

# 行政院國家科學委員會專題研究計畫 成果報告

總計畫：以 TIMSS 2003、PISA 2003、和 TEPS 跨資料庫比較  
為例

計畫類別：整合型計畫

計畫編號：NSC94-2522-S-004-002-

執行期間：94 年 08 月 01 日至 95 年 07 月 31 日

執行單位：國立政治大學教育學系

計畫主持人：余民寧

報告類型：精簡報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 95 年 7 月 26 日

# 行政院國家科學委員會專題研究計畫成果報告

影響科學與數學學習成就因素之國際評比資料分析----總計畫：以 TIMSS 2003、PISA 2003、和 TEPS 跨資料庫比較為例

**International comparison of factors affecting science and mathematics achievement ---- The comparison of TIMSS 2003, PISA 2003, and TEPS used as an example.**

計畫編號：NSC-94-2522-S-004-002

執行期限：94 年 8 月 1 日至 95 年 7 月 31 日

主持人：余民寧 國立政治大學教育學系教授

研究助理：謝進昌 國立政治大學教育學系博士生

## 一、中文摘要

本研究的目的，旨在比較跨資料庫間影響數學學習成就之因素結構是否具有不同的差異模型，以資作為政策制訂與理論模型建立之參考。

本研究根據研究動機及目的，選用 TIMSS 2003 及 PISA 2003 共有的八年級學生之數學學習成就分數，以及測量「自覺數學競爭力」與「自覺數學效用」兩個潛在變項的代表性測量指標為變項，嘗試運用多群組樣本不變性分析技術，建立一個跨資料庫間的心理特質模式。

最後，根據本研究結果提出建議，以供詮釋理論模型與制訂政策之參考。

**關鍵詞：TIMSS 2003、PISA 2003、國際評比**

## ABSTRACT

The purpose of this study is to compare the factors affecting the mathematics achievement that are across different databanks, and to form a model that can explain the results and make suggestions for policy decision making.

According to research motives and purposes, this study selects the TIMSS 2003 and PISA 2003 databanks for data analysis. And it also chooses several indicators that measure the latent variables “self-perceived competence” and “utility of math”. The multi-sample invariance analysis technique of SEM is used to fit the datasets.

Finally, the implication of findings and explanations for the model construction and decision making are proposed. And the applications and suggestions for the future researches are also proposed.

**Key words：TIMSS 2003、PISA 2003、international comparison**

## 二、緣由與目的

國科會科教處為鼓勵國內研究人員在國際教育評比（例如：IEA、OECD 等國際組織所舉辦的各類大型國際教育評比計畫）的基本數據下，針對國內教育

問題（例如：教育政策、課程、教學、學習環境、教育價值等等）及國際比較教育研究等重要議題，進行深度研究，期能於國際教育評量研討會（例如：IEA的教育評量國際研討會）或其他研討會中發表，並提供國內教育政策制定之參考，於是有徵求「國際教育評比資料之深度研究計畫」案的構想提出，並歡迎有興趣的研究人員提出個別型或整合型計畫。

因此，本研究計畫案的提出，原初的目的只是針對上述目標而來而已，後來，經筆者實地上網查詢該等資料庫，及思考研究人員可以著墨的相關問題及實務限制後，發現原初目的的構想太過狹隘，這麼龐大且重要的研究課題怎麼可能在單一個別型計畫裡來完成。因此，一方面為了達成上述目標，另一方面是因為國內在這國際比較領域的研究十分缺乏，於是才有整合型計畫案的構想形成。

本研究子題的目的，即是從國際比較的角度，鎖定八年級學生（相當於台灣國中二年級學生）在 TIMSS 2003 中的數學、PISA 2003 中的數學、與我國中研院剛建置完成並已釋放出第一波（第二波即將釋出）的台灣教育長期追蹤資料庫（Taiwan Education Panel Survey，簡稱 TEPS）中的數學資料為共同對象，作為進一步分析和探索比較各資料庫中影響數學學習成就之因素結構，希望能找出通用於此三種資料庫中影響受試者數學學習成就因素的心理特質模式。

但是，由於深入瞭解 TEPS 資料庫的設計之後，發現它與 TIMSS 2003 和 PISA 2003 的資料庫不完全相同，本研究無法找到共同的測量變項作為建立探索心理特質模式的「通式」（general form）。因此，本研究暫擬以 TIMSS 2003 和 PISA 2003 的資料庫為分析對象，但所發展出的資料分析策略和方法學，卻可以適用與推論到新資料庫的設計內容。

### 三、方法

基於上述目的，有鑑於台灣並未參與 PISA 2003 的評比，因此，本研究考量文化環境、數學成就表現皆與台灣相近的「香港」（Hong Kong），作為探討跨資料庫中影響數學學習成就因素心理特質模式的對照組之用。

於是，為能進行跨資料庫的二級資料分析，本研究特採用台灣與香港受試樣本皆各自參與其中一種測驗的相似定錨試題（similar anchor items）及相似共同受試者（similar common persons）之資料銜接設計，設計方式如表 1 所示（本研究採整列剔除法（listwise deletion）作為缺失值的處理方式）。

表 1. 本研究資料銜接設計與受試樣本的分佈情形

資料庫 \ 國家	台灣	香港
TIMSS 2003	5337	4875
PISA 2003	未參與	4435

本研究特選用 TIMSS 2003 與 PISA 2003 測驗中的數學表現成績作為效標變

項的測量指標代表，以及學生問卷中有關評量受試者對數學學習態度的問卷題目作為預測變項的測量指標代表，並據以作為建立影響數學學習成就之心理特質模式的依據。為考量跨資料庫間的題目都必須要能夠測量到相同的潛在構念（construct）的基本要求，本研究盡力挑選評量心理特質極為接近的類似題目作為預測變項測量指標代表，最後，共計挑選出 9 個題目，作為其背後的潛在構念（分別定義為受試者「自覺數學競爭力（self-perceived competence）」與「自覺數學效用（utility of math）」兩個潛在變項）的測量指標代表。因此，本研究所挑選出之題目如表 2 及表 3 所示，皆採用 Likert 式 4 點評定量表的計分方式，從「非常不同意」至「非常同意」，分別給予 1 至 4 分，得分愈高即表示受試者對於數學學習愈覺得有競爭力，且認為學習數學是有助於未來發展的。

表 2. TIMSS 2003 與 PISA 2003 中「自覺數學競爭力」的測量題目

<b>TIMSS 2003 的題目</b>
I usually do well in mathematics.
Mathematics is more difficult for me than for many of my classmates.
Sometimes, when I do not initially understand a new topic in mathematics, I know that I will never really understand it.
Mathematics is not one of my strengths.
I learn things quickly in mathematics.
<b>PISA 2003 的題目</b>
I am just not good at Mathematics.
I often worry that it will be difficult for me in Mathematics classes.
In my Mathematics class, I understand even the most difficult work.
I have always believed that Mathematics is one of my best subjects.
I learn Mathematics quickly.

註：為有效呈現原始題目，TIMSS 2003 及 PISA 2003 的題目均以英文對照表達。

表 3. TIMSS 2003 與 PISA 2003 中「自覺數學效用」的測量題目

<b>TIMSS 2003 的題目</b>
I think learning mathematics will help me in my daily life.
I need mathematics to learn other school subjects.
I need to do well in mathematics to get into the <university> of my choice.
I need to do well in mathematics to get the job I want.
<b>PISA 2003 的題目</b>
Making an effort in Mathematics is worth it because it will help me in the work that I want to do later on.
Mathematics is an important subject for me because I need it for what I want to study later on.

Learning Mathematics is worthwhile for me because it will improve my career  
<prospects, chances>.

I will learn many things in Mathematics that will help me get a job.

註：為有效呈現原始題目，TIMSS 2003 及 PISA 2003 的題目均以英文對照表達。

本研究使用結構方程式模型 (structural equation modeling, SEM) 中的多群組樣本不變性分析 (multi-sample invariance analysis) 技術，於上述三組資料庫中的資料，企圖建置一個適用於跨資料庫間的心理特質模式通式。

#### 四、結果與討論

在進行跨資料庫之多樣本不變性檢定前，本研究先利用台灣於 TIMSS 2003 資料庫樣本與香港於 PISA 2003 資料庫樣本，進行影響數學學習成就之心理特質模式的驗證性因素分析，並利用 Lisrel 8.5 版程式來執行。茲將結果呈現於下。

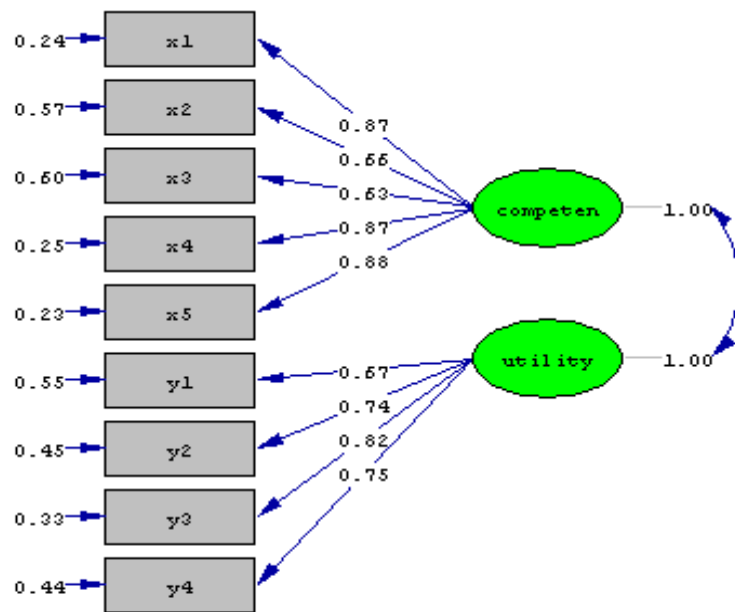
##### 一、台灣樣本於 TIMSS 2003 資料庫之驗證性因素分析結果

以台灣樣本進行影響數學學習成就驗證性因素分析結果如表4所示，在資料模式適配度方面， $\chi^2 = 768.06$ ，顯示本模式經由卡方檢定的結果，資料模式的適配度不佳。但是，卡方檢定在使用上有許多的限制，較適合當作樣本共變數矩陣與模式隱含(model-implied)共變數矩陣兩者之間的適配度衡量，比較不適合當作檢定統計量之用(Jöreskog & Sörborm, 1993)。

由於卡方分配對資料分配有較嚴格的要求，在實務上，資料分配並不容易達到多變量常態分配之要求。此外，大樣本卻是進行概似比檢定(Likelihood Ratio Test)所必需的，但此近似卡方分配的概似比檢定，卻極易受到樣本數與自由度的影響，當估計參數較多或樣本數較大時，模式都容易被拒絕。有鑑於此，卡方檢定不適合當作資料模式適配度的絕對標準，只比較適合在巢狀模式中之模式間互相比較之用。由於本研究樣本數高達5337人，因此，研究者擬參考其他的適配指標來評估資料模式的適配度。在以卡方分配為基礎的模型適配度指標方面，GFI(Goodness of Fit Index) = .9879，大於 .9的基本要求標準。在不以卡方分配為計算基準的替代指標方面，RMSEA (Root Mean Square Error of Approximation)是衡量樣本與最佳化模式兩者間共變數矩陣的差異程度，其差異程度是以自由度為計算基準，因此RMSEA對於模式的複雜度較為敏感，但比較不受樣本數波動的影響(Browne & Cudeck, 1993)。在理想狀況時，RMSEA = 0，RMSEA愈大時，表示適配度愈差；而本模式RMSEA = .073，尚符合Browne & Cudeck (1993)認為RMSEA < .08為可接受的評判標準。本模式的CFI(Comparative Fit Index) = .9511，亦符合要求。整體而言，本模式在完全不依照修正指標(modification index, MI)數據進行修改的狀況之下，顯示本資料與模式之間是適配的。而在內在結構(internal structures)方面，各題目與潛在變項之間的因素負荷量如圖1所示，亦均達顯著水準。

表4. 本研究的適配度指標摘要表

Model	$\chi^2$	df	RMSEA	CFI	GFI
台灣樣本於TIMSS 2003資料庫之驗證性因素分析(CFA)					
台灣 TIMSS 2003	768.06	26	.073	.9511	.9879
香港樣本於PISA 2003資料庫之驗證性因素分析(CFA)					
香港 PISA 2003	342.44	26	.052	.9804	.9939
跨資料庫多樣本因素結構不變性					
台灣-香港-香港	2667.17	116	.067	.9397	.9851

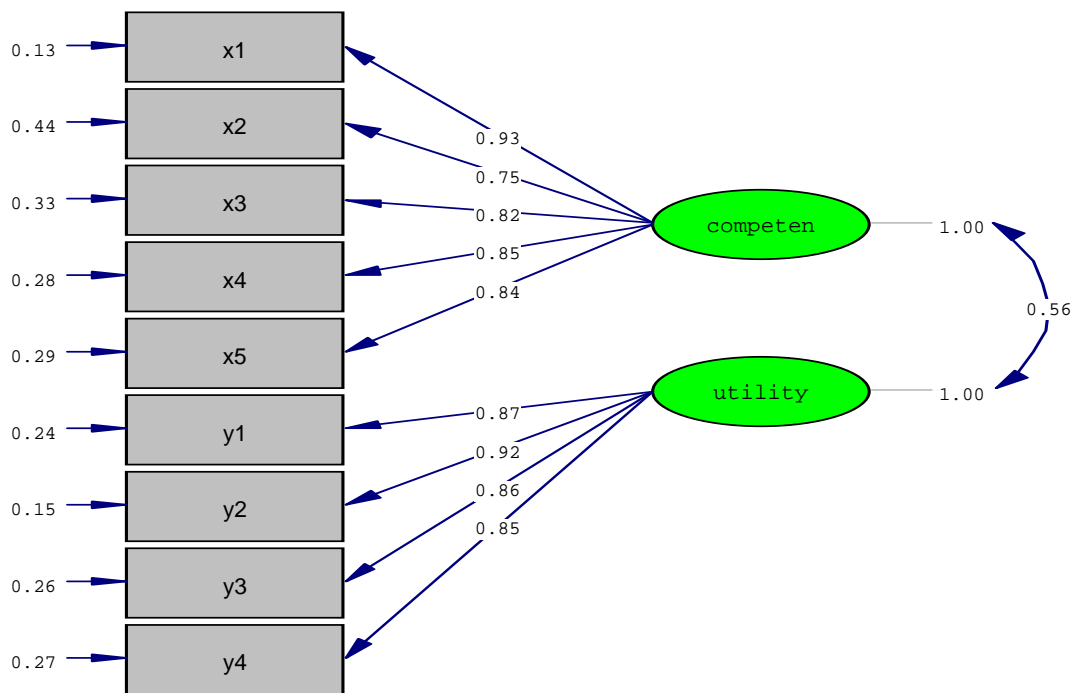


Chi-Square=768.06, df=26, P-value=0.00000, RMSEA=0.073

圖1. 台灣樣本於TIMSS 2003 資料庫的驗證性因素分析

## 二、香港樣本於 PISA 2003 資料庫之驗證性因素分析結果

以香港樣本進行驗證性因素分析的結果，亦如表4所示。在資料模式適配度方面， $\chi^2 = 342.44$ ，雖然模式經由卡方檢定的結果，資料模式的適配度不佳，但如上述，有鑑於卡方檢定的諸多缺點，乃進一步檢視其它適配指標：本研究顯示 GFI(Goodness of Fit Index)=.9939、CFI(Comparative Fit Index)=.9804 均大於.9 的基本要求標準，而 RMSEA=.052 更是符合 Browne & Cudeck (1993)認為.05、Hu & Bentler (1999)認為.06 以下是適配度良好的標準。整體而言，採用香港樣本於 PISA 2003 資料之驗證性因素分析模式，在完全不依照修正指標數據進行修改的狀況之下，其資料與模式之間的適配度是獲得支持的。而在內在結構方面，各題目與潛在變項之間的因素負荷量如圖2所示，亦均達顯著水準。



Chi-Square=342.44, df=26, P-value=0.00000, RMSEA=0.052

圖 2. 香港樣本於 PISA 2003 資料庫之驗證性因素分析(CFA)

### 三、跨資料庫多樣本因素結構不變性分析結果

在多樣本因素結構不變性檢定上，本研究需檢定跨資料庫TIMSS 2003台灣樣本--TIMSS 2003香港樣本--PISA 2003香港樣本等三組樣本的所有參數是否都相等，亦即，所有樣本是否都具有均等的因素負荷量、共變數矩陣、及測量誤差。由表4所示可知，在不作任何修正下，卡方適配度雖不能獲得支持，但在其它如GFI(Goodness of Fit Index )=.9851、CFI(Comparative Fit Index)=.9397均大於.9的基本標準，RMSEA=.067，尚符合Browne & Cudeck (1993)認為RMSEA< .08為可接受的評判標準。整體而言，顯示出三組樣本的所有參數均相等。由上述統計分析可知，在相同的TIMSS 2003資料庫內，台灣與香港兩組樣本具有均等的因素結構，兩者在因素負荷量、因素共變數矩陣、測量誤差上並無顯著差異；而在跨資料庫間，亦即在跨TIMSS 2003和PISA 2003資料庫間，同樣都是香港樣本的兩種資料亦具有均等的因素結構，兩者在因素負荷量、因素共變數矩陣、測量誤差上並無顯著差異。因此，本研究認為，跨資料庫間的不同樣本對於影響數學學習成就上具有均等的心理計量特質模式。

### 五、計畫成果自評

根據本研究發現，跨資料庫間，影響數學學習成就的因素結構模型是均等

的，亦即，TIMSS 2003 的台灣資料、TIMSS 2003 的台灣資料與 TIMSS 2003 的香港資料之間、TIMSS 2003 的香港資料與 PISA 2003 的香港資料之間，在因素結構模型中的因素負荷量、因素共變數矩陣、測量誤差上均無顯著差異，此即表示跨資料庫間，亦具有相同的表現結果。因此，雖然台灣沒有參與 PISA 2003 的評比，但亦可從本研究中推論出其具有與 PISA 2003 香港相同的影響數學學習成就的因素結構模型。本研究成果，不僅對詮釋理論模型貢獻極大，同時亦提供一種推論技術，供類推或預測尚未參與評比的未來結果，此技術對實務應用方面的預測作用深具啟示與應用價值。

## 六、參考文獻

- 余民寧 (2006)。 *潛在變項模式----SIMPLIS 的應用*。台北：高等教育。
- Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- IEA TIMSS網站(2006)。 *TIMSS & PIRLS International Study Center*, 2006 年 4 月 10 日取自 <http://www.timss.com/>。
- Joreskog, K. G., & Sorbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.
- OECD PISA網站(2006)。 *OECD Programme for International Student Assessment*, 2006 年 4 月 12 日取自 <http://www.pisa.oecd.org>。