mathbf{x}_i, \mathbf{y}_i) \propto P(\mathbf{x}_i | \theta, \mathbf{y}_i) p(\theta | \mathbf{y}_i) = P(\mathbf{x}_i | \theta) p(\theta | \mathbf{y}_i)$$

3. 如何計算 posterior mean and variance needed for imputing plausible values

$$p(\theta | \mathbf{x}_i, \mathbf{y}_i) \begin{cases} \text{covariance matrix: } \Sigma_i^p = (\Sigma^{-1} + (\Sigma_i^L)^{-1})^{-1} \\ \text{mean vector: } \tilde{\theta}_i = (\theta_i^c \Sigma^{-1} + \theta_i^L (\Sigma_i^L)^{-1}) \Sigma_i^p \end{cases}$$

4. 如何估計考生題本試題反應(混合模式)的 Likelihood function 參數((μ_i^L 與 σ_i^L)?)

$$l(\theta_i) = \prod_{j=1}^n \left([P_j(\theta_i)]^{x_{ij}} [1 - P_j(\theta_i)]^{1-x_{ij}} \right)^{v_{ij}} \left(\prod_{l=1}^{m_j-1} [P_{jl}(\theta_i)]^{u_{jll}} \right)^{1-v_{ij}}$$

5. EM algorithm ($\theta = \Gamma' \mathbf{y}^c + \varepsilon$) 估計參數過程關鍵統計問題為何？補插法產生五組能力可能值，最佳估計 $T = t(\hat{\theta}, \mathbf{Y})$ 為何？ $Var(\hat{T})$ 意涵為何？

本研究解開 TIMSS 量尺分數『黑盒子』計算程序。TIMSS 的評鑑設計及統計方法適用於大規模教育成就評鑑：瞭解或比較區域或子群體學生學習狀況或成長趨勢。因此本研究結果有助於掌握大規模評鑑的統計方法與技術，可以幫助國內自行研發 12 年國民教育所須的教育績效評鑑工作或作為亞太區域國際評比之總部，提升國際地位。

關鍵詞: Multiple Imputations, Missing Data Analysis, IRT-based Plausible Values Methods, Posterior distribution, IRT-Likelihood Function, Normal Quadrature。