



大型教育資料庫建置 及相關議題研討會論文集



國家教育研究院 編印

大型教育資料庫建置及相關議題 研討會論文集

廖革為、郭俊賢、曾建銘、郭伯臣、張意宗
謝進昌、陳柏霖、余民寧、薛凱方著



國家教育研究院 編印
中華民國100年3月

目次 CONTENTS

序言	III
1A Panel discussion I	01
MH法、SIBTEST、LR法三種DIF檢測方法的比較 與使用迭代式MH法及LR法與非迭代式方法的比較	
..... 一 廖革為 郭俊賢 曾建銘 郭伯臣	
2A Panel discussion II	27
2006年TASA數學科成就測驗國小六年級原住民 與非原住民學生之差別試題功能（DIF）分析研究	
..... 一 張意宗 曾建銘	
3A Panel discussion III	61
2006年TASA國二數學試題、性別的差別試題功能（DIF） 與九年一貫能力指標達成率分析研究	
..... 一 曾建銘	
4A Panel discussion IV	91
新移民與本土子女數學學習成就差別試題功能與多元潛在迴歸分析 —以2007年臺灣學生學習成就評量資料庫為例	
..... 一 謝進昌 陳柏霖 余民寧	
5A Panel discussion V	123
親子關係、憂鬱、偏差行為和幸福感結構關係模式之驗證： 以TEPS資料庫資料為例	
..... 一 薛凱方	

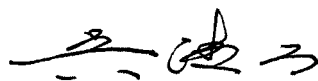
序言

隨著國內外各式大型教育評量或評比資料庫的建置，不啻是為學術圈注入一股新的研究能量，影響所及遍至教育、心理、社會、統計等領域，而國家教育研究院測驗與評量組於98年9月25日舉辦「大型教育資料庫建置及相關議題」學術研討會，希冀藉由國內專家學者闡述有關大型資料庫建置之經驗、相關議題、技術之鑽研與交流，不僅提供本院臺灣學生學習評量資料庫建置之精進，並有效提昇國內整體教育資料庫建置之水平。本專書即針對該次學術研討會中評選為口頭發表之論文，徵求作者意願並依本處學術專書審查辦法進行審查，經外審委員審查通過後之論文予以收錄。

本論文集經外審審查，收錄5篇文章，大致可分為二個面向，分別是試題差別功能以及資料庫分析-模式的驗證等議題，同時兼具技術及實務議題。首先，廖革為、郭俊賢、曾建銘、郭伯臣、張意宗、謝進昌、陳柏霖及余民寧等作者的文章，都是針對試題差別功能議題進行探討，但廖革為等人是以不同差別試題功能檢測方法之比較為取向，而張意宗、曾建銘及謝進昌等人是以實務分析為取向，運用臺灣學生學習評量資料庫探討試題是否可能為因不同性別、原住民/非原住民、本土/新移民子女等身分背景之差異，產生差別的功能，以提供學科專家探索其背後成因之基礎。而薛凱方是採用台灣教育長期追蹤資料庫進行模式的驗證，探討親子關係對於青少年憂鬱、偏差行為及幸福感的影響情形。

伴隨著本論文集有關試題差別功能以及資料庫分析-模式的驗證等文章的出版，期盼這些文獻能夠喚起更多關心資料庫分析的研究社群進一步的回響、啟發。最後，特別要感謝作者們對於各主題撰寫的投入、編輯團隊不辭辛勞的付出與努力，更要感謝每篇兩位外審委員細心的評閱、指教，使得本論文集能以最佳的論述品質，呈現在國內研究同好間。

國家教育研究院院長



1A Panel discussion I

MH法、SIBTEST、LR法三種DIF檢測方法的比較
與使用迭代式MH法及LR法與非迭代式方法的比較

— 廖革為 郭俊賢 曾建銘 郭伯臣

MH法、SIBTEST、LR法三種DIF檢測方法的比較與使用迭代式MH法及LR法與非迭代式方法的比較

廖革為

臺中教育大學教育測驗統計研究所碩士班研究生
lpause@gmail.com

郭俊賢

臺中教育大學教育測驗統計研究所碩士班研究生
keepaloof@hotmail.com

曾建銘

國家教育研究院測驗及評量研究中心
cheng@mail.naer.edu.tw

郭伯臣

臺中教育大學教育測驗統計研究所所長
kbc@mail.ntcu.edu.tw

摘要

在本研究中，採取IRT理論中2PL模式的模擬資料，使用EZDIF及SIBTEST等DIF分析軟體來進行分析，針對二階段MH法、SIBTEST及LR法的結果來進行比較，檢驗三種方法對於DIF試題檢驗的準確性及一致性。研究結果中可以發現到LR法的檢測結果較不理想，與其他的檢測方法比較，DIF試題的檢測率偏低。在此研究中也發現到試題難度對於DIF試題檢測率的影響性。接著針對MH法及LR法來進行迭代法與非迭代法的比較，檢驗統計值達顯著之DIF試題對其他試題的影響。並檢視二種方法中DIF試題的擴散情形。在研究過程中發現到了capitalization on chance的現象，以及在進行迭代的過程中移除顯著試題造成了DIF試題的擴散現象，而這些現象可能會使得DIF試題的檢測產生部分的誤判及影響。

關鍵字：差異試題功能，DIF，SIBTEST，LR，MH

The comparison of three approach to DIF item detection in MH,SIBTEST and LR and The comparison of interaered and non-iterated fashion of MH and LR

Ko-Wei,Liao

Graduate Institute of Educational Measurement and Statistics,
National Taichung University of Education
lpause@gmail.com

Chun-Hsuan Kou

Graduate Institute of Educational Measurement and Statistics,
National Taichung University of Education
keepaloof@hotmail.com

Chien-Ming Cheng

National Academy for Educational Research
E-mail:cheng@mail.naer.edu.tw

Bor-Chen Kuo

Graduate Institute of Educational Measurement and Statistics,
National Taichung University of Education
kbc@mail.ntcu.edu.tw

Abstract

The study used 2-PLM to simulate data and then analyzed the data by EZDIF and SIBTEST software. The purpose of the study is to compare the accuracy and consistency of three DIF item detection methods - MH, SIBTEST and LR in 2-stage.

In the study, the results of LR were worse than the others and we found the item difficulty would affect the accuracy of DIF item detection. In the result of between iterated and non-iterated of MH and LR method we found the significant item would affect other item's statistics and in the procedure of iteration we found that the phenomenon of capitalization on chance. To remove significant item would result in a proliferation of DIF items and will make the result of DIF item detection become erroneous judgment.

keyword : DIF, SIBTEST, LR, MH

壹、緒論

一份良好的測驗應該要能正確的反應出受試者在測驗中所要檢測的能力或特質，而不應該受到其他不相關因素的影響，如種族、性別、民族等因素，這樣的測驗才能展現其公正性，而檢驗試題的公正性及品質的重要指標就是「差別試題功能」（Differential Item Functioning，以下簡稱DIF）。在許多DIF檢測方法中，目前較為常用的有Mantel-Haenszel法（MH法）、logistic regression (LR)、SIBTEST等方法，每一種方法都有其優缺點及其使用的限制性，本研究希望藉由模擬研究來探討各種檢測方法的一致性與準確性，以瞭解各種方法的差異。

在二階段的MH法中使用迭代法（iterative method）來進行試題的淨化（purification），然而目前較常用的DIF檢測方法都是以統計量來檢測試題是否為DIF試題，在進行迭代的過程，是否會產生capitalization on chance (MacCallum, 1986)的現象，而移除達顯著的DIF試題，是否會對其他非DIF試題造成影響，進而使檢測結果發生錯誤，同樣的情況是否在LR法中也會發生，這些都是在本研究中所要探討的問題。

貳、研究背景

歐、美等國家早在1960年代便開始針對試題公正性的議題進行探討，一開始是定義為試題偏誤（item bias），但是由於其字面上的意義容易帶給人負面的感觀，便其他的學者提出「差別試題功能」（DIF）一詞，用以代替試題偏誤（Holland & Thayer, 1988）。

DIF試題的分析一直以來都是備受學者重視的領域，針對測驗試題的效度及測驗的公平性，DIF更是一樣重要的檢驗指標，當一份測驗中有著DIF試題的存在，則可能會產生對能力相同的不同群體造成不公平的現象。以目前國內的國民中學學生基本學力測驗（以下簡稱國中基測）為例，國中基測結果為考生將來入學的重要依據，若是測驗的題目經由DIF分析並且經過專家分析檢定，確實產生了試題偏誤（item bias），不僅測驗的效度及公平性將會備受質疑，對於參加測驗的考生影響更是重大。而國外大型測驗機構，如ETS（Educational Testing Service）、也都使用DIF方法來分析試題，以確定試題、測驗是否公平（Zieky, 2003）。而相關的研討會，如：American Educational Research Association (AERA)、American Psychological Association (APA)、National Council on Measurement in Education (NCME)也有針對DIF研究的領域，進行相關的討論及學術發表。

DIF分析僅是統計分析或數量分析的結果（盧雪梅，1999），針對具有DIF的試題，我們並不能直接判定其為「試題偏誤」（item bias）而將其排除在測驗之外，仍須經過進一步的審查及判定，或是經過學科專家的分析，確定該試題有與測驗不相關的其他因素所影響，而造成不同群體間的得分不同，才可以將其排除於測驗之外。

如何正確、有效及快速的檢測出DIF試題是學者們不斷努力及追尋的方向。目前已有眾多檢測DIF的方法被發展出來，隨著理論的演進及技術的進步，已有許多不同的學者針對不同的檢測方法提出改良及修正。DIF分析方法，大致上可以分為IRT(Item Response Theory)取向的DIF檢定方法與非IRT取向的DIF檢定方法。進行檢測之前必須將受試者團體進行分組，一般分為

焦點組 (focal group) 及參照組 (reference group)，焦點組為研究者感興趣的受試團體 (大多為少數族群(minority))，參照組則是用來作為對照用途的 (大多為多數族群(majority))。

IRT取向的DIF檢定方法有：Lord卡方檢驗法 (Lord, 1980)；兩組受試團體的IRF或ICC區域量數 (Camilli & Shepard, 1994; Lord, 1980; Millsap & Everson, 1993)；概似率考驗法 (likelihood ratio test, 簡稱 LR_{irt} 考驗) (Thissen, Steinberg, & Wainer, 1988)。這三種方法，Lord卡方檢驗法以及 LR_{irt} 考驗法都只提供了統計考驗上是否達顯著，並沒有辦法區別差異的程度大小，ICC區域量數雖然指出了出兩組間的差異程度，而在考驗顯著的有Raju (1990)提出的Z分配的統計量，但是仍須更多的研究來評估其檢定效能。IRT取向的檢定方式必須符合IRT理論的各項嚴格假設，同時也需要大樣本以獲得精確的參數估計，研究者必須熟悉IRT的相關理論以及IRT的程式使用，在實際運用上相當的不便。

非IRT取向的DIF檢定方式有：Mantel-Haenszel法 (Holland & Thayer, 1988, 以下簡稱MH法)；標準化法 (Dorans & Kulick, 1986)；邏輯式回歸分析法 (logistic regression, 以下簡稱LR法) (Swaminathan & Rogers, 1990)；SIBTEST (Shealy & Stout, 1993; Cheng, 2001)。其中MH法與標準化法計算都非常簡單，而且有一套完整的解釋DIF的系統，廣為測驗界使用。而LR法以及SIBTEST都有對應的程式軟體可供應用，在檢測DIF試題方面更為快速及方便 (盧雪梅, 1999)。

若是在連續的能力水平上，某一個組別相對於另一個組別都具有一定的優勢，我們將這樣的DIF稱為單向性的DIF (uniform DIF, 以下簡稱 uniform DIF) (Hambleton, Swaminathan, & Rogers, 1991; Mellenbergh, 1982)。若是這樣的優勢會隨著能力水平的不同而跟著改變，使得兩個組別的ICC曲線呈現交叉的現象，則這樣的DIF我們稱為非單向性的DIF (nonuniform DIF, 以下簡稱 nonuniform DIF) (Hambleton et al., 1991; Mellenbergh, 1982; Narayanan & Swaminathan, 1996)。

目前較多人使用的方法有：Mantel-Haenszel (MH)、logistic regression (LR)及SIBTEST。MH法過程較為簡單，而且依照ETS的分類方式可以依DIF效果量大小分為A、B、C三類，A類為輕微；B類為中度；C類為嚴重，但是MH法針對nonuniform DIF檢驗的效能非常低，僅能檢驗出uniform DIF。針對此點缺點，Swaminathan和Rogers (1990)提出logistic regression (LR)，可以有效地檢測出nonuniform DIF和uniform DIF，不過由於MH法及LR都需要逐題檢驗，程序上較為繁瑣。為此Shealy 和 Stout (1993)亦發展出 SIBTEST 用以偵測 DIF，不過SIBTEST只能針對uniform DIF進行檢測，所以針對non-uniform的DIF試題，Li和Stout (1996)提出修正的SIBTEST—Crossing SIBTEST來進行偵測。

上述方法皆以測驗總分來進行分類，若測驗中已含有DIF試題，表示測驗總分具有偏誤的情況，若發生這種情況，就無法對兩團體能力進行有效的匹配，這時便需要進行「淨化」的程序 (purification) (Holland&Thayer, 1988)，所謂淨化便是將第一階段中的DIF試題排除，將測驗總分重新計算，將兩團體的能力重新匹配，再進行一次檢定。Holland and Thayer (1988)也建議在MH法中採用二階段淨化程序，使用非DIF試題組成的總分來進行最後的DIF分析。與沒有進行淨化的測驗總分比較，使用淨化程序的測驗總分可以達到較高的DIF檢測率 (Ackerman, 1992; Fidalgo, Mellenbergh, & Muñiz, 2000)。另外也有研究指出，與未使用的淨化程序比較之下，使用迭代的淨化程序會產生更佳的DIF檢測率，特別是在DIF試題比率較高的情況下 (Candell & Drasgow, 1988; Kok, Mellenbergh, & Van Der Flier, 1985; Van der Flier, Mellenbergh, Ade`r, & Wijn, 1984)。蘇雅蕙、王文中 (2001)、蘇雅蕙 (2001)、Fidalgo et al. (2000)都曾以模擬資料檢驗MH法的淨化程序，結果顯示測驗中存在著大於10% DIF 試題時，二階段MH 法可提昇DIF 的檢驗效率。所以在本研究中決定採用二階段的MH法來進行DIF試題的檢測，以求能達到較佳的DIF試題檢定率。

Mantel-Haenszel (MH)方法是一種列聯表的分析方法，在k個分數層中，各個分數層的受試者的作答反應可以整理出一個2X2的列聯表，如表1，表中

T_k 表示得分為k的總人數， A_k 代表得分為k且答對某試題的參照組人數， B_k 代表得分為k且答錯某試題的參照組人數， C_k 代表得分為k且答對某試題的焦點組人數， D_k 代表得分為k且答錯某試題的焦點組人數， n_{Rk} 和 n_{Fk} 分別代表參照組與焦點組的人數， m_{1k} 和 m_{0k} 分別代表答對及答錯的人數

表1 總分為K的2X2列聯表

組別	試題反應		
	答對(1)	答錯(0)	合計
參照組	A_k	B_k	n_{Rk}
焦點組	C_k	D_k	n_{Fk}
合計	m_{1k}	m_{0k}	T_k

MH 法檢定的虛無假設為：這K個分數層的參照組和焦點組的共同勝算比參數 α_{MH} 的值等於1.0。 α_{MH} 的估計值如下：

$$\hat{\alpha}_{MH} = \frac{\sum_k A_k D_k / T_k}{\sum_k B_k C_k / T_k}$$

Mantel 和Haenszel(1959)提出一卡方統計數來考驗 α_{MH} 等於1.0 的假設：

$$x_{MH}^2 = \frac{\left[\sum_k A_k - \sum_k E(A_k) \right] - 0.5}{\sum_k Var(A_k)}$$

$$\text{其中： } E(A_k) = \frac{n_{Rk} m_{1k}}{T_k}, \quad Var(A_k) = \frac{n_{Rk} n_{Fk} m_{1k} m_{0k}}{T_k^2 (T_k - 1)}$$

當虛無假設為真的時候，服從自由度為1的卡方分配，若是拒絕虛無假設，代表試題呈現DIF。

在實際應用上，為了便於解釋分析的結果，ETS已發展出一套理論架構，將 α_{MH} 的值轉換成另一種形式的DIF量數，稱為MH D-DIF()，轉換公式如下：

$$\Delta_{MH} = -2.35 \ln(\alpha_{MH})$$

Holland和Thayer(1988)進一步提出其標準誤的公式如下：

$$SE(\Delta_{MH}) = 2.35 \sqrt{\widehat{Var}(\ln(\alpha_{MH}))}$$

上式中：
$$\widehat{Var}(\ln(\alpha_{MH})) = \frac{\sum_k \frac{U_k V_k}{T_k^2}}{2 \left(\sum_k \frac{A_k D_k}{T_k} \right)^2}$$
，其中：

$$U_k = (A_k D_k) + \hat{\alpha}_{MH} (B_k C_k), \quad V_k = (A_k + D_k) + \hat{\alpha}_{MH} (B_k + C_k)。$$

Δ_{MH} 為一個等距變數，是以ETS的難度量尺指標來解釋參照組與焦點組的難度差異，若 Δ_{MH} 值為正，表示試題有利焦點組，若 Δ_{MH} 值為負，表示試題有利參照組。(余民寧 & 謝進昌，2006；盧雪梅 & 毛國楠，2008)

logistic regression (LR)方法採用跟MH法類似的列聯表，但是與MH法所不同的是，LR法將測驗總分視為連續變數，根據測驗總分推估組別的預測答題表現，Swaminathan 和 Rogers (1990)提供了一個基本模式如下：

$$p(u = 1 | \theta, g) = \frac{e^{\beta_0 + \beta_1 \theta + \beta_2 g + \beta_3 (\theta g)}}{1 + e^{\beta_0 + \beta_1 \theta + \beta_2 g + \beta_3 (\theta g)}} \quad \text{模式一}$$

$p(u=1 | \theta, g)$ 表示再觀察能力 θ 下， g 組別的受試者答對第 i 題的機率。 θ 表示觀察到的能力水準，通常以測驗總分來代表， g 表示群組標記，通常以0和1來代表。 (θg) 表示 θ 和 g 的乘積，回歸係數 β_0 、 β_1 、 β_2 、 β_3 分別代表難度、觀察能力、組別在答題表現上的差異、組別和能力的交互作用。另外在LR法中還有兩種模式，分別為：

$$p(u = 1 | \theta, g) = \frac{e^{\beta_0 + \beta_1 \theta + \beta_2 g}}{1 + e^{\beta_0 + \beta_1 \theta + \beta_2 g}} \quad \text{模式二}$$

$$p(u = 1 | \theta, g) = \frac{e^{\beta_0 + \beta_1 \theta}}{1 + e^{\beta_0 + \beta_1 \theta}} \quad \text{模式三}$$

模式二是由模式一縮減而來，少了組別與能力的交互作用項。模式三是由模式二再縮減而來的，比較之下，少了組別項。LR法的DIF檢定便是進行每一試題的三個模組的比較，以差異對數概似率考驗(difference log-likelihood ratio test)來進行比較，比較的步驟是先比較模組一和模組二，檢定 β_3 是否等於0，並產生一個自由度為1的卡方分配來進行考驗，若拒絕 β_3 等於0的假設，則表示能力與組別之間的交互作用達到顯著，試題呈現非單向性DIF (nonuniform DIF)。若接受 β_3 等於0的假設，繼續比較模式二和模式三，同樣也以自由度為1的卡方分配來進行考驗，若是拒絕 β_2 等於0的假設，表示試題出現單向性DIF (uniform DIF)。

SIBTEST的主要作法是將整份測驗試題分為可疑部分測驗 (suspect subtest) 及有效部分測驗 (valid subtest)。可疑部分測驗是由可能為DIF的試題所組成，題數可能為一題或是多題。有效部分測驗則是由未受懷疑有DIF的試題所組成。假設地1題至第n題為有效部分測驗，第 n+1 題至第N題為可疑部分測驗。 U_{gij} 為在組別g (焦點組或是參照組) 中第 j 位受試者在第i題的作答反應。分別計算第 j 位受試者在有效部分測驗 ($X_{gj} = \sum_{i=1}^n U_{gij}$) 及可疑部分測驗 ($Y_{gj} = \sum_{i=n+1}^N U_{gij}$) 的總分。接著進行相同能力下兩群體的表現差異，與分別代表每組受試者在可疑部分試題下的平均得分，計算方式如下：

$$\bar{Y}_{RK} = \frac{1}{J_{Rk}} \sum_{j=1}^{J_{Rk}} Y_{Rj} \quad k=0, 1, 2, \dots, n$$

$$\bar{Y}_{FK} = \frac{1}{J_{Fk}} \sum_{j=1}^{J_{Fk}} Y_{Fj} \quad k=0, 1, 2, \dots, n$$

其中的 \bar{Y}_{RK} 與 \bar{Y}_{FK} 分別代表將的分數從0分到n分，分為n+1個組別，每個組別中包含參照組與焦點組成員得到k分的人數。進而計算DIF估計值：

$$\hat{\beta}_U = \sum_{k=0}^n \hat{p}_k (\bar{Y}_{RK} - \bar{Y}_{FK}) \quad , \quad \text{其中} \quad \hat{p}_k = \frac{(J_{Fk} + J_{Rk})}{\sum_{k=0}^n (J_{Fk} + J_{Rk})} \quad .$$

接著計算在有效部分測驗中獲得k分的受試者於可疑部分測驗得分的變異數：

$$\hat{\sigma}^2(Y|k, g) = \frac{1}{J_{gk} - 1} \sum_j^{j_{gk}} (Y_{gj} - \bar{Y}_{gk})^2, \text{ g代表焦點組或參照組。}$$

最後定義DIF統計檢定量 $B = \frac{\hat{\beta}_U}{\hat{\alpha}(\hat{\beta}_U)}$ 。其中 $\hat{\alpha}(\hat{\beta}_U)$ 為：

$$\hat{\alpha}(\hat{\beta}_U) = \left(\sum_{k=0}^n \hat{p}_k^2 \left(\frac{1}{J_{Rk}} \hat{\sigma}^2(Y|k, R) + \frac{1}{J_{Fk}} \hat{\sigma}^2(Y|k, F) \right) \right)^{1/2}$$

當DIF不存在時，統計檢定值B會接近標準化常態分配N(0,1)，若表示拒絕沒有DIF的虛無假設，表示某試題具有DIF的現象。

在上述的DIF檢定方式中，使用MH法及LR法來進行DIF檢測，可以看到是完全依賴統計數據的，而這種依數據處理(data-driven)的特質可能會產生capitalization on chance的現象 (MacCallum, 1986)，造成某些結果的錯誤和誤判，這種現象就如同SEM (Structural Equation Modeling)中的模式搜尋 (specification search)，它是以迭代的方式進行搜索。

在本研究中，將採取模擬資料，以二階段MH法、SIBTEST及LR法來進行比較，檢驗三種方法對於DIF試題檢驗的準確性。並且針對MH法及LR法來進行迭代法的比較，檢驗DIF試題對其他試題的影響。

參、研究目的與研究問題

一、研究目的

依據上述的研究背景，本研究以模擬資料為主，主要研究目的如下：

- (一)比較二階段MH法、SIBTEST及LR法三種DIF試題分析方法的準確性。
- (二)在LR法及MH法中經統計檢驗達顯著之試題，對其他試題之影響。
- (三)進行試題分析時，capitalization on chance的現象對DIF試題判定的影響。

二、研究問題

依據上述的研究目的，本研究的研究問題如下：

- (一)二階段MH與SIBTEST及LR法三種DIF試題分析方法是否有差異？
- (二)進行迭代的LR法及MH法是否會對DIF試題造成影響？
- (三)進行迭代法將最顯著試題移除，移除順序是否與未進行迭代法的前四題最顯著試題的順序是否有差異？

肆、研究設計與方法

一、研究工具

本研究使用Microsoft Excel 2003產生試題參數、受試者能力及計算出試題反應。使用美國Vanderbilt 大學教授 N.G.Waller (1998) 所發展的EZDIF軟體來進行MH法及LR法的DIF分析，使用美國Illinois 大學教授 W.F.Stout所發展的SIBTEST軟體來進行SIBTEST法的DIF分析。使用SPSS及MATLAB來進行模擬資料的整理與分析。

二、資料模擬

本研究使用IRT理論的二參數對數模式來進行資料的模擬，模擬試題參數及受試者的能力值，**a**參數的模擬採用均勻分佈，數值介於0.8~2之間。**b**參數的模擬也是採用均勻分佈，數值介於-2~2之間。受試者的能力參數的模擬採用常態分佈， $N(0,1)$ 。利用模擬出來的試題參數及能力值定配合IRT理論的二參數對數模式(two parameter logistic model)的公式，計算出模擬受試者的作答反應，針對這樣的條件進行模擬操作20次。詳細模擬說明如下：

- (一)測驗長度：本研究採用40題的測驗長度來進行模擬分析。
- (二)樣本大小：本研究採用的樣本數為2000人，對照組及焦點組的比例為1：1，樣本數為(1000，1000)。
- (三)DIF比率：本研究模擬試題數為40題，設定DIF試題為5題，DIF試題比率為12.5%。分別為第5、12、18、27、35題，相關試題參數如表2所示。
- (四)DIF類型：本研究只模擬單向性DIF的情境，針對單向性DIF來進行模擬操作。
- (五)DIF強度：根據操弄**b**參數來產生DIF試題，本研究將兩組試題的**b**參數調整為相差0.6，試題對於參照組有利。**a**參數保持不變。

表2 模擬資料DIF試題之試題參數

試題	焦點組		參照組	
	a參數	b參數	a參數	b參數
試題5	1.701	1.209	1.701	0.609
試題12	1.072	-0.616	1.072	-1.216
試題18	0.967	0.967	0.967	0.367
試題27	0.882	1.766	0.882	1.166
試題35	1.526	-0.435	1.526	-1.035

三、 研究方法

(一)資料模擬：

步驟一：利用前面所述的資料模擬的條件，模擬出參照組及焦點組各1000人，及40題題目的a參數及b參數。

步驟二：根據IRT理論的公式計算出各能力下的答對機率 $P(\theta)$ 。

步驟三：產生一個介於0,1之間的隨機亂數，與步驟二的答對機率進行比較，若是 $P(\theta)$ 大於該隨機亂數，則記錄為答對，反之記錄為答錯。

步驟四：重複步驟二至三，進行20次的資料模擬。

(二)DIF檢測方法比較：

1. 將模擬資料以SPSS整理為EZDIF及SIBTEST的檔案格式，進行分析，EZDIF檢測出的MH法及LR法數據，MH法採用2-step的MH法，將達顯著(p值小於0.01)的試題記錄為DIF試題。由於EZDIF軟體可依照ETS的分類標準來標記試題，所以也將在結果中標記為B或C的試題視為DIF試題並記錄之，以利後續的差異比較。而LR法以p值等於0.01為顯著水準，只針對單向性DIF進行分析，將p值小於0.01的試題視為DIF試題。SIBTEST的檢測，先將所有試題假定為DIF試題來進行分析，同樣以0.01的p值為顯著水準，分析DIF試題。接著再將同樣的資料進行一次SIBTEST的檢測，將第一次檢測為DIF試題的題

目重複檢驗，找出 p 值達顯著水準的試題。

2. 計算出各檢測方法的第一類型錯誤率(將非DIF試題判定為DIF試題的比率)， $power$ 檢定力(正確檢測出DIF試題的比率)。

(三) 進行迭代法刪除顯著試題：

1. 迭代式MH法：

使用EZDIF進行分析，將2-step的MH法的結果記錄下來，依照各題的 p 值由小到大排序，再依照各題的卡方值由大到小排序，將前四題的題目記錄下來。將排序第一的題目移除後，使用EZDIF進行分析，依照一樣的排序方式進行排序，在將排序後結果排名第一的試題移除，進行EZDIF的分析，如此迭代至移除四題題目後停止，將移除題目的順序記錄下來，與未進行迭代之前排序前四題的題目進行比較，並觀察其他題目 p 值的變化。

2. 迭代式LR法：

與迭代式MH法一樣的作法，但是因為本研究只針對單向性DIF試題來進行分析，所以記錄的 p 值，以單向性DIF的 p 值為主。一樣進行刪去的題目與非迭代法中排名前四名的題目進行比較。

伍、研究結果

一、DIF試題分析結果

經過20次的模擬研究，針對單向性DIF來進行分析，結果如表3所示，可以很明顯的發現LR法的檢測率非常低，在0.01的顯著水準下，在總計100題的DIF試題中，僅檢測出29題，雖然LR法的第一類型錯誤率是最低的，但過低的power檢定力，其整體檢定的結果是令人失望的。

其他三種的檢定方法，檢測率都在0.95以上，第一類型錯誤率也都控制良好，在MH法中可以看到，原先採用p值來判斷是否為DIF試題的數據與後來採用ETS分類標準的數據比較，可以發現第一類型錯誤率降低至原來的一半，不過檢測率也稍為降低。所以在表3中的各項數據中，除了LR法之外，其他的方法都有著不錯的檢測率和控制良好的第一類類型錯誤率。

表3 DIF檢測方式準確率比較表

檢測法	第一類型錯誤率(題數)	檢測率(題數)
MH法(p值)	0.021(15)	0.990(99)
MH法(ETS)	0.010(7)	0.950(95)
SIBTEST	0.016(11)	0.980(98)
LR法	0.003(2)	0.290(29)

在表4中可以看到本研究模擬的DIF試題在各種方法中被檢測到的次數，可以看到在LR法中各題的檢測次數都是較低的，尤其是題目27只有被檢測到一次，在其他的檢測方法中，各DIF試題幾乎都可以檢測出來，不過在試題27，可以發現其檢測次數是偏低的，針對這個現象，對照表2，可以發現試題27的b值是偏高的，屬於高難度的題目，即使試題對參照組有利，參照組的難度也是偏高的，這樣的情況可能會造成普遍得分偏低，使得兩組間的差異縮小，可能是造成檢測次數偏低的原因。

表4 DIF試題在各種方法的檢測次數

檢測方法	試題35	試題12	試題5	試題18	試題27
MH法(p值)	20	20	20	20	19
MH法(ETS)	20	20	20	18	17
SIBTEST	20	20	20	20	18
LR法	9	7	5	7	1

二、 迭代式MH法及LR法與非迭代式MH法及LR法

(一)迭代式MH法

在進行四次迭代之後的結果如表5所示，表5所記錄的數據分為非迭代及迭代的順序，分別記錄下試題在兩種情況下排序或被刪除的次數，可以看到在題目35，及題目18在兩種情況下產生了不同的順序排列，但是其差異並沒有很明顯，可能的原因是因為本模擬研究的難度差異為0.6，所以在使用MH法進行DIF試題檢測時，檢測出正確DIF試題的比率非常高，而在模擬結果中也可以發現這些題目的卡方值非常大，所以對題目順序的影響可能會有所降低。但是在MH法中可以觀察到另一個現象，就是DIF試題的擴散，在幾次的模擬分析中可以發現，某些試題在進行未刪題的DIF檢測試時，並未達到顯著，但是當刪除某幾題後，這些題目卻達顯著，成為DIF試題，如同表6所示，除了題目27是模擬研究中設定的DIF試題之外，其他試題都並非DIF試題，但是在進行若干次迭代之後，可以發現都達顯著，反而使原先的非DIF試題變成了DIF試題。

表5 非迭代與迭代MH法比較

題目	非迭代 MH法排序				迭代MH法刪題次序			
	1st	2nd	3rd	4th	1st	2nd	3rd	4th
5	14	4	2	0	14	4	2	0
35	5	8	6	1	5	8	5	2
18	1	4	6	5	1	4	7	4
12	0	4	4	7	0	4	4	7
27	0	0	2	7	0	0	2	7

表6 非迭代與迭代MH法DIF試題的擴散

模擬次數	題目	非迭代式MH法		迭代式MH法		
		p值	χ^2 值	迭代次數	p值	χ^2 值
2	29	0.026	4.938	4	0.01	6.713
9	8	0.013	6.204	2	0.01	6.662
14	20	0.012	6.318	2	0.007	7.19
20	27	0.012	6.329	4	0.009	6.915

(二)迭代式LR法

在LR法進行四次迭代後，將刪題的順序及未刪題前的排序狀況加以統計，結果如表7所示，可以看到在某一些題目與進行迭代式的MH法中的部份題目類似，也發生了capitalization on chance的現象，進行迭代的刪題順序與未刪題前達顯著試題的排序不一樣，如題目18與題目16都有發生這樣的現象，但是在MH法中DIF試題的擴散現象，並未在LR法中發生。

表7 非迭代與迭代LR法比較

題目	非迭代 LR法排序				迭代LR法刪題次序			
	1st	2nd	3rd	4th	1st	2nd	3rd	4th
35	7	4	1	2	7	4	1	2
12	6	4	1	1	6	4	1	2
18	3	4	2	2	3	4	2	3
5	3	2	3	1	3	2	3	1
10	1	1	0	2	1	1	0	2
27	0	0	2	2	0	0	2	2
16	0	0	2	0	0	0	3	1
19	0	0	1	1	0	0	1	1
37	0	0	1	1	0	0	1	1
1	0	0	0	1	0	0	0	1
4	0	0	1	0	0	0	1	0
7	0	1	0	0	0	1	0	0
8	0	0	1	0	0	0	1	0
13	0	1	0	0	0	1	0	0
15	0	0	1	0	0	0	0	1
17	0	0	1	0	0	0	1	0
21	0	0	0	1	0	0	0	1
25	0	1	0	0	0	1	0	0
29	0	0	1	0	0	0	1	0
32	0	0	1	0	0	1	0	0
33	0	1	0	0	0	0	1	0
34	0	0	0	2	0	0	0	0
38	0	0	0	1	0	0	0	1
39	0	1	0	0	0	1	0	0
3	0	0	0	1	0	0	0	0
9	0	0	0	1	0	0	0	0
14	0	0	0	1	0	0	0	0
31	0	0	0	0	0	0	0	1
36	0	0	1	0	0	0	0	0
40	0	0	0	0	0	0	1	0

陸、結論與建議

一、研究結論

(一)在使用模擬資料的研究中，三種DIF檢測方法，使用p值來檢驗DIF試題的MH法、使用ETS分類來檢驗DIF試題的MH法、SIBTEST，這些方法的結果都呈現了高度的power檢定力，而且第一類型錯誤率也都有良好的控制，整體來說，檢測的準確度是非常良好的，在模擬資料中的DIF試題幾乎都可以準確的檢測出來。

(二)在DIF檢測方法中可以看到LR法的結果並不是很理想，整體的power檢定力低於0.3，明顯的偏低，所以在本模擬研究的條件設定之下，LR法的表現相較於其他的檢測方法並不理想，對於DIF試題並不能有效的檢測出來。

(三)在本次模擬研究中，可以發現到試題的難度會影響到檢測結果，試題本身若是屬於高難度的題目，相對的會增加其檢測的難度，在模擬研究的結果中可以看到，試題27在各種方法的檢測率，相較於其他的試題，都是偏低的。

(四)在迭代式的MH法與LR法的結果可以看到capitalization on chance的現象，也顯示了在迭代過程中刪除統計值達顯著的試題，對於剩下的試題的統計值是會產生變化，尤其在迭代式MH法中可以看到DIF試題擴散的現象，在二階段的MH法是使用迭代式的淨化，來獲得乾淨試題(pure item)，但是本模擬研究的結果中可以看到，移除了達顯著的DIF試題之後，可能會使得原先並非DIF試題的統計值達顯著，而成為DIF試題，MH法及LR法都是依靠統計檢定值來檢驗DIF試題，而這樣的結果，讓我們往後在進行DIF試題分析的時候，必須更謹慎嚴謹的看待分析結果的檢定值，以求檢測結果的準確度。

二、研究建議

(一)本模擬研究只針對樣本數2000人的條件下進行模擬，DIF試題難度差異固定為0.6，DIF試題比率固定為12.5%，建議可以針對不同樣本數及不同的難度差異來進行模擬研究，也可以調整DIF試題的比率，比較這些檢測方法

在不同條件的表現情況。進而探討影響DIF檢定結果的因素為何？

(二)在本模擬研究的條件設定下，LR法的表現不甚理想，建議可以調整相關的條件設定，進行模擬研究，檢驗是否在其他的條件設定之下，LR法可以有較佳的檢定效果。

(三)迭代式的MH法與LR法的比較，由於本模擬研究設定的難度差異較大，DIF試題較容易檢測出來，建議可以調整難度差異，增加樣本數，減少DIF試題的比率，來檢測capitalization on chance的現象是否會更為明顯，及DIF試題擴散的現象是否會更為嚴重。

參考文獻

- 余民寧、謝進昌(2006)。國中基本學力測驗之 DIF 的實徵分析：以 91 年度兩次測驗為例。《國立高雄師範大學教育學系教育學刊》，26 期，頁 241-276。
- 盧雪梅 (1999)。試題差別功能 (DIF) 的檢定方法。《台北市立師範學院學報》，30 期，頁 149-165。
- 盧雪梅、毛國楠(2008)。國中基本學力測驗數學科之性別差異和差別試題功能(DIF)分析。《教育實踐與研究》，21(2)，95-126。
- 蘇雅蕙(2001)。多分題差異試題功能之檢定。國立中正大學心理學研究所碩士論文，未出版，嘉義縣。
- 蘇雅蕙、王文中(2001)。三種 MANTEL-HAENSZEL DIF 檢驗程序的效果。第五屆華人社會心理與教育測驗學術研討會。台北：台灣師大教育學院。
- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29(1), 67-91.
- Camilli, G., & Shepard, L.A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Candell, G. L., & Drasgow, F. (1988). An iterative procedure for linking metrics and assessing item bias in item response theory. *Applied Psychological Measurement*, 12, 253-260.
- Cheng, C. M. (2001). *A study on Differential Item Functioning (DIF) of the Basic Mathematical Competence Test for Junior High Schools in Taiwan*. Unpublished doctoral dissertation, Michigan State University.
- Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the scholastic aptitude test. *Journal of Educational Measurement*, 23, 355-368.
- Fidalgo, A. M., Mellenbergh, G. J., & Muñiz, J. (2000). Effect of amount of DIF, test length, and purification type on robustness and power of Mantel-Haenszel. *Methods of Psychological Research Online*, 5(3), 43-53.

- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun(Eds.), *Test validity* (pp.129-145). Hillsdale, NJ: Lawrence Erlbaum.
- Kok, F. G., Mellenbergh, G. J., & Van Der Flier, H. (1985). Detecting experimentally induced item bias using the iterative logit method. *Journal of Educational Measurement, 22*, 295-303.
- Li, H. H., & Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika, 61*(4), 647-677.
- Lord, F. M. (1980). *Applications of item response theory to practical problems*. Hillsdale, NJ: Lawrence Erlbaum.
- MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100*(1), 107-120.
- Mantel, N., & Haenszel, W. M. (1959). Statistical aspects of the analysis of data from respective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics, 7*, 105-118.
- Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement, 17*, 297-334.
- Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement, 20*, 257-274.
- Raju, N. S. (1990). Determining the significance of estimated signed and unsigned area between two item response functions. *Applied Psychological Measurement, 14*, 197-207.
- Shealy, R., & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects that bias/DIF as well as item bias/DIF. *Psychometrika, 58*(2), 159-194.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370.
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I Braun (Eds.), *Test validity* (pp. 147-169). Hillsdale, NJ: Erlbaum.

Van Der Flier, H., Mellenbergh, G. J., Ade`r, H. J., & Wijn, M. (1984). An iterative item bias detection method. *Journal of Educational Measurement*, *21*, 131-145.

Zieky, M. (2003). *A DIF Primer*. Princeton, NJ: Educational Testing Service.

▶2A Panel discussion II

2006年TASA數學科成就測驗國小六年級原住民 與非原住民學生之差別試題功能（DIF）分析研究

— 張意宗 曾建銘

2006年TASA數學科成就測驗國小六年級原住民與非原住民學生之差別試題功能（DIF）分析研究

張意宗

國立台中教育大學教育測驗統計研究所碩士班研究生

E-mail : tr.ciz@msa.hinet.net

曾建銘

國家教育研究院測驗及評量研究中心

cheng@mail.naer.edu.tw

摘 要

本研究根據2006年TASA數學科成就測驗之國小六年級施測後所蒐集到的學生作答資料進行分析。

本研究結果摘述如下：

- 一、二族群學生在2006年TASA小六數學科成就評量的學習成就表現，存在著顯著的差異，小六非原住民學生的數學學習成就高於原住民學生。
- 二、分析小六原住民與非原住民學生二族群的DIF現象，結果有6題呈現DIF，約占全部試題103題的6%。但由於這6題DIF試題內容並未出現在公告版，所以無法進一步分析是否真正屬於偏誤試題。
- 三、於存在有DIF試題的試題區塊中，進行DIF試題刪除前與刪除後之小六原住民與非原住民學生的平均總分比較，結果顯示每一試題區塊均呈現顯著差異，即非原住民學生的TASA測驗平均總分高於原住民學生的平均總分之情況，並未受到DIF試題存在與否的因素而受影響，但在本研究的研究對象中，因原住民學生（294名）與非原住民學生（7663名）的人數相差太多，所以於進行DIF試題的檢測前，已先經過學生能力的配組程序，所以此項測驗平均分數比較的結果亦有可能是能力配組抽樣所造成。

關鍵詞：臺灣學生學習成就評量資料庫、差異試題功能、偏誤試題。

Analysis of DIF on 2006 TASA Mathematics Achievement Test For Aboriginal and Non-indigenous Students in Sixth Grade

Chang Yi-Tsung

Master degree student, Graduate Institute of Educational Measurement and Statistics,
National Taichung University of Education
E-mail:tr.ciz@msa.hinet.net

Chien-Ming Cheng

National Academy for Educational Research
E-mail:cheng@mail.naer.edu.tw

Abstract

The study is to analyze the students' response data gathered from the sixth grade in 2006 TASA mathematics achievement test.

The results are as follows:

- (A)The test result performance on 2006 TASA mathematics achievement test for aboriginal and non- aboriginal students in sixth grade is significantly different. The result shows that the non- aboriginal students' achievement is higher than the aboriginal students'.
- (B)Analyzing the DIF items of aboriginal and non-aboriginal students, it shows that 7 test items are DIF, which are 7% of the 103 test items. However, the three DIF items are not on the bulletin board, it is not possible to do further analysis whether these are biased items or not.
- (C)In the block of DIF test items, compare the 6th grade aboriginal and non-aboriginal participants' average scores (without DIF data) with the previous score (with DIF data), and found that the average score is significantly different. The result shows: whether the DIF items were deleted or not, they don't influence the result that non-aboriginal students' average score is higher than the aboriginal students'. In the study, the number of the aboriginal (n=294) and non-aboriginal (n=7663) differs greatly; therefore, students are matched for abilities before DIF test, so the average test score result possibly is caused by sampling.

Keywords: Taiwan Assessment of Student Achievement , TASA, Differential Item

壹、研究背景與目的

教育是謀求國家社會進步的主要動力，我國憲法強調教育機會均等，發展多元文化教育理念。其中，台灣原住民的教育問題，一直是社會所關心的問題。台灣社會有著特別重視教育的特殊文化傳統（黃毅志，2002），但是，僅佔台灣地區人口2%（1.9%）不到的原住民（行政院原住民族委員會，2006），受教育的機會卻非常有限，根據調查研究顯示，台灣原住民學生的就學率及升學率未臻理想，在學成績普遍偏低（教育部，1996）。張建成、黃鴻文、譚光鼎（1993），在研究中歸納分析原住民教育的劣勢原因：可能是家庭經濟困難，家裡讀書環境不佳及家長持反對態度；可能是學習新語言的困難、學習異文化價值觀念的問題及家庭與社會學習環境差異的壓力；可能是學生學習態度不佳及自我發展不利；可能是教師心態偏差、師資人才結構失調及教育經費與設備短缺。但原住民學生的教育問題，除了以上所述的因素之外，是否還有其他的因素影響著原住民學生的學習成就？

余民寧(1993)曾提到社會大眾對心理測驗或教育測量有個相當迫切的關注，那就是測驗的公平性(test fairness)問題。由於編製測驗試題的專家，受到自己本身的專業素養、國學程度、文化認知、甚至主觀偏見等限制和影響，以致所編製出的試題有時只會有利於某些族群的考生，而不利於另一些族群的考生。所以，在以漢族文化為背景的課程設計中，根據九年一貫數學課程能力指標所編製的試題，對於具有相同能力的兩族群，非原住民學生與少數族群(原住民)學生，是否會在答題表現上呈現顯著的差異？進而造成原住民學生答題上的不公平。

在針對試題對於不同背景的受測群體之公平性，「差別試題功能」(Differential Item Functioning，簡稱 DIF)的研究一直受到重視。來自不同的族群或團體，但能力相同的個人，在做答某試題上的機率卻有所不同，則該試題便具有DIF現象（余民寧，1993）。學者今多以DIF這個詞來取代先前研究所謂的「試題偏誤」，並且對DIF與試題偏誤做進一步的區分（Camilli

and Shepard, 1994)，不再視二者為同義詞。DIF和因受試群體能力不同所造成的表現差異要加以區別，若受試群體未經任何的配組程序，而導致兩群體表現有所差異稱為impact。DIF僅是統計分析或數量分析的結果，DIF是試題偏誤的必要但非充分條件，若試題呈現DIF，尚須經過學科專家的審視與分析，確實發現試題包含與測驗構念無關之因素，導致試題難度對於不同背景之受測群體的不相等，才能確認該試題是否為「偏誤試題」（biased item）。

DIF的存在表示試題對於測驗結果的公平性可能有負面的影響，即同一試題於能力或表現相配比的兩團體或兩族群中，在答題表現上呈現顯著的差異，可能是因不同性別、族群或是地區性教學差異、生活經驗不同、資源分配不公等因素所造成。所以，DIF反映的不只是試題適不適合的依據，而且有可能是課程的內容與教學需要做改變。

由於DIF的研究多以中學生為研究對象，少有以國小學生為對象進行DIF探討（Lane, Wang and Magone, 1996；Ryan and Fan, 1996），而且大部分都針對性別的試題DIF情形進行實徵資料的研究，對於原住民與非原住民學生的DIF研究與文獻資料卻付之闕如。基於以上的動機敘述，再加上數學科學習成就的議題一直都是家長們與教育單位所注重與關心的科目之一，所以本研究擬根據教育部委託國家教育研究院籌備處所建置之「臺灣學生學習成就評量資料庫，TASA」2006年數學科國小六年級施測後所蒐集到的學生作答資料來進行分析，以評量中共同問卷內容之「是否為原住民」作為研究背景變項，進一步探討小六原住民與非原住民學生在2006年TASA國小六年級數學科成就評量的學習成就表現與試題DIF的情形。而根據行政院原住民族委員會2006年5月所公佈的原住民人口數統計資料（行政院原住民族委員會，2006），經台灣政府認定的原住民族有：阿美族（Amis）、泰雅族（Atayal）、排灣族（Paiwan）、布農族（Bunun）、卑南族（Puyuma）、魯凱族（Rukai）、鄒族（Tsou）、賽夏族（Saisiyat）、達悟族（Yami）、邵族（Thao）、噶瑪蘭族（Kavaian）、太魯閣族（Truku）等12族。

但因TASA的試題具有保密性，所以對於存在有DIF現象的試題，將只提供統計數據，作為TASA命題小組進行試題質的分析，以確認該DIF試題是否

為試題偏誤（item bias），若為偏誤試題，應該如何修改、避免、刪除或針對教學內容與教學方法作改進，以達到考試的公平性。

貳、研究方法

本研究目的主要是以原住民與非原住民作為分組變項，探討原住民學生與非原住民學生在2006年TASA國小六年級數學科成就表現之差異性，以及該成就評量之試題對於二受測族群學生是否存在有DIF現象與DIF試題對於二族群學生數學學習成就的影響性。以下就本研究的研究對象、研究工具以及資料處理與分析分別加以說明：

一、研究對象

2006年TASA國小六年級數學科學習成就評量之實徵資料中，評量試題共分為13個試題區塊，每一個試題區塊分別有8題試題，試題數總計有 $13 \times 8 = 104$ 題，但因M12_2試題（第12個試題區塊第2題）於實徵資料中被刪除，所以本研究所能採用的實際試題數作答資料共計有103題。

成就評量總共有13種試題本，每位受測學生只作答其中的一種試題本，每一種試題本包含3個試題區塊，因此，每一種試題本有 $3 \times 8 = 24$ 題選擇題；但試題本8、試題本11與試題本12，因刪除M12_2試題，所以此三種試題本均只有23題選擇題的學生作答資料進行資料統計與分析。試題本中的試題區塊分配，詳見表1。

表1 2006年TASA國小六年級數學試題本的試題區塊分配

試題本編號	試題數	前	中	後
1	24	區塊01 (M01)	區塊02 (M02)	區塊05 (M05)
2	24	區塊02 (M02)	區塊03 (M03)	區塊06 (M06)
3	24	區塊03 (M03)	區塊04 (M04)	區塊07 (M07)
4	24	區塊04 (M04)	區塊05 (M05)	區塊08 (M08)
5	24	區塊05 (M05)	區塊06 (M06)	區塊09 (M09)
6	24	區塊06 (M06)	區塊07 (M07)	區塊10 (M10)
7	24	區塊07 (M07)	區塊08 (M08)	區塊11 (M11)
8	23	區塊08 (M08)	區塊09 (M09)	區塊12 (M12)
9	24	區塊09 (M09)	區塊10 (M10)	區塊13 (M13)
10	24	區塊10 (M10)	區塊11 (M11)	區塊01 (M01)
11	23	區塊11 (M11)	區塊12 (M12)	區塊02 (M02)
12	23	區塊12 (M12)	區塊13 (M13)	區塊03 (M03)
13	24	區塊13 (M13)	區塊01 (M01)	區塊04 (M04)

本研究的研究對象為接受2006年TASA國小六年級數學科學習成就評量測驗，共7957名學生之作答資料，其中，原住民學生共計有294名，而非原住民學生共計有7663名。

二、研究工具

為達到本研究的目的，本研究採用的資料分析工具為統計分析軟體 SPSS for Windows 12.0、試題分析軟體 BILOG-MG3.0、EZDIF軟體與DIFPACK軟體，藉此進行實徵資料之分析、檢測及考驗。

三、資料處理與分析

(一) IRT模式參數分析

由於2006TASA小六數學所提供的資料並未包含各試題的IRT參數，所以本研究以BILOG-MG3.0 分析軟體進行IRT模式參數分析，利用此軟體計算出2006年TASA國小六年級數學科學習成就評量的103題試題之每一題之難度參數、鑑別度參數與猜測度參數，以及7957名受測學生每一名學生之學生能力值，以分析評量中各試題的適切性與做為試題DIF偵測前學生能力配組的依據。

（二）受測原住民與非原住民學生學習成就之差異情況

根據已計算出的所有受測學生能力值之資料，利用SPSS之描述性統計與獨立樣本t檢定，檢測受測原住民學生（294名）與非原住民學生（7663名）的學生能力值之顯著差異情況，但因二族群學生人數相差太多，所以需再以效果量（Effect Size）的大小判斷檢測結果實際顯著性（Practical significance）之程度。藉此檢定結果，來進一步瞭解二族群國小六年級學生於2006年TASA小六數學科成就評量中學習成就之差異情況與差異顯著性。

（三）受測原住民與非原住民學生之能力配組

由於實徵資料中每一試題區塊之小六原住民學生與小六非原住民學生的人數相差太多，且為了避免兩受試群體未經任何的配組程序，而導致兩群體表現有所差異之impact的DIF分析結果發生，所以於進行DIF檢定前，宜針對受測之焦點組與參照組學生進行能力的配組，而在本研究中，所設定的焦點組是研究者所要探討的「國小六年級原住民學生」，參照組則是「國小六年級非原住民學生」。

本研究所採用的DIF檢測法皆為非IRT取向的方法，而非IRT取向的DIF檢測程序是以觀察分數的總分為配組變項，所以於進行DIF檢測之前，通常應是以測驗總分為能力配組的依據，但因TASA的設計是採用平衡不完全區塊（BIB）設計題本模式，且每位受測學生作答的題本皆為3個不同的試題區塊所組成，因此，在各個試題區塊中，得分高的受測學生並不一定就代表其能

力也是高的。所以，為了符合DIF分析必須要建立在二組能力或表現相比（comparable）的群體之原則，本研究以受測學生之IRT能力值作為能力配組的變項。且於簡茂發、劉湘川、許天維、郭伯臣和殷志文（1995）的研究結果也顯示，在進行MH法檢定時，以IRT能力值為分組準則的DIF分析結果較傳統總分下的分組準則稍微穩定。

因此，本研究是以受測學生之IRT能力值為分組依據，於進行DIF檢測分析前，分別依序進行以下學生能力配組之動作：

- 1、分別將每個試題區塊內的原住民學生能力值（ θ ），依據能力值（ θ 介於+3與-3之間），每間隔0.5為一組，平均分成12組，由於每一試題區塊的原住民學生人數並不多，為了避免抽樣的人數再減少，所以第12組的能力值範圍變更為 $-2.5 < \theta > -3.5$ 。能力值組別之詳細內容如表2所示：

表2 學生能力值（ θ ）之組別

組別	學生能力值範圍	組別	學生能力值範圍
1	+3.0 $\theta > +2.5$	7	0 $\theta > -0.5$
2	+2.5 $\theta > +2.0$	8	-0.5 $\theta > -1.0$
3	+2.0 $\theta > +1.5$	9	-1.0 $\theta > -1.5$
4	+1.5 $\theta > +1.0$	10	-1.5 $\theta > -2.0$
5	+1.0 $\theta > +0.5$	11	-2.0 $\theta > -2.5$
6	+0.5 $\theta > 0$	12	-2.5 $\theta > -3.5$

- 2、在每一試題區塊中，分別根據每一組之原住民學生的能力值範圍與人數，從相同能力值範圍的非原住民學生中，利用SPSS for Window 12.0，以隨機抽樣的方式，依序在12組能力值中選取與原住民學生相同的人數，然後予以整理成本研究所需的13個試題區塊學生作答資料。

(四) 原住民與非原住民學生之DIF分析

由於任何DIF檢測法的統計考驗力與正確DIF偵測率都受樣本人數的影響，而非IRT取向之DIF程序，如Mantel-Haenszal（以下簡稱MH）、Logistic Regression（以下簡稱LR）與Simultaneous Item Bias Test（以下簡稱SIBTEST），各組人數大約200至300人，即可以得到適當的結果（盧雪梅，1999）。

且林坤昌（1998）於研究中提到三種非IRT取向DIF檢測法的優缺點，MH檢定法的優點為計算容易、樣本不需很大與型一誤差較小，缺點是對於non-uniform DIF試題的檢定效果較差；LR檢定法的優點則是能檢驗uniform與non-uniform DIF試題，缺點是當資料為三參數模式時，型一誤差率會變大；而SIBTEST與Crossing SIBTEST的優點是能檢驗uniform與non-uniform DIF試題。所以，DIF分析應用於實徵資料，如能同時以不同檢測法進行DIF偵測，其分析結果取其交集，更能確保DIF試題判定的可靠性（林奕宏、林世華，2004）。

根據上述之理由，因本研究所採用的實徵資料，焦點組(國小六年級原住民學生)受測人數只有294位，且各個試題區塊之焦點組學生人數也分別只有72~99位，在樣本人數很少的情形之下，本研究決定採用非IRT取向之MH、LR與Crossing SIBTEST分析法，進行2006TASA國小六年級數學科成就測驗中之原住民與非原住民學生的DIF試題檢測。

完成各試題區塊的二族群學生能力配組後，以每一試題區塊為單位，分別以EZDIF（MH-2、LR）與DIFPACK（Crossing SIBTEST）軟體，逐一檢驗每一試題區塊中各個試題之DIF情形。本研究之DIF試題為三種DIF分析結果取其交集，當某一試題同時被二種以上DIF檢測方法測出有DIF現象時，本研究即將該試題列為DIF試題。但因MH-2對於偵測non-uniform DIF題的偵測力較弱，因此為避免non-uniform DIF題的被排除，所以本研究將分別針對Crossing SIBTEST所偵測到的DIF題，檢視二族群學生的答對率折線圖交叉情形，如判定後為non-uniform DIF試題，則列入本研究所欲探討的DIF試題。

(五) DIF試題對二族群受測學生數學學習成就之影響

針對存在有DIF試題的試題區塊，將偵測到的DIF試題之該題學生作答資料分別從該試題區塊中刪除，於DIF試題刪除後，進行SPSS軟體的獨立樣本t檢定，檢驗具有DIF試題的試題區塊中，經能力配組之小六原住民和非原住民學生的數學學習成就之顯著差異情況，與該試題區塊二族群全部受測學生的數學學習成就之顯著差異情況；同時，再分別與該試題區塊DIF試題刪除前的測驗平均分數獨立樣本t檢定顯著差異情況做比較，藉以瞭解DIF試題對於二族群之受測學生數學學習成就的影響性。

參、研究結果與討論

一、IRT模式參數分析

(一) 試題參數分析

以BILOG-MG3.0 分析軟體，針對2006年TASA國小六年級數學科學習成就評量之103題試題進行IRT模式之試題參數分析，其分析結果如表4所示：

1、難度參數分析

針對全部103題試題，各試題之難度參數介於-2.02~3.12，平均難度為0.00，標準差1.02。而試題的難度參數與測驗的效率有關，難度適當的試題是構成優良測驗的必要條件，IRT難度參數值介於-2至2之間是較為適當的。所以，全部試題屬於難度適中的試題，唯試題M03-6之 $b = 2.49$ 、M10-5之 $b = 3.12$ 、M13-2之 $b = 2.35$ ， b 值皆大於2，尤其M10-5之 b 值更大於3，故此4題屬難度較高之試題。而M12-7之 $b = -2.02$ ，則為太過於簡單之試題。

2、鑑別度參數分析

各試題的鑑別度參數介於0.21~2.56，平均鑑別度為0.85，標準差0.34。根據美國測驗學者Baker（2001）之鑑別度判斷標準，茲將檢定結果等級分類分析整理成表3。由分析結果可知，除試題M01-4之鑑別度參數為0.25，可能需大幅度修改或刪除之外，其餘102題試題皆具有良好的鑑別度。全部試題屬於具有中、高鑑別度的試題。

表3 103題之試題鑑別度檢定結果等級分類

鑑別度參數值	鑑別度等級	題數	百分比 (%)
$0.01 \leq a < 0.35$	需大幅度修改或刪除	1	1
$0.35 \leq a < 0.65$	尚可但需部份修改	29	28
$0.65 \leq a < 1.35$	優良但需小幅度修改	67	65
$1.35 \leq a$	優良	6	6

3、猜測度參數分析

各試題猜測度參數介於0.11~0.27，平均猜測度為0.18，標準差0.03，平均而言，不至於過高。但試題M11-2的猜測度則高於選項數的倒數0.25，較不理想，其他試題的猜測度皆小於0.25，符合猜測度參數估計值的一般範圍。

表4 2006年TASA小六數學103題IRT模式之試題參數

題號	鑑別度	難度	猜測度	題號	鑑別度	難度	猜測度
M01-1	1.27	-1.26	0.18	M07-5	0.69	-1.03	0.17
M01-2	0.74	0.09	0.17	M07-6	0.89	-1.21	0.17
M01-3	0.54	-0.72	0.16	M07-7	0.86	-0.56	0.18
M01-4	0.25	1.69	0.19	M07-8	0.98	-1.09	0.21
M01-5	0.43	-1.62	0.18	M08-1	1.12	0.88	0.13
M01-6	0.56	-0.24	0.20	M08-2	0.82	-0.46	0.18
M01-7	0.98	-1.49	0.22	M08-3	0.58	1.34	0.19
M01-8	0.93	-1.06	0.16	M08-4	1.42	0.55	0.16
M02-1	0.56	1.28	0.19	M08-5	1.04	0.78	0.16
M02-2	1.12	0.64	0.17	M08-6	0.78	0.55	0.16
M02-3	0.82	-0.10	0.18	M08-7	0.88	0.04	0.19
M02-4	0.66	-0.67	0.17	M08-8	1.12	0.43	0.18
M02-5	0.93	-0.61	0.14	M09-1	0.37	0.28	0.18
M02-6	0.69	-1.83	0.18	M09-2	0.64	-0.56	0.20

M02-7	1.10	0.165	0.20	M09-3	0.69	0.55	0.18
M02-8	1.37	0.39	0.18	M09-4	0.75	-0.89	0.17
M03-1	0.38	0.17	0.20	M09-5	0.74	-0.34	0.15
M03-2	0.49	-1.28	0.18	M09-6	0.67	1.44	0.18
M03-3	0.84	0.37	0.21	M09-7	1.21	-0.09	0.21
M03-4	0.60	-0.94	0.19	M09-8	1.28	0.04	0.14
M03-5	0.70	-0.20	0.22	M10-1	1.44	0.21	0.13
M03-6	0.42	2.49	0.19	M10-2	0.95	1.28	0.15
M03-7	0.81	-0.57	0.21	M10-3	2.56	1.23	0.10
M03-8	1.03	1.19	0.15	M10-4	0.84	-0.79	0.18
M04-1	0.56	0.92	0.16	M10-5	0.21	3.12	0.19
M04-2	1.06	0.91	0.18	M10-6	1.03	0.22	0.14
M04-3	1.44	0.73	0.15	M10-7	1.00	-0.72	0.19
M04-4	1.46	1.52	0.14	M10-8	0.49	-0.12	0.19
M04-5	0.51	-0.36	0.23	M11-1	0.82	0.18	0.19
M04-6	1.07	-0.40	0.20	M11-2	1.18	-0.40	0.27
M04-7	0.58	0.74	0.17	M11-3	1.55	-0.21	0.15
M04-8	0.66	0.28	0.19	M11-4	1.42	1.17	0.13
M05-1	1.09	-0.15	0.17	M11-5	0.71	-0.11	0.17
M05-2	1.14	0.52	0.12	M11-6	0.67	-0.15	0.16
M05-3	0.58	-1.72	0.19	M11-7	0.83	-0.98	0.20
M05-4	0.52	0.80	0.18	M11-8	0.65	-1.57	0.18
M05-5	0.63	-0.72	0.18	M12-1	0.89	-0.51	0.20
M05-6	1.07	0.31	0.18	M12-3	1.03	-0.84	0.19
M05-7	0.57	-1.97	0.18	M12-4	1.25	0.21	0.17
M05-8	0.80	0.79	0.13	M12-5	0.61	1.31	0.20
M06-1	0.94	0.86	0.16	M12-6	0.86	-0.04	0.20
M06-2	0.66	-0.73	0.16	M12-7	0.55	-2.02	0.18
M06-3	0.82	0.50	0.20	M12-8	0.75	0.53	0.18
M06-4	0.89	-0.09	0.17	M13-1	0.37	-0.15	0.19
M06-5	1.11	-0.43	0.15	M13-2	0.43	2.35	0.19

M06-6	0.95	-0.70	0.11	M13-3	0.57	-1.65	0.20
M06-7	1.29	-0.40	0.19	M13-4	0.52	0.02	0.18
M06-8	0.63	0.39	0.19	M13-5	1.26	0.37	0.15
M07-1	0.79	-0.55	0.19	M13-6	0.42	1.84	0.19
M07-2	0.69	-1.17	0.16	M13-7	0.98	0.94	0.19
M07-3	1.11	-1.41	0.20	M13-8	0.50	1.92	0.19
M07-4	0.55	-1.85	0.16	平均	0.85	0.00	0.18
				標準差	0.34	1.02	0.03

(二) 學生能力值參數分析

在IRT模式之學生能力值（abilities）參數分析方面，其分析結果經SPSS統計整理之後，如表5所示。全部7957名國小六年級受測之學生的學生能力值介於3.12至-3.04，平均能力值為-0.03，標準差為1.08；且由表6可知，其中294名受測之原住民學生的學生能力值介於2.88至-3.02，平均能力值為-0.68，標準差為1.09；7663名受測之非原住民學生的學生能力值介於3.12至-3.04，平均能力值為-0.01，標準差為1.07。

表5 原住民與非原住民學生之學生能力值（ θ ）統計量數

是否為原住民	個數	總N的百分比	平均數	平均數的標準誤	標準差	最大值	最小值
原住民	294	3.7%	-0.68	.06	1.09	2.88	-3.02
非原住民	7663	96.3%	-0.01	.01	1.07	3.12	-3.04
總和	7957	100.0%	-0.03	.01	1.08	3.12	-3.04

二、原住民與非原住民學生學習成就之差異

(一) 描述性統計

在進行兩族群學生的學生能力值之獨立樣本t檢定之前，本研究擬先進行兩族群學生的能力估計值之描述性統計，以得知其數學學習成就的差異情況，其分析結果如表6所示。從表中之資料可發現，非原住民學生之平均學生能力值 $=-.01$ ，較原住民學生之平均學生能力值 $=-.68$ 為佳。其中，原住民學生能力值之偏態（skewness）係數 $=.68 > 0$ ，為正偏態（右偏態）分配，且偏態係數 > 0.5 ，代表該樣本資料分配出現大的偏態，且較多學生的分數集中在低分部分，如圖1所示；而非原住民學生能力值之偏態係數 $=-.06 < 0$ ，為負偏態（左偏態）分配，但幾近常態分布，如圖2。所以，由描述性統計結果可得知，非原住民高能力值的學生人數比率是高於原住民高能力值的學生人數比率，即非原住民學生的數學學習成就表現較原住民學生學習成就表現為佳。但此差異是否達到顯著差異，就必須再進行兩受測族群學生的學生能力值之獨立樣本t檢定。

表6 受測原住民與非原住民學生數學學習成就之描述性統計

是否為原住民	個數	平均數	標準差	中位數	偏態	峰度
原住民	294	-.68	1.09	-.79	.68	.24
非原住民	7663	-.01	1.07	.01	-.06	-.38

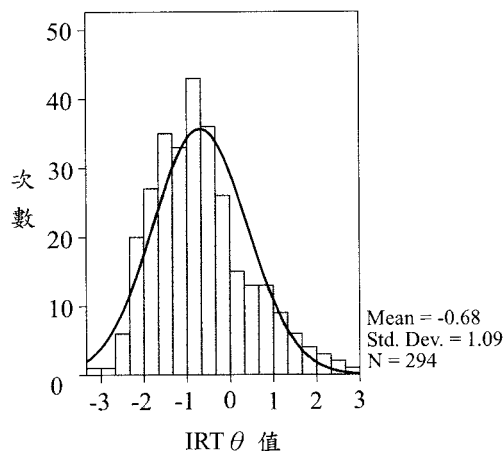


圖1 原住民學生能力值（ θ ）之次數分配曲線圖

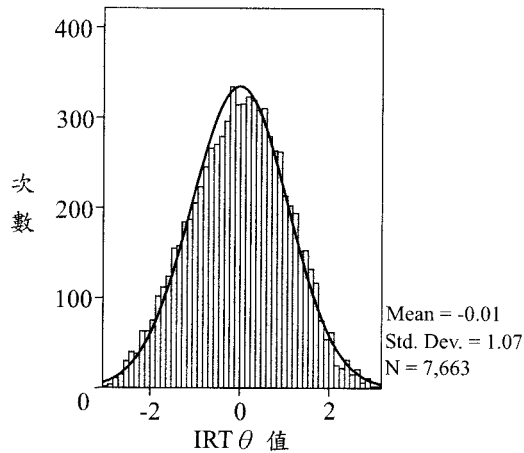


圖2 非原住民學生能力值（ θ ）之次數分配曲線圖

（二）獨立樣本t檢定

兩受測族群學生的學生能力值之獨立樣本t檢定結果，如表7所示。從表中資料可知，變異數相等的 Levene 檢定之 $F=.08$ 、 $p=.77$ ，因為 p 值 $>.05$ ，所以應以「平均數相等的t檢定」之「假設變異數相等」的資料為準， $p=.00 <.05$ ，因此受測原住民學生與非原住民學生之學生能力值達到統計上的顯著差異；接著，計算其效果量（Effect Size）的大小，計算結果得知， $Effect\ Size=0.63$ ，根據Cohen的標準，介於0.5至0.8表示實際顯著性為中至高等，故此顯著差異程度代表兩群受測學生之能力值差異頗大。

所以，從上述之檢定結果可知，兩受測族群學生在2006年TASA國小六年級數學科評量的學習成就表現，存在著顯著的差異，且為國小六年級非原住民學生之數學學習成就表現優於原住民學生之學習成就表現。

表7 兩受測族群學生的學生能力值（ θ ）之獨立樣本t檢定結果

自變項	個數	平均數（標準差）	t	p
小六原住民學生	294	-0.68	-10.57	.00
小六非原住民學生	7663	-0.01		

$\alpha = .05$

三、DIF試題之偵測結果

完成2006年TASA國小六年級數學科成就評量之13個試題區塊二族群學生的能力配組後，分別以EZDIF（MH-2、LR）與DIFPACK（Crossing SIBTEST）軟體，依序檢驗經能力配組之試題區塊M01至區塊M13中的每一試題是否存在有DIF現象。以下就針對本研究偵測的結果予以討論。

（一）MH-2法DIF檢測之結果

利用EZDIF軟體之MH-2進行DIF試題偵測，在設定 $\alpha = .05$ 的標準下，所偵測到具有DIF現象的試題共有5題，表8為ETS分類法（MH-2）DIF檢測之結果統計表，並分述如下：

- 1、試題M03-6的ETS效果量編碼為B，檢定達.05顯著水準，顯示該題呈現中度的DIF現象；且該題的 $\alpha < 1.0$ ，MH D-DIF > 0.0 ，表示該試題對表示該試題對焦點組（原住民學生）較為有利。
- 2、試題M06-6的ETS效果量編碼為B，檢定達.05顯著水準，顯示該題呈現中度的DIF現象；且該題的 $\alpha > 1.0$ ，MH D-DIF < 0.0 ，表示該試題對參照組（非原住民學生）是較為有利的。
- 3、試題M06-7的ETS效果量編碼為CR，檢定達.05顯著水準，顯示該題呈現重度的DIF現象；且該題的 $\alpha > 1.0$ ，MH D-DIF < 0.0 ，表示該試題對參照組（非原住民學生）是較為有利的。
- 4、試題M11-3的ETS效果量編碼為B，檢定達.05顯著水準，顯示該題呈現中度的DIF現象；且該題的 $\alpha < 1.0$ ，MH D-DIF > 0.0 ，表示該試題對焦點組（原住民學生）較為有利。
- 5、試題M11-6的ETS效果量編碼為B，檢定達.05顯著水準，顯示該題呈現中度的DIF現象；且該題的 $\alpha > 1.0$ ，MH D-DIF < 0.0 ，表示該試題對參照組（非原住民學生）是較為有利的。

綜合以上偵測結果，存在DIF現象的試題共計有5題，分別為試題M03-6、M06-6、M06-7、M11-3與M11-6，其中有3題試題是對參照組（非原住民學生）有利，而有2題試題則為對焦點組（原住民學生）有利。

表8 MH-2法所偵測到之DIF試題 $\alpha = .05$

ETS標準	題號	α	X^2	P-Value	MH D-DIF
B	M03-6*	0.38	4.52	0.03	2.30
B	M06-6*	3.85	4.12	0.04	-3.17
CR	M06-7***	6.55	8.71	0.00	-4.42
B	M11-3***	0.30	8.04	0.00	2.80
B	M11-6*	2.79	4.81	0.03	-2.41

***p<.001 *P<.05

(二) LR法DIF分析之結果

利用EZDIF軟體之LR法，在設定 $\alpha = .05$ 的標準下，所偵測到具有DIF現象的試題共計有10題，如表9所示，分別為試題M01-7、M06-7、M08-3、M10-3、M10-6、M11-2、M11-3、M11-5、M11-6以及M12-1。其中，試題M01-7、M11-3、M11-6屬於uniform，而試題M06-7、M08-3、M10-3、M10-6、M11-2、M11-5與M12-1則屬於non-uniform。

表9 LR法所偵測到之DIF試題 $\alpha = .05$

Item	Group		Trait x Group	
	Estimate	p-value	Estimate	p-value
M01-7	2.12	0.04	-0.35	0.15
M06-7	-1.85	0.12	1.06	0.01
M08-3	-1.66	0.04	0.52	0.02
M10-3	3.27	0.05	-0.75	0.03
M10-6	4.45	0.01	-1.05	0.01
M11-2	-1.96	0.04	0.56	0.02
M11-3	-2.80	0.02	0.46	0.09
M11-5	2.32	0.03	-0.52	0.04
M11-6	2.60	0.04	-0.44	0.14
M12-1	2.29	0.03	-0.68	0.03

(三) Crossing SIBTEST法DIF分析之結果

利用DIFPACK軟體之Crossing SIBTEST法，在設定 $\alpha = .05$ 的標準下，所偵測到具有DIF現象的試題共計有4題，分別為試題M02-8、M04-6、M08-5與M11-6，且皆屬於non-uniform DIF，表10為Crossing SIBTEST法所偵測到之DIF試題統計表。

表10 Crossing SIBTEST法所偵測到之DIF試題 $\alpha = .05$

Item Number	Beta -cro	Magnitude of Beta below Crossing crossing	%	Magnitude of Beta above Crossing point	%	p-value
M02-8	0.07	0.03 F	47	0.04 R	53	0.00
M04-6	0.15	0.00	0	0.15 R	100	0.00
M08-5	-0.08	0.00	0	0.08 F	100	0.00
M11-6	0.15	0.00	0	0.15 R	100	0.00

(四) 2006年TASA小六數學科成就評量之DIF試題

依據上述三種DIF分析法在每一試題區塊所偵測到的DIF試題，分別將偵測結果整理成表11。共有15題試題分別被不同DIF檢測法檢測出存在有DIF現象，其中，LR所檢測到的DIF試題就有10題，且有7題是其他二種檢測法都沒有偵測到的DIF試題，原因可能是因本研究之實徵資料所採用的是三參數IRT模式分析，所以導致分析結果的第一類型誤差率升高（林坤昌，1998）。

綜合統計表之三種DIF分析法的偵測結果，在15題試題中，試題M06-7與M11-3因同時有二種DIF分析法偵測到DIF現象，且試題M11-6同時更有三種DIF分析法偵測到DIF現象，所以將試題M06-7、M11-3與M11-6列入本研究所要探討之DIF試題。

表11 三種DIF分析法偵測到之DIF試題統計表 $\alpha = .05$

	MH-2	LR	Crossing SIBTEST
01-7		DIF	
M02-8			DIF
M03-6	DIF		
M04-6			DIF
M06-6	DIF		
M06-7	DIF	DIF	
M08-3		DIF	
M08-5			DIF
M10-3		DIF	
M10-6		DIF	
M11-2		DIF	
M11-3	DIF	DIF	
M11-5		DIF	
M11-6	DIF	DIF	DIF
M12-1		DIF	

但由於MH-2對於non-uniform DIF試題的偵測力較弱，而Crossing SIBTEST法所偵測到的 DIF題，僅試題M11-6有列入本研究之DIF試題，為了避免non-uniform DIF的被排除，所以本研究將檢視Crossing SIBTEST法所偵測到的 DIF試題M02-8、M04-6與M08-5之二族群學生答對率折線圖，確認其是否為non-uniform DIF試題，以決定是否列入DIF試題。檢視結果如下：

1、試題M02-8

圖3為試題M02-8的二族群學生答對率折線圖，由圖可知，試題M02-8為 non-uniform DIF，為有利於非原住民學生作答之試題。

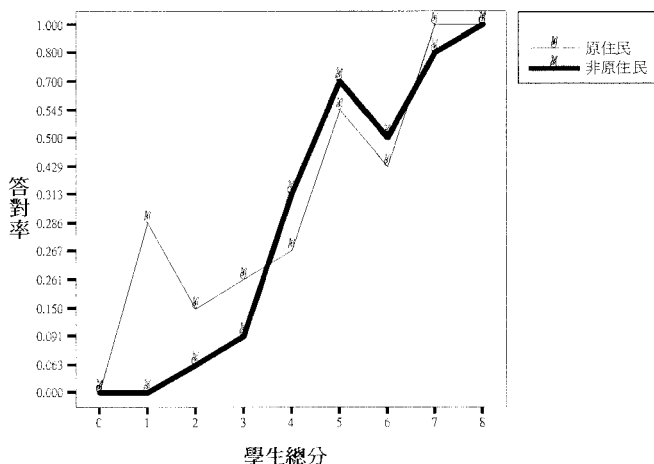


圖3 試題M02-8的二族群學生答對率折線圖

2、試題M04-6

圖4為試題M04-6的二族群學生答對率折線圖，由圖可知，試題M04-6為 non-uniform DIF，為有利於非原住民學生作答之試題。

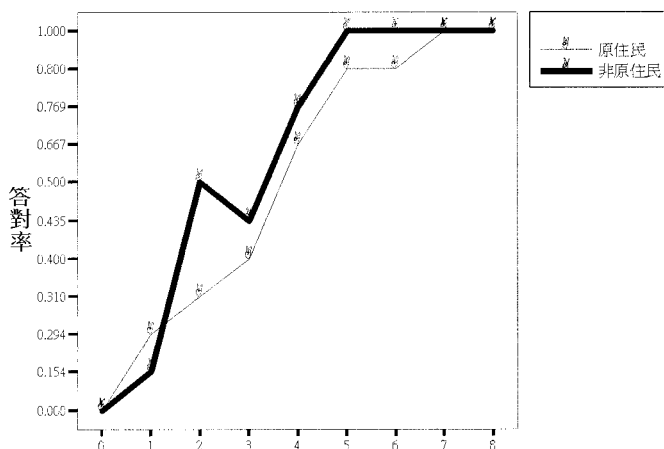


圖4 試題M04-6的二族群學生答對率折線圖

3、試題M08-5

圖5為試題M08-5的二族群學生答對率折線圖，由圖可知，試題M08-5為non-uniform DIF，為有利於原住民學生作答之試題。

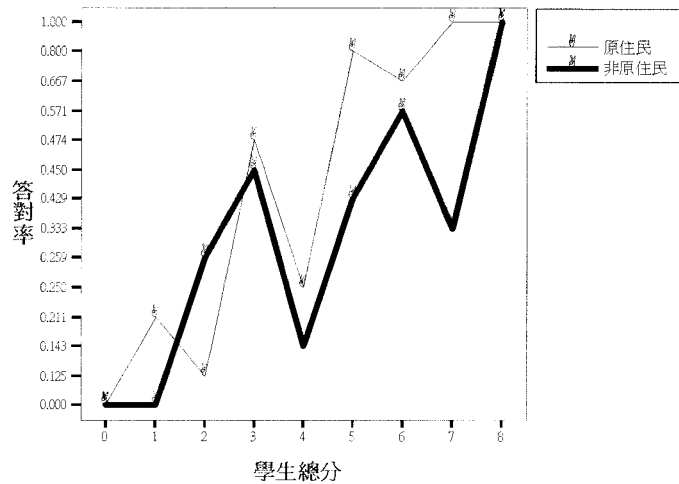


圖5 試題M08-5的二族群學生答對率折線圖

檢視試題M02-8、M04-6與M08-5之二族群學生答對率折線圖後可得知，由Crossing SIBTEST所偵測到的三題試題皆為non-uniform DIF試題，所以將此三題列入本研究所要探討分析的DIF試題。

綜合以上DIF偵測結果，本研究的DIF試題共計有6題，分別為試題M02-8、M04-6、M06-7、M08-5、M11-3與M11-6。

四、DIF試題對受測學生學習成就的影響

本研究為了瞭解DIF試題對受測學生的數學學習成就是否產生影響，於是針對五個有出現DIF試題的試題區塊（M02、M04、M06、M08、M11）中之二族群受測學生的平均分數分別進行比較。

（一）試題區塊中經過能力配組的學生

首先，先分別針對具有DIF題之試題區塊中經過能力配組的二族群學生進行DIF試題刪除前與刪除後的平均分數獨立樣本t檢定，比較結果如表13所示。從表12可得知，經過能力配組的原住民與非原住民學生在此五個試題區塊之DIF題刪除前與刪除後的學生平均分數均沒有足夠的證據可以說明兩者之間是有顯著的差異，但在試題區塊M06與M11中，原住民與非原住民學生的得分表現卻有些微的變化。

於試題區塊M06中，經過能力配組的原住民與非原住民學生在該試題區塊之DIF題刪除前與刪除後的數學學習成就表現雖沒有足夠的證據可以說明兩者之間是有顯著的差異，但從表12中的t值，我們可發現，原住民學生原本較低的得分平均數，於刪除DIF試題後，反而變得比非原住民學生的平均得分還要高，此一情形亦符合ETS的檢定結果，即M06-7為對小六非原住民學生作答有利的DIF試題，所以在刪除此試題後，減少了對原住民學生不公平的影響，使得原住民學生的數學學習成就表現高於非原住民學生的表現。

而在試題區塊M11中，經過能力配組的原住民與非原住民學生在該試題區塊之DIF題刪除前與刪除後的數學學習成就表現雖也沒有足夠的證據可以說明兩者之間是有顯著的差異，且從t值可發現，原住民學生於DIF試題刪除前與刪除後的平均得分分別與非原住民學生的平均得分都差距不大，原因可能在於ETS的結果為M11-3於ETS是對原住民學生的作答有利，而M11-6則是對非原住民學生較為有利，所以互相抵銷其對二族群學生作答的不公平性。

表12 DIF試題刪除前後經能力配組學生的平均分數之比較 $\alpha = .05$

試題 區塊	自變項	DIF試題刪除前 平均數相等的t檢定			DIF試題刪除後 平均數相等的t檢定		
		平均數 (標準差)	t	p	平均數 (標準差)	t	p
M02	原住民 學生	3.51 (1.82)	-.27	.79	3.18 (1.68)	-.48	.63
	非原住 民學生	3.58 (1.86)			3.30 (1.63)		

M04	原住民 學生	2.83 (1.91)	-1.32	.19	2.39 (1.74)	-.98	.33
	非原住 民學生	3.20 (2.00)			2.64 (1.77)		
M06	原住民 學生	3.11 (1.93)	-0.56	.58	2.76 (1.76)	0.22	.83
	非原住 民學生	3.28 (1.89)			2.70 (1.62)		
M08	原住民 學生	2.68 (1.82)	-.43	.67	2.34 (1.64)	-.71	.48
	非原住 民學生	2.80 (1.78)			2.52 (1.70)		
M11	原住民 學生	3.95 (2.00)	1.11	.27	3.03 (1.50)	0.99	.32
	非原住 民學生	3.63 (1.93)			2.82 (1.47)		

(二) 試題區塊中全部受測學生

分別再針對具有DIF題之試題區塊中的二族群全部受測學生進行DIF試題刪除前與刪除後的平均分數獨立樣本t檢定，檢定結果如表13所示。從表中數據可發現，於此五個試題區塊中，各試題區塊全部的原住民與非原住民學生在DIF題刪除前與刪除後的數學學習成就表現均在統計上呈現顯著的差異，而且從各Effect Size可知，其顯著差異程度屬中等，甚至很大，且在六個試題區塊中，皆為非原住民學生於DIF試題刪除前與刪除後的數學學習成就表現明顯都高於原住民學生。

所以，從具有DIF題之五個試題區塊中的二族群全部受測學生平均得分獨立樣本t檢定的結果可得知，小六原住民與非原住民學生於此五個試題區塊之數學學習成就表現皆存在著顯著差異的情形，並不會因DIF試題的存在與否，而有所影響。

但從表中之t值可發現，在試題區塊M02、M04、M06與M11中，DIF題刪除前學習成就明顯較低的原住民學生，在分別刪除DIF題後，其學習成就稍有拉近了與非原住民學生學習成就表現的差距程度。這也顯示了DIF試題對

於原住民學生的作答，確實存在著不公平的影響。

表13 DIF試題刪除前後受測二族群學生的平均分數之比較 $\alpha = .05$

試題 區塊	自變項	DIF試題刪除前 平均數相等的t檢定			DIF試題刪除後 平均數相等的t檢定		
		平均數 (標準差)	t	p	平均數 (標準差)	t	p
M02	原住民 學生	3.51 (1.82)	-6.48	.00	3.17 (1.68)	-5.87	.00
	非原住 民學生	4.78 (2.02)			4.28 (1.76)		
	效果量 (Effect Size)		0.63		0.63		
M04	原住民 學生	2.83 (1.91)	-4.79	.00	2.39 (1.74)	-3.98	.00
	非原住 民學生	3.84 (1.98)			3.15 (1.78)		
	效果量 (Effect Size)		0.51		0.42		
M06	原住民 學生	3.11 (1.93)	-7.11	.00	2.76 (1.76)	-6.46	.00
	非原住 民學生	4.91 (2.11)			4.20 (1.86)		
	效果量 (Effect Size)		0.86		0.78		
M08	原住民 學生	2.68 (1.82)	-5.78	.00	2.34 (1.63)	-6.06	.00
	非原住 民學生	3.88 (2.14)			3.47 (1.91)		
	效果量 (Effect Size)		0.56		0.60		
M11	原住民 學生	3.95 (2.00)	-5.39	.00	3.03 (1.50)	-5.06	.00
	非原住 民學生	5.06 (2.00)			3.81 (1.49)		
	效果量 (Effect Size)		0.55		0.52		

肆、結論與建議

一、結論

依據研究結果與討論，將本研究之結論整理如下所述：

(一) 試題參數分析

本研究針對2006年TASA國小六年級數學科學習成就評量中103題試題之學生作答資料進行IRT模式參數分析，以檢核評量中各試題的適切性與做為試題DIF偵測前學生能力配組的依據。分析結果顯示，103題試題的鑑別度參數介於0.214~2.567，平均鑑別度為0.852；難度參數介於-2.025~3.126，平均難度為-0.002；猜測度介於0.112~0.272，平均猜測度為0.181。其中，試題M03-6、M10-5、M13-2與M10-5等4題屬難度較高之試題，而試題M12-7則屬於過於簡單之試題；試題M01-4之鑑別度小於0.35，需大幅度修改或刪除；試題M11-2的猜測度則高於0.25，較不理想。整份測驗平均而言，屬於難度適中、猜測值不至於過高，且具有良好鑑別度的測驗。

(二) 原住民與非原住民學生能力值參數分析與學習成就之差異

以IRT模式參數分析，針對測驗的全部7957名受測小六學生之作答實徵資料，進行學生能力值之分析，分析結果顯示，其中294名受測之小六原住民學生的學生能力值介於2.8828至-3.0270，平均能力值為-0.6888；而7663位受測之小六非原住民學生的學生能力值介於3.1218至-3.0479，平均能力值為-0.0128，從數據可得知，非原住民學生的學生能力值高於原住民學生的學生能力值；且在二族群學生的測驗成績之描述性統計分析資料中，統計結果也發現非原住民高能力值之學生人數比率是高於原住民高能力值之學生人數比率，並經由進行兩受測族群學生的學生能力值之獨立樣本t檢定與Effect Size，得知二族群學生的學生能力值存在著顯著的差異，而且顯著差異頗大。

因此，綜合以上研究結果，就針對二族群學生於2006年TASA國小六年級數學科學習成就評量的平均能力值來做比較，可得結論：小六非原住民學生的數學學習成就是高於原住民學生的學習成就。

（三）2006年TASA小六數學科成就評量之DIF試題分析

由於實徵資料中，二族群學生人數相差太多，為了避免因能力未經任何配組程序的impact 分析結果發生，於進行DIF檢測前，先分別於各個試題區塊中，將原住民與非原民學生依據「IRT模式學生能力值參數」進行隨機抽樣與能力配組的程序，完成能力配組後，再以MH-2、LR與Crossing SIBTEST之DIF檢定程序同時於每一試題區塊進行DIF試題的檢測，偵測小六原住民學生與非原住民學生二族群於成就評量中的DIF現象。檢測結果顯示，在設定 $\alpha = .05$ 的標準下，103題試題中，共有15題呈現DIF，但其中只有三題試題，分別為M06-7、M11-3與M11-6同時被二種以上的DIF檢定法偵測到DIF現象，故本研究將此三題列入DIF題；但為了補強MH-2只能偵測uniform DIF的缺點，避免non-uniform DIF的被排除，所以本研究亦將Crossing SIBTEST法所偵測到的DIF試題M02-8、M04-6與M08-5列入DIF試題。

經資料整理後，本研究所偵測到的DIF試題總計有六題，分別為M02-8、M04-6、M06-7、M08-5、M11-3與M11-6，約占全部試題103題的6%。其中，根據ETS分類法（MH-2）分析之數據顯示，試題M06-7之ETS效果量編碼為CR，顯示該試題呈現重度的DIF現象，且該試題對參照組（非原住民學生）的作答情況是較為有利的；而試題M11-6的ETS效果量編碼為B，顯示該題呈現中度的DIF現象，且該試題對參照組（非原住民學生）的作答也是較為有利的；試題M11-3的ETS效果量編碼為B，顯示該題呈現中度的DIF現象，但分析數據顯示，該試題對焦點組（原住民學生）的作答卻是較為有利。

由於這六題DIF試題並未出現在公告版，所以無法進一步分析是否屬於試題偏誤。所以，本研究的研究數據與結果，可供TASA與未來的研究做為參考。

(四) DIF試題對學習成就的影響

1、試題區塊中經過能力配組的學生

於存在DIF試題（M02-8、M04-6、M06-7、M08-5、M11-3與M11-6）的試題區塊M02、M04、M06、M08與M11中，經能力配組的小六原住民與非原住民學生在DIF試題刪除前與刪除後之數學學習成就，統計上雖沒有呈現顯著的差異，但還是存有些微的變化。其中，試題區塊M06中，原住民學生原本較低的學習成就，於刪除DIF試題M06-7後，反而變得比非原住民學生的學習成就還要高，可知DIF試題M06-7對於原住民學生的作答有較為不公平的情形發生，也符合MH-2的檢定結果。

而在試題區塊M11中，原住民學生於DIF試題刪除前與刪除後的平均得分都些微的高於非原住民學生，其原因可能在於：依MH-2的檢測結果，在二題DIF試題分別是對不同族群學生較為有利的作用之下，因而互相抵消其對二族群學生作答的不公平性。

2、試題區塊中全部的學生

在試題區塊M02、M04、M06、M08與M11中，DIF試題刪除前與刪除後的所有受測小六原住民與非原住民學生之測驗平均分數獨立樣本t檢定的結果顯示，二群受測學生之測驗平均分數於DIF試題刪除前與刪除後，均呈現顯著的差異，即非原住民學生的測驗平均分數高於原住民學生的平均分數之情況，並未受到DIF試題存在與否的因素而受影響。而此研究結果有可能是以下二種原因所造成，一是由於在本研究之研究對象中，因原住民學生（294名）與非原住民學生（7663名）的人數相差太多，而為了取得較為正確精準的DIF試題檢測數據，所以在進行DIF試題之檢測前，已先經過學生能力的配組程序，經配組抽樣後，非原住民學生人數也為294名，與原本7663名原住民學生人數相差甚大，所以此項測驗平均分數比較的結果亦有可能是能力配組與抽樣之後所造成；而第二種原因有可能就是因DIF試題的相銷所造成，即經檢測的結果，有些DIF試題是對非原住民學生的作答較為有利，但也有一些試題卻是對原住民是有利的，所以在互相抵銷各DIF試題分別對二族群學

生的影響之後，二族群學生的成績表現於統計上依然呈現顯著的差距。

綜合以上DIF試題對測驗成績影響的結論可得知，在2006年TASA國小六年級數學科成就評量中，DIF試題對於原住民與非原住民學生學習成就之顯著差異情況，雖沒有絕對性與大幅度的影響，但DIF試題的存在，的確會對受測學生的作答產生不公平的情形，進而影響學習成就。但DIF試題並不代表該試題就必須予以刪除，而是應進一步做質的分析，討論是否需就針對教學內容或教學方式做改進，進而確實的對學生實施補救教學，達到教學的目的。

二、建議

在採用MH法進行DIF分析時，為了要可以符合期望值大於5的基本假設，TASA的每一試題區塊之二族群的學生樣本人數分別至少要大於 $(8+1) \times 2 \times 2 \times 5 = 180$ 人（余民寧、謝進昌，2006）。雖然可以減少分組組數的方法，來減少所需的學生樣本人數，但DIF之統計檢定力可能也會因樣本人數的減少而降低。所以，為了能更準確分析DIF試題對於小六原住民學生與非原住民學生在TASA測驗平均分數的影響，因此建議TASA可針對原住民學生增加隨機抽樣的樣本人數來進行施測，以達到更穩定的DIF分析結果與統計數據。

且在本研究的研究結果中提到，DIF試題的存在與否，對於受測國小六年級原住民學生與非原住民學生的數學學習成就表現並沒有顯著的影響；且在DIF試題刪除前與刪除後，非原住民學生的數學學習成就表現均高於原住民學生的學習成就，這個研究結果與之前一些針對原住民學生學習成就所做的相關研究結論類似，所以，該如何提升原住民學生的學習成就，應是實施TASA測驗的重要目的之一。故建議TASA測驗在施測完之後，應可以提供學校相關的測驗結果與統計資料，以作為學校實施補救教學與教師改進教學的依據與參考，讓評量可以發揮應有的效果。

參考文獻

一、中文部分

- 行政院原住民族委員會（2006）。*原住民人口數統計資料*。線上檢索日期：2009年1月18日。網址：http://www.apc.gov.tw/main/docDetail/detail_TCA.jsp?isSearch=&docid=PA000000000535&cateID=A000297&linkSelf=161&linkRoot=4&linkParent=49&url=
- 余民寧（1993）。試題反應理論的介紹（13）－試題偏差的診斷。*研習資訊*，10(6)，7-11。
- 余民寧、謝進昌（2006）。國中基本學力測驗之DIF的實徵分析：以91年度兩次測驗為例。*國立高雄師範大學教育學系教育學刊*，26，241-276。
- 林坤昌（1998）。*DIF檢定方法之探討與比較*。國立台中師範學院國民教育研究所碩士論文，未出版，台中市。
- 林奕宏、林世華（2004）。國小高年級數學科成就測驗中與性別有關的DIF現象。*台東大學教育學報*，15，67-96。
- 張建成、黃鴻文、譚光鼎（1993）。少數民族教育的理論與實際。載於中國教育學會（主編），*多元文化教育*。台北：台灣書店。
- 教育部（1996）。*全國原住民教育會議參考資料（16-38）*。台北：教育部。
- 黃毅志（2002）。*社會階層、社會網絡與主觀意識－台灣地區不公平的社會階層體系之延續（2版）*。台北：巨流。
- 盧雪梅（1999）。差別試題功能（DIF）的檢定方法。*台北市立師範學院學報*，30，149-166。
- 簡茂發、劉湘川、許天維、郭伯臣和殷志文（1995）。以Mantel-Haenszel法檢定試題區別功能之相關因素探討。*測驗年刊*，42，85-102。

二、英文部分

- Baker, F. B. (2001). *The basics of item response theory*. Portsmouth, N.H.: Heinemann.
- Camilli, G., & Shepard, L.A. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Lane, S., Wang, N., & Magone, M. (1996). Gender-related differential item functioning on a middle-school mathematics performance assessment. *Educational Measurement: Issues and Practice*, 15(4), 121-127.
- Ryan, K. E. & Fan, M (1996). Examining Gender DIF on a Multiple-Choice Test of Mathematics: A Confirmatory Approach. *Educational Measurement: Issues and Practice*, 15(4), 15-20.

3A Panel discussion III

2006年TASA國二數學試題、性別的差別試題功能（DIF） 與九年一貫能力指標達成率分析研究

— 曾建銘

2006年 TASA國二數學試題、性別的差別試題功能 (DIF) 與九年一貫能力指標達成率分析研究

曾建銘

國家教育研究院測驗及評量研究中心
cheng@mail.naer.edu.tw

中文摘要

本研究主要目的，是利用2006TASA國二數學抽測之學生作答資料，來分析國二學生在九年一貫數學各能力指標的學習成就及測驗試題分析，並探討試題本身對於男女兩群體是否有差異功能現象，若在其公告版試題中有DIF試題將進一步對試題內容做質性研究，以判斷該試題是否有偏誤。

整份測驗以傳統測驗理論分析，試題平均難度為0.56，難易度屬於中等者(0.4~0.8)約佔88%；平均鑑別度為0.7，除了一題低於0.3外其餘之鑑別度皆屬於良好。以試題反應理論(IRT)分析，整份測驗以三參數(3PL)模式下估計，試題平均鑑別度為1.97，除第一題為0.63稍低外，其餘介於0.65~4.09皆符合中、高鑑別度標準。難度方面，平均難度為0.19，大致符合中間難度。猜測度方面，平均猜測值為0.2，不至於過高。全部試題算是難度中稍微偏易且有高鑑別度的試題。

國二學生數學能力指標學習成就狀況，在數與計算、量與實測、幾何、統計與機率、代數等五大測驗內容中，以統計與機率領域的平均答對率最高達69.4%，其依序為幾何58.1%，數與計算54.9%，代數53.4%，最低為量與實測為53.3%。

以SIBTEST軟體分析試題性別的DIF現象，在設定 $\alpha=.01$ 的標準下，78題中有7題呈現DIF，約佔總題數的9%，其中3題屬於數與計算、幾何範圍對男生有利，其餘4題對女生有利且都屬於代數範圍。在機率與統計、量與實測則無DIF題出現。7題呈現DIF的試題有2題出現在公告版中，經進一步就試題內容做質的分析，結果此兩題不屬於試題偏誤。

最後，根據上述研究之結果與發現，提出對TASA資料庫分析及未來研究的建議。

關鍵字：TASA、IRT、九年一貫課程、數學能力指標、SIBTEST、DIF。

The Gender DIF Study of the Grade 2th Junior High School Mathematics items of 2006 TASA and the Achievement Rate of Mathematics Competence of Grade 1-9 curriculum

Chien-Ming Cheng

National Academy for Educational Research

E-mail:cheng@mail.naer.edu.tw

Abstract

The purpose of this study is to understand the students' achievement on the core mathematics competence of Grade 1-9 curriculum by the second grade response data of the junior high school students of Taiwan Assessment of Student Achievement (TASA), to analyze the test items, and to investigate whether or not there are items existing Differential Item Function (DIF) items between genders.

By classical test theory analysis, the average of item difficulty index is 0.56. The difficulty index of all items belongs to median difficulty (0.4~0.8). The average of item discrimination index is 0.7. All of the items have good discrimination index except one (item 60) is lower than 0.3. By item response theory analysis, the average of item difficulty index is 0.19. The average of discrimination index is 1.97. The average of item guessing index is 0.2. The quantity of whole test items could be said with median difficulty as well as high discrimination.

The test content can be divided to five dimensions. They are numbers and calculation, quantity and measurement, geometry, probability and statistics, as well as algebra. The highest average correct response rate is probability and statistics with 69.4%. The followings are geometry, numbers and calculation, algebra, as well as quantity and measurement with 58.1%, 54.9%, 53.4%, 53.3%, separately.

There are 7 DIF items after detecting by SIBTEST method. It is around 9%. Three DIF items belong to numbers and calculation, as well as geometry content which favor boys. The other four DIF items which belong to algebra are advantage to girls.

Key words: TASA、IRT、Grade 1-9 curriculum、mathematics competence indicators、SIBTEST、DIF

壹、緒論

第一節 研究背景與動機

回顧剛開始實施「九年一貫暫行綱要數學課程」時，出現學校數學教師，因不了解能力指標的具體意義，產生不知如何拿捏教學的深度，以及對新課程教學方式的不熟悉，而引發教學專業的焦慮現象，家長及社會大眾對於子弟在學校數學的學習內容，以及學習成效皆無法了解，社會媒體的報導更是褒貶不一。數學老師與家長對於九年一貫實施後一學生的數學程度降低的反映時有所聞，此外家長與學生更因有別於傳統的數學評量方式，無法直接判別學習情況的優劣而對學習成效茫然無知，終究引發各界對九年一貫課程新方向的質疑。因此，如何能解除學校老師與學生的疑惑，清楚地向家長及社會大眾解釋，便是修訂後數學課程綱要一項重要的課題。

既然各界困惑的原因是不知道要教（學）什麼、教（學）到什麼程度，那麼若是能有一份根據能力指標內容來編排的測驗，社會大眾、學校教師便能根據學生在測驗上的表現，分析教學目標與學生實際學習之成效，來檢討教師本身的教學內容和技術，如此一來能得到教學上的回饋，自然能提供教師的自信心與安全感，肯定自我的教學專業而不再焦慮（蔡清田，2003）。另外，家長及社會大眾對於子弟的學習情形及受教權的關心，更是有賴於具有公信力、客觀且有效的評量結果了。若能針對全國學生進行抽測建立資料庫以了解學生在能力指標學習成就的結果與變化趨勢，即可解答社會大眾對九年一貫實施後，學生學習成效結果的初步疑慮。

美澳等先進國家的教育系統，對於學生基本學習能力的表現水準，都有相當深切的關懷及具體明確的認知，這些國家對於其學生的學術成就指標，例行性地進行資訊的搜集與統整。國內長期缺乏量化指標和標準化測量工具來檢視學生學習成就的表現及其差異，以致無法確實瞭解課程實施的成效，亦不利於課程發展之進行與相關教育政策之研擬。目前教育資料庫的建立普遍受到重視，「臺灣學生學習成就評量資料庫（TASA）」，從2005年起採建

置題庫模式，進行組題每年對全國小四、小六、國二、高中二及高職二等五個年段，依人口密度進行二階段分層抽樣，對國語文、英語文、數學、社會及自然等五科施行抽測(小四英語文、社會除外)，2005年先行並對全國國小六年級學生的國、英、數三科抽測，2006年起則按原計劃對五個年段五科進行抽測，以了解全國國民中小學、高中職學生之學習成效，並建立年級、跨學科之學生學習成就長期性的資料庫，可提供國內專家學者或學術單位進行基礎性研究(國立教育研究院籌備處，2007)。

為了達成九年一貫課程目標，好的教材與教學法是不可或缺的，而教材的編寫則需根據課程綱要，課程綱要亦隨著世界潮流與社會文化政治等環境不斷地演進，在課程綱要修訂的過程中，學生對於目前所實施課程中各項能力指標的達成狀況，可反映出學生的學習成就，並進一步可反應出教材內容安排的適切性，即教材內容的編序是否符合學生心智的成長，欲進行檢測需要一份具有良好效度的工具，來了解學生學習成就概況。這好比人在接受健康檢查時，會透過儀器來檢查，例如抽血後經由儀器進行血液分析，得出血液分析報告書，每一個檢驗項目會得到一個數值，醫師依據檢驗項目的判讀參考值，來判讀檢驗項目的功能是否異常，醫師再依據診斷結果開處方或治療方式。學生的學習若能比照醫學檢查的模式，建立科學、客觀的標準，以供將來課程綱要修訂時參考。TASA所建立之資料庫恰可提供此資訊，因此本研究擬利用2006年TASA數學科目國二年段施測後所蒐集到的學生作答資料來進行分析，以了解學生在國二數學各能力指標之學習成效，並進一步分析試題在量方面的統計資料如古典理論與現代測驗理的難度、鑑別度、內部一致性效度、選項誘答力等。

另外本研究擬進一步研究試題本身是否呈現不公的資訊，即具有相同能力的兩群體，理論上他們的答對率應當是大概相等的，但若是實際的結果卻顯示有差異，我們稱為此問題存在著差異功能(Differential Item Functioning)，值得我們進一步研究其造成的因素，若發現是試題偏誤(item bias)，則該刪除以達考試的公平性。試題本身對於男女兩群體是否存在DIF的現象亦是為歐美測驗界所關注的議題，因TASA之試題具有保密性，所以將只提供統計後具

有DIF之試題資訊，作為TASA研發試題時，對具有DIF的試題就內容對於男女兩群體是否存在進行檢視，以確認該試題是否偏誤，若為偏誤題該如何避免、修改或刪除。

基於上述之研究動機，本研究擬達成下列兩項目的，第一，國二數學各能力指標之學習成效，利用2006TASA國二數學抽測所得之學生作答資料，根據「國民中小學九年一貫課程數學領域綱要」國二的能力指標來分析學生在各能力指標的學習成就，並分析測驗之試題。第二，探討試題本身對於男女兩群體是否有差異功能現象，若在其公告版試題中有DIF試題將進一步對試題內容做研究，以判斷該試題是否有偏誤。最後，針對結果提供建議，供未來研究及實務參考。

貳、研究方法

一、研究對象

為能確保「臺灣學生學習成就評量資料庫」之建立所抽取之樣本具有全國代表性，且因臺灣各縣市人口多寡各異，為充分顯現教改後臺灣學生學習成就實際情形，故採二階段隨機抽樣設計。第一階段採分層叢集隨機抽樣，根據縣市、人口密度、及班級數等三個變項進行分層。第二階段則根據所抽取到之樣本學校，對每一層以學生個人為單位，進行簡單隨機抽樣(國立教育研究院籌備處，2007)。

二、測驗編製過程(國立教育研究院籌備處，2007)

a、命題過程

2006TASA國二數學測驗的編製主要聘請現職教師協助，根據國中二年級的課程內容以及課程標準架構完成命題工作，測驗編製前先行舉辦命題研習。透過命題研習課程，讓命題教師進一步瞭解2006TASA數學測驗的研究目的，以及不同型態測驗題型之編製原則。透過範例試題分享，讓命題教師明確掌握命題規範。

b、修審題

2006 TASA數學從命題結束後開始進行修審題工作，為期約2個月。現職教師完成的命題初稿，先由2位數學內容專家審查試題修訂內容的精確性和可讀性。而後再由2位測驗專家審查修訂試題的形式。修審題之後，進行題庫的整理，針對測驗結構與各內容實際試題的數量進行檢查，題數不足的細格再請相關人員進行補充命題、審題的工作直到試題數量分配達到預期標準。

c、臨床施測

試題完成內容及形式審查後，研究小組著手進行試用測驗編輯之作業。預試學生用國中三年級，人數約有750人。透過預試之實徵資料進行統計分析，修改或刪除不適用的試題，以進行後續正式測驗的編輯。

d、命題架構

2006 TASA數學國中二年級測驗內容，是以教育部所公布的「國民中小學九年一貫課程暫行綱要」為依據，同時參酌各版本的教材。測驗內容皆側重學生生活數學經驗的連結，試題的設計以數學概念理解、程序執行、和問題解決能力為主要評量的目標。測驗架構是以數學內容及作業複雜度兩個向度建構而成，數學內容又可區分為數與計算、量與實測、幾何、統計與機率、以及代數五個內容領域。

表1是2006TASA數學國中二年級測驗內容與題型的題數分配架構，表中各內容領域所占的題數比例，主要參酌課程綱要與現行教科書的內容領域比重。國中二年級以「數與計算」、「幾何」與「代數」比重較重，而「量與實測」、「統計與機率」之相對比重較輕。

表1 2006 TASA數學國中二年級測驗各內容領域題數分配百分比

	選擇題	題數%
數與計算	16	21%
量與實測	8	10%
幾何	16	21%
統計與機率	8	10%
代數	30	38%
總題數	78	100%

三、正式測驗題本試題分配及BIB量尺化設計

所有2006 TASA數學試題分為13個試題區塊，每個區塊6題，每個學生作答3個試題區塊，也就是說一份測驗題本包含3個試題區塊。因此，每一份題本中有18題選擇題，施測時，題本即依事先派定的順序分發給學生。測驗時間為45分鐘。其中包含指導語說明及回收時間各5分鐘，故學生實際作答時間為35分鐘。於測驗科目結束後，另有一節問卷填答時間，供學生回答測驗相關問卷問題，預計學生問卷填答時間約為5分鐘(國立教育研究院籌備處，2007)。

然而受限於學生受測時間以35分鐘為主的因素，並無法使每位學生都能受測到所有試題，因此TASA採用平衡不完全區塊（Balanced Incomplete Block, BIB）設計題本模式，其優點為試題區塊與題本的配置方式，採螺旋式排列方式，可使每一個試題區塊的施測次數相同（van der Linden, Veldkamp, & Carlson, 2004），且IRT參數估計是採用同時估計法（concurrent calibration），如此可減少等化程序所介入的額外誤差（Kim, & Cohen, 1998; Kim, & Cohen, 2002; Kolen, & Brennan, 1995; Petersen, Cook, & Stocking, 1983; Stocking, & Lord, 1983）。因此TASA將測驗題78題分為13個試題區塊（treatment），每個區塊中包含6題試題，且區塊中的試題不重複。將13個試題區塊（1~13）編製成13個題本，而每個題本中的區塊可能部分相同或完全不同，但每個區塊出現的次數是一樣的，亦即每個試題所受測的學生約是相同的。

四、研究工具與資料處理

本研究採用的資料分析工具為統計分析軟體SPSS for Window 11.0，以及試題分析軟體BILOG-MG 3.0(Zimowski, Muraki, Mislevy, & Bock, 2003)、TESTER 2.0（余民寧，2002）與SIBTEST (Li & Stout, 1994)等，藉此進行資料分析及考驗，以達研究目的。

以統計分析軟體SPSS for Windows 11.0與試題測驗分析程式TESTER 2.0進行傳統試題及測驗分析，求得試題統計數（含答對率、點二系列、試題難度、試題鑑別指數等）以及選擇題之選項分析。

以BILOG-MG 3.0分析軟體進行試題參數分析，本研究利用此軟體計算測驗每一題之難度參數（b）、鑑別度參數（a）、猜測度參數（c）、試題測驗訊息、資料與試題的適合度。在兩種模式分析下，分別求其鑑別指標、難度指標、受試者能力等指標的相關情形。統計分析學生在每一能力指標之平均答對率。最後以SIBTEST軟體進行性別的DIF現象分析。

五、SIBTEST方法

檢驗試題是否為試題差異功能的方法中，較常用且有名的包括區域測

量法(Area Measure)(Camilli & Shepard, 1994; Lord, 1980; Millsap & Everson, 1993)、概似率考驗法(Model Comparison Measure)(Thissen, Steinberg & Wainer, 1988, 1993)、Mantel-Hanszel法(Mantel & Hanszel, 1959)及SIBTEST法(Shealy & Stout, 1993)。根據Cheng(2005)的研究發現，在以上的試題差異功能檢驗方法中，對於國中基本學力測驗數學科而言，SIBTEST具有最佳效率，因此本研究將採用SIBTEST來檢定2006TASA國二數學科測驗試題對男女的受試者是否有試題差異功能發生。

雖然SIBTEST是根據IRT所發展出來的檢定DIF方法，但SIBTEST並未涉及IRT的試題參數估計，而是假設具有相同配對分數的受試者，其潛在能力也相同，因此檢定DIF是以配對子測驗之分數作為分組計算的依據。應用SIBTEST進行DIF檢定，先將試題分為兩部份，其中一部份由無DIF試題組成稱為有效部份測驗(valid subtest)；另一部份稱為可疑部份測驗(suspect subtest)，這部份的試題將接受DIF檢定。SIBTEST DIF的計算如下：假定共有N個試題，其中前n個試題為有效部份測驗的試題，以 U_i 表示答對或答錯，則每個受試者在有效部份測驗的總分為 $X = \sum_{i=0}^n U_i$ ，受試者在可疑部份測驗的總分為 $Y = \sum_{i=n+1}^N U_i$ 。SIBTEST DIF統計數的計算公式如下：

$$\hat{\beta}_U = \sum \hat{P}_k (\bar{Y}_{Rk}^* - \bar{Y}_{Fk}^*) ,$$

\hat{P}_k 是在有效部份測驗得分 $X=k$ 之受試者中焦點組人數所佔的比例， $(\bar{Y}_{Rk}^* - \bar{Y}_{Fk}^*)$ 是兩組受試者在可疑部份測驗試題之得分的調整平均數的差異，這差異分數可以是單一試題得分或一組試題得分的差異，若是為單一試題得分的差異則為DIF檢定，在此研究為此種單一試題DIF檢定，如果無DIF存在，值將會是0。

參、研究結果分析與討論

一、試題分析

本研究資料刪除部份未答題經重新整理後，78題之受試學生人數介於1685人與1647人之間，以區塊5即第25~30題之受試學生最多，區塊13(第73~78題)之受試學生最少。研究者以SPSS for Window 11.0及TESTER 2.0進行試題分析結果如表2所示。

(一)、古典理論部份：

1.答對率

各試題答對率(P值)介於0.21至0.84之間，其中答對率低於0.4者有第3、37、42、53、60、71、73等七題（佔總題數9%），以第60題（ $P=0.21$ ）最低；高於0.8者有第5、35題等二題（3%），以第5題（ $P=0.84$ ）最高；介於0.4至0.8者有六十九題（88%），整份測驗的平均答對率為0.56，難易度適中偏易。

2.鑑別度

就鑑別度而言，低於0.25者有第60題（佔總題數不到1%），其餘各題（99%以上）均在0.3以上，整份測驗的平均鑑別度（D）高達0.7。

3.選項分析

研究者將所得資料，以TESTER 2.0統計軟體分析每一試題的選項，以進行選擇題的選項有效性分析，其結果除第53題的選項4，高分組有34%選，而低分組只有33%的人選；以及第60題的選項3，高分組有52%選，而低分組只有41%的人選外，餘皆符合「每題不正確的選項至少有一個低分組的受試者選它」以及「低分組比高分組更多人選擇不正確的選項」兩個選項誘答力原則（郭生玉，1999）。此兩題的答對率分別只有30.4%與21.3%，值得進一步對此兩題中的兩個選項做質性的分析。

題號:53

選 項	1	2*	3	4	其他	
選項率	.18	.30	.14	.37	.01	通過率：30.4
高分組	.08	.55	.04	.34	.00	難 度：0.33
低分組	.29	.11	.27	.33	.01	鑑別度：0.44

題號:60

選 項	1	2*	3	4	其他	
選項率	.16	.21	.48	.12	.01	通過率：21.3
高分組	.12	.29	.52	.06	.01	難 度：0.22
低分組	.25	.14	.41	.18	.02	鑑別度：0.15

4. 試題間的一致性

將全體受試者在每一道試題的答題情形（答對或者答錯）與其得分（答對的總題數），分別求其積差相關，亦即求每一道試題的點二系列相關係數，如表2所示，結果顯示測驗總分與每一道試題的得分除第60題為負外餘皆成正相關。但第51、71、72、73等四題的點二系列相關係數皆低於0.2。

(二)、試題反應理論部份：**1. 資料-模式適合度**

本研究以BILOG-MG 3.0軟體所計算的卡方值，作為三參數模式的適合度的指標，結果如表2所示，表中列出模式的卡方的p值，在此因為總受試人數達7503人，所以顯著機率標準訂為.01，當機率值小於.01的顯著水準，則表示試題適合度不佳；表中數字左上方標有「*」者表示達.01顯著水準。由表2的結果顯示，卡方的p值大於.01以上顯著水準者，亦即適合度較佳的試題，共有56題佔總試題數的72%。

2. 試題參數估計

在參數估計方面，本研究將7503名受試者在測驗的作答資料經重新整理成一7503X78的大矩陣後，以BILOG-MG 3.0軟體進行IRT三參數模式之試題參數估計，其結果如表2。鑑別度參數介於0.63~4.09之間，平均鑑別度為1.97；難度參數介於-1.34~2.81，平均難度為0.19；猜測值介於0.07~0.36，平均猜測值為0.20。

3. 試題最大訊息函數

在試題所能提供最大訊息函數方面，以第44題所能提供的訊息量最大達9.34，其相對的最大能力值為0.16，而以第1題所能提供的訊息量最少，只有0.19。

表2 試題分析表

題號	受試人數	答對率	鑑別度	點二系列相關	a	b	c	X ² 的P值	此能力值所得訊息量最大值	最大訊息量
1	1667	70.5	0.53	0.21	0.63	-0.88	0.21	0.00*	-0.63	0.19
2	1667	60.5	0.75	0.37	1.41	0.04	0.21	0.05	0.16	0.95
3	1667	27.4	0.59	0.31	2.91	1.24	0.15	0.00*	1.28	4.62
4	1667	43.6	0.78	0.37	2.13	0.78	0.21	0.02	0.85	2.18
5	1667	84.3	0.43	0.34	1.44	-1.34	0.19	0.00*	-1.23	1.04
6	1667	64.7	0.76	0.42	2.49	0.13	0.33	0.14	0.22	2.35
7	1665	43.8	0.62	0.27	1.32	1.06	0.24	0.09	1.20	0.79
8	1665	56.4	0.83	0.49	1.99	0.04	0.14	0.60	0.10	2.18
9	1665	72.6	0.62	0.39	1.75	-0.34	0.31	0.45	-0.22	1.21
10	1665	57.7	0.75	0.43	1.38	-0.08	0.12	0.00*	0.00	1.09
11	1665	67	0.75	0.49	2.33	-0.25	0.21	0.05	-0.18	2.61
12	1665	61.4	0.81	0.50	2.15	-0.11	0.17	0.34	-0.04	2.43
13	1660	44	0.79	0.48	2.01	0.45	0.12	0.02	0.50	2.31
14	1660	58.9	0.75	0.41	1.67	0.01	0.19	0.11	0.11	1.39
15	1660	64	0.74	0.47	1.99	-0.21	0.19	0.01	-0.14	1.99
16	1660	49.9	0.66	0.31	1.22	0.61	0.22	0.87	0.75	0.70
17	1660	78.4	0.48	0.29	1.13	-1.04	0.25	0.00*	-0.88	0.57

18	1660	52.3	0.64	0.24	1.45	1.02	0.36	0.01	1.18	0.75
19	1663	44.9	0.74	0.37	1.92	0.74	0.21	0.77	0.83	1.76
20	1663	54.3	0.75	0.43	1.58	0.17	0.16	0.48	0.25	1.34
21	1663	59.5	0.81	0.47	1.89	-0.01	0.18	0.04	0.06	1.84
22	1663	65.7	0.67	0.42	1.51	-0.27	0.19	0.00*	-0.17	1.13
23	1663	74	0.65	0.43	1.79	-0.64	0.19	0.00*	-0.56	1.60
24	1663	59.7	0.74	0.44	1.63	0.00	0.19	0.03	0.09	1.34
25	1685	44	0.64	0.28	1.92	1.07	0.28	0.13	1.17	1.56
26	1685	46.9	0.80	0.45	2.65	0.57	0.21	0.05	0.63	3.39
27	1685	64.8	0.74	0.45	1.82	-0.14	0.23	0.23	-0.05	1.55
28	1685	67.7	0.80	0.53	2.36	-0.32	0.19	0.04	-0.25	2.80
29	1685	71.3	0.70	0.53	2.36	-0.52	0.15	0.00*	-0.47	3.01
30	1685	48.1	0.77	0.42	2.04	0.54	0.20	0.30	0.62	2.02
31	1668	73.7	0.63	0.35	1.24	-0.66	0.24	0.00*	-0.52	0.70
32	1668	62.5	0.84	0.55	2.37	-0.19	0.14	0.46	-0.14	3.09
33	1668	64.6	0.83	0.56	2.16	-0.35	0.10	0.05	-0.30	2.77
34	1668	74.6	0.71	0.51	2.11	-0.78	0.09	0.00*	-0.74	2.72
35	1668	82.6	0.55	0.46	2.58	-0.89	0.26	0.01	-0.82	2.93
36	1668	64.9	0.75	0.41	1.27	-0.41	0.13	0.02	-0.32	0.91
37	1667	26.7	0.61	0.43	2.22	1.01	0.07	0.01	1.04	3.08
38	1667	57.2	0.80	0.40	2.88	0.44	0.34	0.20	0.52	3.11
39	1667	48.2	0.71	0.30	1.73	0.86	0.29	0.70	0.98	1.23
40	1667	48.4	0.72	0.33	3.15	0.81	0.32	0.04	0.88	3.84
41	1667	55.4	0.76	0.47	1.93	0.14	0.18	0.37	0.21	1.92
42	1667	39.4	0.67	0.28	1.68	1.14	0.23	0.18	1.24	1.30
43	1679	53	0.84	0.47	1.49	0.08	0.10	0.00*	0.14	1.34
44	1679	53.1	0.91	0.62	4.09	0.13	0.13	0.70	0.16	9.34
45	1679	49.1	0.84	0.49	2.86	0.45	0.21	0.24	0.50	3.95
46	1679	74.5	0.68	0.45	1.95	-0.67	0.18	0.00*	-0.59	1.97
47	1679	59	0.79	0.46	2.00	0.05	0.20	0.46	0.12	1.99
48	1679	69.1	0.75	0.52	3.18	-0.22	0.26	0.43	-0.16	4.39
49	1669	52	0.79	0.46	1.81	0.26	0.16	0.41	0.33	1.74
50	1669	48.5	0.84	0.46	1.57	0.24	0.09	0.00*	0.30	1.49

51	1669	64.8	0.78	0.47	2.32	-0.07	0.26	0.18	0.02	2.35
52	1669	51.3	0.84	0.53	3.01	0.29	0.18	0.21	0.33	4.64
53	1669	30.4	0.44	0.12	0.81	2.81	0.22	0.05	3.01	0.32
54	1669	63.6	0.81	0.55	2.75	-0.23	0.14	0.01	-0.19	4.17
55	1668	49.9	0.74	0.29	1.26	0.65	0.23	0.06	0.78	0.74
56	1668	63	0.79	0.48	2.36	-0.10	0.21	0.01	-0.03	2.69
57	1668	62.1	0.68	0.45	1.90	-0.17	0.15	0.04	-0.11	1.93
58	1668	62.1	0.66	0.37	1.21	-0.29	0.13	0.00*	-0.19	0.83
59	1668	50.8	0.67	0.30	1.03	0.52	0.19	0.09	0.67	0.53
60	1668	21.3	0.15	-0.11						
61	1667	57.7	0.70	0.31	0.98	0.06	0.17	0.27	0.20	0.50
62	1667	66.5	0.70	0.40	2.08	-0.04	0.30	0.07	0.06	1.76
63	1667	44.3	0.58	0.22	1.74	1.26	0.31	0.47	1.38	1.19
64	1667	43.9	0.75	0.44	2.16	0.56	0.16	0.05	0.62	2.48
65	1667	69.5	0.61	0.29	0.94	-0.54	0.22	0.00*	-0.37	0.42
66	1667	78.3	0.60	0.40	1.71	-0.91	0.16	0.00*	-0.83	1.56
67	1652	55	0.76	0.43	1.54	0.09	0.15	0.00*	0.17	1.28
68	1652	49.6	0.82	0.51	2.60	0.28	0.16	0.77	0.33	3.60
69	1652	66.9	0.67	0.36	1.40	-0.28	0.24	0.00*	-0.15	0.89
70	1652	47.2	0.75	0.47	2.65	0.40	0.17	0.19	0.46	3.64
71	1652	34.7	0.51	0.15	2.15	1.73	0.28	0.12	1.83	1.93
72	1652	41.1	0.58	0.18	1.46	1.56	0.30	0.24	1.70	0.85
73	1647	25.5	0.34	0.09	2.55	2.07	0.22	0.00*	2.13	3.09
74	1647	48.2	0.82	0.52	2.30	0.32	0.14	0.22	0.37	2.94
75	1647	65.5	0.81	0.55	3.02	-0.28	0.15	0.00*	-0.24	4.91
76	1647	69.4	0.71	0.49	2.37	-0.40	0.19	0.00*	-0.33	2.82
77	1647	41.7	0.62	0.23	1.94	1.32	0.30	0.00*	1.43	1.53
78	1647	52.3	0.81	0.43	2.19	0.42	0.24	0.89	0.50	2.19
平均		56.3	0.70	0.40	1.97	0.19	0.20			

註：第60題因點二系列相關為負值，因此在IRT三參數的模式下，並不
做進一步的估計，所以該題在IRT部份皆為空白。

4.CTT與IRT 結果之比較

(一) 難度指標

就難度指標而言，以 CTT之難度值(答對率)與IRT三參模式難度參數作一比較，兩者之間的相關，如表3所示：

表3 IRT與CTT之難度指標相關

CCT難度	三參數難度	
CCT難度	1.00	
三參數難度	.93**	1.00

** $p < .01$

結果顯示CTT與IRT三參模式的難度相關為0.93之間，相關值甚高，此應為在CTT中較難的試題通過人數較少相對的在IRT中較高難度的試題答對機率較低，結果表示CTT的難度參數指標與IRT的難度參數指標具有相似性。

(二) 鑑別度指標

在鑑別度方面，在表4所呈現的是CTT的點二系列相關係數與三參數模式的鑑別度，兩者之間的交叉相關：

表4 IRT與CTT之鑑別度指標相關

	點二系列相關係數	三參數鑑別度
點二系列相關係數	1.00	
三參數鑑別度	.54**	1.00

** $p < .01$

結果顯示點二系列相關係數與三參數模式的鑑別度相關為0.54。此一結果符合Lord (1980) 的論述，亦即在考慮猜測值的情形下，由於猜測值 c 的增加，使得原本為0~1的答題可能機率（ p ），變為 c ~1之間，壓縮了答對試題與否機率的空間，由於每一道試題的猜測值不同，使得每一道試題的鑑別度，因而產生變化，這可以解釋何以三參數模式鑑別度與CTT鑑別度指標相關如呈現中度而非高相關的原因。

二、TASA2006國中二年級各能力指標達成狀況

資料經整理後統計分析得到下表5中各能力指標的平均答對率，首先就一數與計算、量與實測、幾何、統計與機率、代數等五大測驗內容觀之，以統計與機率領域的平均答對率最高達69.4%，其依序為幾何58.1%，數與計算54.9%，代數53.4%，最低為量與實測為53.3%。

再更進一步就「數與計算」測驗內容來看，平均答對率最高的能力指標為N_3_15能在情境中理解比、比例(包括正比例和反比例)、比值、率(百分率、ppm)的意義，平均答對率為66.5%，最低的為N_3_18能察覺整數的因數、倍數、公因數、公倍數，只有21.3%。但此題為第60題其點二系列相關的值為負的，因此，似乎有必要檢視此試題的內容，以確定其試題是否能有效代表此能力指標。

在「量與實測」測驗內容中則以N_3_14 能將各種柱體，變形成長方柱而計算其體積，形成柱體之體積計算公式的平均答對率最高達67%，而以N_3_17 能掌握米/秒和千米/時之間的關係，並利用此關係作化聚的平均答對率最低為49.1%。

接著在幾何的測驗內容中，以S_3_8 能瞭解平面圖形線對稱的意義平均答對率最高達74.5%，最低的為S_3_11能操作圖形之間的轉換組合與S_4_1 能根據給定的性質作局部推理，平均答對率皆為44%。

在統計與機率測驗內容中，以D_3_4 能報讀生活中有序資料的統計圖表，其平均答對率最高達82.6%，此亦為所有國中二年級各能力指標中平

均答對率最高者，在此類測驗內容中以D_4_3能進行簡單的實驗，以瞭解機率、抽樣的初步概念的平均答對率最低為55%。

最後在代數測驗內容中以A_4_1 能利用等量公理解從生活情境問題中列出的一元一次方程式的平均答對率最高達66.6%，最低為A_3_10 能瞭解幾何圖形及形體變動時，其幾何量對應變動情形，其平均答對率只有26.7%，此能力指標雖屬於代數測驗內容，但又牽涉到幾何問題，是否在試題中同時測兩個概念而導致困難度大增，值得進一步就試題內容做檢。

表5 國中二年級各能力指標平均答對率

測驗內容	能力指標	題數	平均答對率(%)
數與計算	N_2_16能知道先乘除後加減的約定，並能用來列式及簡化計算式子。	2	62.3
	N_3_1能延伸非負整數的認識。	1	49.9
	N_3_6在具體情境中，能用分數、小數表示除的結果(除的結果為有限小數)。	1	48.1
	N_3_7能用分數倍的概念，整合以分數為除數的包含除和等分除的運算格式。	3	61.5
	N_3_15能在情境中理解比、比例(包括正比例和反比例)、比值、率(百分率、ppm)的意義。	2	66.5
	N_3_18能察覺整數的因數、倍數、公因數、公倍數。	1	21.3
	N_3_20能察覺整數的最大公因數、最小公倍數、質數和合數，並能將一個數做質因數分解。	2	46.7
	N_4_1能掌握命數系統，並以科學符號表示一個數。	4	53.8
合計題數與平均答對率		16	54.9

	N_2_10 能認識各種量的普遍單位，應用在生活中的實測和估測活動，並培養出量感(普遍單位：千米、毫米、公升、毫公升、時、分、秒)。	2	52.2
	N_3_12 能對非直線形的平面區域，選定適當的正方形單位，估計其概略面積，並檢驗圓面積公式(π ， r 為圓的半徑)。	1	55.4
量與實測	N_3_13 能理解容量和容積(體積)之間的關係，並利用此關係計算大容器(如游泳池)之容量。	1	50.8
	N_3_14 能將各種柱體，變形成長方柱而計算其體積，形成柱體之體積計算公式。	1	67
	N_3_15 能在情境中理解比、比例(包括正比例和反比例)、比值、率(百分率、ppm)的意義。	2	49.7
	N_3_17 能掌握米/秒和千米/時之間的關係，並利用此關係作化聚。	1	49.1
	合計題數與平均答對率	8	53.3
	S_2_4 能運用東西南北的語詞描述位置及方向。	1	44.9
	S_3_4 能利用構成要素間的可能關係，描述複合形體要素間的可能關係。	2	60.5
	S_3_5 能利用形體的性質解決幾何問題。	6	58.6
	S_3_8 能瞭解平面圖形線對稱的意義。	1	74.5
幾何	S_3_10 能透過實測辨識三角形、四邊形、圓的性質。	2	53.1
	S_3_11 能操作圖形之間的轉換組合。	1	44
	S_4_1 能根據給定的性質作局部推理。	1	44
	S_4_5 能瞭解垂直、平行的定義。	2	71.6
	合計題數與平均答對率	16	58.1
	D_3_4 能報讀生活中有序資料的統計圖表。	1	82.6
	D_3_5 能將有序資料整理成折線圖，並抽取折線圖中有意義的資訊加以解讀。	1	65.7
統計與 機率	D_4_2 能將資料整理成圓形百分圖，並抽取圓形百分圖中有意義的資訊，加以解讀。	1	71.3
	D_4_3 能進行簡單的實驗，以瞭解機率、抽樣的初步概念。	1	55
	D_4_5 能解讀現成資料之折線圖、圓形百分圖、及與百分位數有關的統計圖表。	4	70.2
	合計題數與平均答對率	8	69.4

	A_3_1 能用 x 、 y 、 \dots 的式子表徵生活情境中的未知量及變量。	1	64.9
	A_3_2 能將生活情境中的問題表徵為含有 x 、 y 、 \dots 的等式或不等式，透過生活經驗檢驗、判斷其解，並能解釋式子及解與原問題情境的關係。	4	54.3
	A_3_5 能察覺簡易數量模式與數量模式之間的關係。	1	43.8
代數	A_3_6 能瞭解幾何量的各種表徵模式。	1	64.8
	A_3_10 能瞭解幾何圖形及形體變動時，其幾何量對應變動情形。	1	26.7
	A_4_1 能利用等量公理解從生活情境問題中列出的一元一次方程式。	2	66.6
	A_4_2 能解從生活情境問題中列出的二元一次聯立方程式。	3	59.4
	A_4_3 能檢驗、判斷不等式的解並描述其意義。	1	46.9
	A_4_4 能利用一次式解決生活情境中的問題。	2	56.7
	A_4_5 能畫出形如 $y=ax+b$ 的坐標平面圖形。	6	43.7
	A_4_7 能認識平方根以及用電算器看出其近似值。	1	49.9
代數	A_4_8 能使用乘法公式。	2	56.5
	A_4_9 能認識商高定理及其生活中的應用。	1	59.5
	A_4_11 能利用配方法或十字交乘法解一元二次方程式。	4	57.2
	合計題數與平均答對率	30	53.4

三、性別的差別試題功能 (DIF) 分析

本研究將所得之資料重新整理，根據考試區塊劃分共有13個區塊，每個區塊有六題試題，再根據每一區塊的受試者分成男女兩群，而將人數略少的女生設定為焦點組，因每個區塊只有六題試題，因此，並不另外設定有效子測驗(valid subtest)，而採用軟體內訂之自動DIF分析(automatic DIF analysis; ADA)，亦即分別以每一個試題作為受評試題，而以另外五題做為探索性分析之有效子測驗，以找出可能有DIF的試題，本研究以SIBTEST軟體進行單向性行(uniform)DIF檢定，結果如表6所示，若設定 $\alpha=.05$ 則在13個區塊中

只有2個區塊沒有DIF題，在78個試題中共有18題呈現DIF，因為對每一試題男女兩群體之總人數皆大於1600人，為了不高估DIF的題數，因此於本研究擬採較小之 $\alpha=.01$ 為判斷試題是否呈現DIF的標準，結果共有第7(02_01;代數)、18(03_06;數與計算)、20(04_02;代數)、25(05_01;幾何)、43(08_01;數與計算)、48(08_06;代數)、62(11_02;代數)等7題呈現DIF，約佔總題數的9%，其中第18、25、43等3題屬於數與計算、幾何範圍對男生有利，其餘4題對女生有利且都屬於代數範圍。在機率與統計、量與實測則無DIF題出現。

表6 SIBTEST檢定試題DIF之結果統計表

區塊 題號	Beta-hat	p_value	區塊 題號	beta-hat	p_value	區塊 題號	beta-hat	p_value
01_01	0.053	0.051	06_01	-0.013	0.687	10_01	0.06	0.034
01_02	-0.019	0.488	06_02	0.051	0.124	10_02	-0.02	0.443
01_03	0.024	0.317	06_03	0.018	0.585	10_03	-0.025	0.346
01_04	-0.048	0.085	06_04	0.042	0.133	10_04	0.004	0.89
01_05	-0.044	0.023	06_05	-0.024	0.318	10_05	0.038	0.181
01_06	-0.009	0.728	06_06	-0.026	0.437	10_06	-0.007	0.774
02_01	-0.083	0.01	07_01	0.062	0.027	11_01	0.035	0.226
02_02	0.024	0.408	07_02	0.049	0.115	11_02	-0.086	0.001
02_03	0.049	0.067	07_03	-0.028	0.396	11_03	0.038	0.2
02_04	-0.057	0.057	07_04	-0.004	0.91	11_04	-0.007	0.801
02_05	0.015	0.59	07_05	0.002	0.954	11_05	-0.048	0.068
02_06	0.043	0.125	07_06	-0.018	0.573	11_06	-0.05	0.025
03_01	0.037	0.19	08_01	0.131	0	12_01	0.058	0.046
03_02	0.001	0.974	08_02	-0.015	0.633	12_02	-0.07	0.016
03_03	-0.003	0.92	08_03	0.084	0.011	12_03	-0.026	0.362
03_04	-0.003	0.92	08_04	-0.065	0.018	12_04	0.028	0.35
03_05	-0.048	0.045	08_05	0.013	0.689	12_05	0.01	0.74
03_06	0.087	0.004	08_06	-0.127	0	12_06	-0.015	0.632
04_01	0.031	0.309	09_01	-0.038	0.197	13_01	0.007	0.788
04_02	-0.113	0	09_02	0.045	0.125	13_02	-0.003	0.912
04_03	0.016	0.588	09_03	-0.058	0.026	13_03	-0.002	0.941
04_04	0.012	0.686	09_04	0.05	0.078	13_04	-0.047	0.07

04_05	0.036	0.173	09_05	-0.001	0.98	13_05	0.033	0.274
04_06	0.027	0.366	09_06	-0.058	0.024	13_06	-0.034	0.232
05_01	0.085	0.008						
05_02	-0.037	0.214						
05_03	-0.022	0.435						
05_04	-0.019	0.482						
05_05	0.032	0.207						
05_06	0.003	0.931						

在7題DIF題中恰有第20、25題在公告測驗版中出現，如下：

第20題(04_02)、()化簡 $\frac{2}{3}(5x+4) - \frac{5}{2}(x-1)$ 的結果為何？

- (1) $5x+1$
- (2) $5x+31$
- (3) $\frac{5x+1}{6}$
- (4) $\frac{5x+31}{6}$

★正確答案：4

選 項	1	2	3	4*	其他	
選項率	.07	.16	.23	.54	.00	通過率：54.3
高分組	.01	.04	.04	.91	.00	難 度：0.53
低分組	.13	.27	.41	.16	.00	鑑別度：0.75

此題是單純的代數化簡問題屬於難度中等，由題幹觀之並無任何對女生有利的地方，但從上面的選項分析可以發現有23%的學生選3，學生會選到3的原因乃計算時後面括弧裡的負號忘記乘所造成的結果，但經進一步對兩群的選項做分析，發現男女兩群體選3的比例是22.8%與22.7%幾乎相同，但選擇正確答案4的比例分別為51.6%與57.4%有明顯的不同；選項2亦有16%的人選取，會選取2的原因為將化簡的通分與解方程式的去分母搞混，再分析兩群

體選取該項的比例分別為18.3%與12.9%，明顯地造成此題DIF的原因乃男女兩群體對「化簡的通分與解方程式的去分母」有不同比例的迷思概念所致。因此，本題並不能視為DIF題而是教師在教學過程中要特別強調此一迷思，尤其是男生。

第25題(05_01)、() 將下列哪一種四邊形四邊中點所連成的新四邊形不是菱形？

- (1) 長方形
- (2) 正方形
- (3) 鳶形
- (4) 等腰梯形

★正確答案：3

選 項	1	2	3*	4	其他	
選項率	.14	.06	.44	.37	.00	通過率：44.0
高分組	.03	.02	.81	.14	.00	難 度：0.49
低分組	.21	.09	.17	.53	.00	鑑別度：0.64

此試題是屬於幾何問題，難度為中稍為偏難，主要考的概念為四邊形中對角線是否相等的概念，再加上三角形兩邊中點連線等於第三邊的一半的概念。兩群體對於常見的長方形與正方形兩種圖形概念差不多，但對於鳶形與等腰梯形明顯地似乎有所差異，進一步就兩群體的選項分析可發現，男女選擇正確選項3的比例分別為46%與41.7%明顯有差異，選擇4的比例亦分別為34.7%與38.8%，因此就此試題而言亦不應列為DIF題。雖然是考幾何但並無將此問題中的四種圖形表出，基本上男生對於幾何圖形的概念優於女生，可嘗試在各選項中加入圖形，或許可以增加女生答對的比例；另一方面此問題亦可做為檢視女生是否了解幾何問題的好試題。

基於試題保密原則，本研究無法進一步對其餘5題DIF試題做質性內容分析，在此提供給TASA作為內部質性研究時參考。

肆、結論與建議

本研究主要目的，是利用2006TASA國中二年級數學施測的資料，來分析學生在各九年一貫各能力指標的學習成就，並且就試題量的方面做分析，以及利用SIBTEST軟體對性別分析試題是否存在差異功能（DIF），結果發現具有DIF的試題亦出現公告版的試題中，因此進一步就DIF問題進行質的探討。

一、結論

本研究對2006TASA所編製的國中二年級數學成就測驗施測後所得的資料，除了進行傳統試題分析，求得統計數（答對率、鑑別指數、難度、點二系列相關）外，還進一步以試題反應理論分析，估計每一題的難度、鑑別度、猜測度、試題訊息以及分析資料與試題之間的適合度，並探討性別DIF現象，最後再比較CTT與IRT的分析結果。茲將本研究主要結果摘述如下：

- 1、整份測驗以傳統測驗理論分析，其試題平均難度為0.56，難易度屬於中等（0.4~0.8）約佔88%；平均鑑別度為0.7，除了一題低於0.3外其餘之鑑別度皆屬於良好，全部試題算是難度中稍微偏易且有高鑑別度的試題。但其中第60題因點二系列相關為負的，因此此試題需進一步做內容效度的檢驗。
- 2、選項分析，除兩題各有一選項違反「低分組比高分組更多人選擇不正確的選項」的選項誘答力原則外，其餘76道試題的選項皆符合標準。
- 3、以試題反應理論（IRT）模式考驗，整份測驗在三參數模式下估計各參數分別為
 - （一）鑑別度方面，平均鑑別度為1.97，依Baker（2001）的觀點，除第一題為0.63稍低外，其餘介於0.65~4.09皆符合中、高鑑別度標準。
 - （二）難度方面，平均難度為0.19，大致符合中間難度。
 - （三）猜測度方面，平均猜測值為0.2，不至於過高。
- 4、CTT與IRT的分析結果之關係：

(一) 在難度方面，CTT與IRT三參數模式的難度相關為0.93。

(二) 在鑑別度上，三參數模式與點二列相關係數呈現中度相關值為0.54。

- 5、國二學生數學能力指標學習成就狀況，在數與計算、量與實測、幾何、統計與機率、代數等五大測驗內容中，以統計與機率領域的平均答對率最高達69.4%，其依序為幾何58.1%，數與計算54.9%，代數53.4%，最低為量與實測為53.3%。在個別數學能力指標以D_3_4能報讀生活中有序資料的統計圖表，其平均答對率最高達82.6%，而以N_3_18能察覺整數的因數、倍數、公因數、公倍數，只有21.3%。但此題為第60題其點二系列相關的值為負的，因此，似乎有必要檢視此試題的內容，以確定其試題是否能有效代表此能力指標。
- 6、以SIBTEST軟體分析試題性別的DIF現象，在設定 $\alpha=.01$ 的標準下，78題中有7題呈現DIF，約佔總題數的9%，其中3題屬於數與計算、幾何範圍對男生有利，其餘4題對女生有利且都屬於代數範圍。在機率與統計、量與實測則無DIF題出現。7題呈現DIF的試題有2題出現在公告版中，經進一步就試題內容做質的分析，此兩題不屬於試題偏誤。

二、建議

(一)、TASA測驗

- 1、選項分析，好的誘答選項可發現學生迷思概念，藉由學生錯誤的試題，了解其在哪個（些）能力指標，需要再學習，而施予補救教學。亦可作為教師改進教學的依據，以及課程至教材內容的編排。
- 2、藉由不同群體的DIF分析可以了解哪些能力指標對於哪些群體產生試題差異功能，再進一步做質的內容分析，以了解是試題偏誤或其他學習或教學所產生之結果。
- 3、以上之建議可擴及TASA其它年段與科目。

- 4、經由試題分析將各種類型的優良試題納入題庫中，將來可以藉助IRT的方法針對不同的測驗目標來設計不同的標準化測驗工具，以及發展電腦化適性測驗。

(二)、未來的研究

- 1、本研究所用之SIBTEST檢定DIF問題，只是眾多DIF方法中的一種，未來亦可用其它DIF方法來做檢定並比較，將存在試題偏誤之試題找出修改或刪除，以確保測驗的公平性。
- 2、不同的等化設計所估出之能力差異性如何？亦是將來可以探討的議題。
- 3、如何結合TASA資料庫來建立本國學生學習成就長期趨勢變化，並與國際作比較也是將來可發展的研究方向。

參考文獻

一、中文部份

余民寧(2002)。《教育測驗與評量》。台北：心理出版社。

郭生玉(1999)。《心理與教育測驗》。台北縣：菁華。

國立教育研究院籌備處(2007)。《2006工作成果報告書》。台北：國立教育研究院籌備處(2007)。

蔡清田(2003)。《九年一貫課程的評鑑意義、範圍與行動途徑》。檢索日期：民國95年5月10日。取<http://www.trd.org.tw/main.htm>。

二、英文部份

Baker, F. B. (2001). *The basics of item response theory*. Portsmouth, N. H.: Heinemann.

Kim, S., & Cohen, A. S. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurement, 22*, 131-143.

Kim, S., & Cohen, A. S. (2002). A comparison of linking and concurrent calibration under graded response theory. *Applied Psychological Measurement, 26*, 25-41.

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York: Springer-Verlag.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates, Inc.

Li, H.-H., & Stout, W. (1994). SIBTEST: A FORTRAN-V Program for Computing the Simultaneous Item Bias DIF Statistics [Computer program]. Urbana-Champaign, IL: University of Illinois, Department of Statistics.

- Petersen, N. S., Cook, L. L., & Stocking, M. L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. *Journal of Educational Statistics, 8*, 137-156.
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7*, 201-210.
- van der Linden, W. J., Veldkamp, B. P., & Carlson, J. E. (2004). Optimizing balanced incomplete block designs for educational assessments. *Applied Psychological Measurement, 28*, 317-331.
- Zimowski, M. F., & Muraki, E., & Mislevy, R. J. & Bock, R. D. (2003). BILOG-MG. [Computer program]. Chicago, IL: Scientific Software.



4A Panel discussion IV

新移民與本土子女數學學習成就差別試題功能
與多元潛在迴歸分析
— 以2007年臺灣學生學習成就評量資料庫為例

— 謝進昌 陳柏霖 余民寧

新移民與本土子女數學學習成就差別試題功能 與多元潛在迴歸分析 —以2007年臺灣學生學習成就評量資料庫為例

謝進昌

國家教育研究院測驗及評量研究中心助理研究員
jin@mail.naer.edu.tw

陳柏霖

國立政治大學教育學系博士研究生
97152516 @nccu.edu.tw

余民寧

國立政治大學教育學系教授
mnyu@nccu.edu.tw

摘要

本研究目的主要在探討本土與新移民子女間，數學學習成就是否存在顯著差異表現。然而要瞭解是否這兩族之間有差異，需從測驗題目進行差別試題功能分析。因此，本研究採用ConQuest軟體，分析2007年臺灣學生學習成就評量資料庫數學科成就測驗試題。運用差別試題功能與多元潛在迴歸進行分析。研究結果發現在進行族群間數學成就表現比較結果時，顯示在國小六年級內，並未存在顯著族群差異，但在國小四年級中，是以本土子女的表現平均優於新移民子女，而在控制家庭社經地位後，兩者的差異就消失了，顯示族群間家庭社經之不平衡可能才是影響成就差異之主因。最後根據研究結果，於文後提出相關建議與省思，供讀者參考。

關鍵字：差別試題功能、新移民子女、臺灣學生學習成就評量資料庫、
數學學習成就

Differential Item Functioning and Latent Regression Analysis in Mathematics Learning Achievement between New Immigrant and Native Children: A Case Example of 2007 TASA

Jin-Chang Hsieh

Assistant Research Fellow, Research Center for Testing and Assessment,
National Academy for Educational Research
jin@mail.naer.edu.tw

Po-Lin Chen

Ph. D. student, Department of Education, National Chengchi University
97152516 @nccu.edu.tw

Min-Ning Yu

Professor, Department of Education, College of Education, National Chengchi University
mnyu@nccu.edu.tw

Abstract

The purpose of this study is to investigate the phenomena of differential item functioning in mathematics learning achievement on 2007 Taiwan Assessment of Student Achievement (TASA) among different local children with new immigrant children. The analysis method for differential dimensionality was chiefly based on the ConQuest statistical package. The conclusions were as follows: Results based on mathematics learning achievement contents and performance indicated that there were no visible ethnic effect sizes in elementary school students for grade 6. But, local children were performed slightly better than new immigrant children amongst high mathematics learning achievement groups for grade 4. Because, to control personal socioeconomic status indicated that there were no visible ethnic effect sizes in elementary school students for grade 6 and grade 4. According to the results, the research made the suggestions and reflections for future studies are proposed.

Keywords: Differential item functioning, Mathematics learning achievement, new immigrant children, Taiwan Assessment of Student Achievement

壹、緒論

一、緣起

近十餘年來，台灣人口的自然成長已呈現逐年減少的趨勢。根據教育部統計處（2007）針對國民小學一年級新生人數的推估，至民國100年，人數將從現有的29萬人降至約20萬人，「少子女化」的時代儼然已經來臨！在出生人口中，新移民子女所占的人口比例逐漸提升，因此社會所衍生關注的議題，不再只是侷限於總人口數下降之問題，還包含社會人口結構多元及可能因應而來的社會衝擊現象。

然而，在這股人口成長逐年減緩的趨勢中，卻存在另一股新興力量，即是「新移民子女」的繁衍。依據內政部統計顯示，從1998年至2009止，歷年出生嬰兒之生母國籍，本國籍者所占比率是逐年趨緩，而至於非本國籍者，則是從5.12%，逐步上升至12.88%，再降至8.68%（教育部，2010a）。尤其，隨著新移民家長來臺時間的增長，進入國民教育階段就讀人數亦日益增加，從九十四學年度新移民子女就讀國中小人數為60,258人，到了現今（九十八學年度）合計為155,144人（教育部，2010b）。面對這群「新臺灣之子」，其學習成長狀況、家庭結構、生活適應、甚至是對未來國家政治、經濟、社會、文化衝擊的影響等問題，都值得深入分析。再加上，過去針對新移民子女的學習成就結果不一（王世英等，2006；周秀潔，2005；教育部，2005；黃琬玲，2005；劉秀燕，2003），更是激發研究者擬透過大型資料庫的分析，瞭解此一現況。並且，在國際教育成就調查委員會（IEA）公佈「國際數學與科學教育成就趨勢調查」（TIMSS 2007）結果中指出，臺灣國中小學生的數學分居全球第一、三名，表現優異（教育部，2008），引發研究者深思本國籍子女與新移民子女在數學學習成就是否有差異，相關文獻（何美瑤，2007；林佩蓉，2008；柯淑慧，2004；張慧貞，2005；賴翠媛，2006；鍾文悌，2005）的結果指出，新移民子女在數學科學習成就較低於本國籍子女，但蕭彩琴（2005）研究卻指出，兩者之間是無差異的；而引發研究者深思果真出現此差異嗎？還是其數學測驗題目，對於新移民子女與本地子女，其試題本質上就可能有不公平的現象出現嗎？或是家庭社經地位之平

等與否影響新移民子女與本地子女？

在國內國家教育研究院籌備處所建置之臺灣學生學習成就評量資料庫（Taiwan Assessment of Student Achievement, TASA）之2007年臺灣學生學習成就評量資料庫（TASA2007），主要在透過大規模的施測，瞭解不同學生層級學習成就的表現，本研究擬以此作為研究分析的樣本，以探討新移民子女數學學習成就表現，及其與本土子女進行比較；其次，瞭解數學科是否存在試題差異功能？希冀以此作為未來長期分析之基礎。

以下茲就影響新移民子女學習弱勢的可能因素與本研究所援用之數學成就評量架構，分別進行陳述：

二、影響新移民子女學習弱勢的可能因素

對於透過國際通婚而締結姻緣的新移民族群而言，離鄉背井來到異鄉，建立新的家庭，除經歷人生階段的轉變外，還要面對各類接踵而來文化差異的挑戰，所承擔者不僅是新移民女性本身，更廣及至整個家庭、社會環境等，如此，無論是直接或間接，都會影響到其子女的發展。過去相關研究結果（李嘉莉，2008；呂念慈，2007；那昇華，2007；黃森泉、張雯雁，2003）歸納，由於新移民女性之家庭地位、經濟狀況、語言能力、環境適應等種種弱勢及限制。影響新移民子女學習弱勢的可能因素，包括學校適應、學習成就、學習態度、語言學習、人際互動等方面（王振世、蔡清中，2008；高玉靜，2005；張芳全主編，2007；陳玉娟，2009；陳柏霖、陳書農，2007；黃森泉、張雯雁，2003）。目前研究新移民子女學習的領域，大多著重在語文或數學領域為主。換言之，新移民子女可能在上述因素的影響下，間接或直接造成學業表現上的弱勢，而對照本研究標的而言，即是數學成就之表現。

目前相關研究指出文獻（周炳言，2009；林佩蓉，2008；柯淑慧，2004；張慧貞，2005；教育部，2005；賴翠媛，2006；鍾文悌，2005）的結果指出，新移民子女在數學成就、或是整體數感測驗，較低於本地子女，但

蕭彩琴（2005）研究卻指出，新移民子女與本地子女及配對組相較之下，在五大課程領域並無差異。而（王振世、蔡清中，2008；何美瑤，2007）的研究分別提出，大陸籍、東南亞籍的新移民子女及本籍配偶子女，在數學成就都是女生高於男生、大陸籍與東南亞籍配偶子女在數學領域之學業成就有顯著差異。溫明麗（2007）更是指出，國中新移民子女在數學學習成就下降最多。

綜歸上述，新移民子女在數學學習成就低於本國籍子女，但亦有研究發現是無差異的。實際上，學習成就常使用紙筆測驗作為評量的依據，然而一份測驗其題目的情境內容，必須對同等能力的每一種學生群體而言，不會因不熟習題目中某些與主要測量目標無干的詞句，而影響某群體答對該題的機會。如果與主要測量目標無關的詞句之干擾，而影響某群體的受試者在某一題的答對機會，測驗的內部效度就會降低（楊孟麗、譚康榮、黃敏雄，2001）。那麼，測驗是否具有公平性，即受到質疑，其結果亦造成研究推論是否合理。因此，本研究試圖運用TASA資料庫進行，新移民子女數學成就差別試題功能的檢定，在檢定前，需瞭解其數學評量架構，以下加以述之。

三、數學評量架構

2007年臺灣學生數學學習成就評量國民小學四年級的測驗內容，主要遵循教育部所公布之「國民中小學九年一貫數學學習領域課程綱要」（簡稱課程綱要）能力指標（教育部，2003）所制定，側重於日常生活中學生數學經驗的連結，而其評量的架構係由數學內容領域（content domain）與數學能力（math ability）兩個向度所建構而成，進而再輔以各類等化設計概念以進行施測（國家教育研究院籌備處，2007a）。

TASA2007數學科的測驗內容，其題目的設計是為了有效涵蓋所有數學課程內容與能力，因此，在進行施測之題目通常需高達數百題，但這對於受試樣本於時間、體力上的負荷多過於龐大，因此，在施測上為求便利，多會於組卷時採用試題等化（item equating）概念進行題目的安排，就TASA2007

而言，除需滿足等化能力估計目的外，為有效連結2005、2006年評量的結果，乃預先將全部數學試題分成10個題本（Booklet）（以B1-B10表示），而其下每個題本則包含有24題選擇題與2題應用題，但每個題本之間，僅有3分之2的題目是屬2007年新編試題（以M1-M10表示），其中，保留約3分之1的題目是從TASA2005、2006抽取之具高鑑別度優良試題，以作為各個題本之共同定錨試題（anchor item），以便共同能力的估計（國家教育研究院籌備處，2007b年，3月）。但有鑑於各題本間並未存在應用題之等化試題，且為簡化分析的複雜性，於後續分析上，本研究是僅以選擇試題作為數學能力之估計。

四、研究目的

根據前述動機和評閱相關文獻之後，本研究目的為：（一）分析2007年臺灣學生學習成就評量資料庫之新移民子女與本土子女之數學學習成就差異。（二）探討2007年臺灣學生學習成就評量資料庫數學科試題可能存在之新移民子女與本土子女的差別試題功能。

貳、研究方法

一、研究樣本

本研究所分析之新移民子女樣本，主要源自於TASA資料庫，經研究者向承辦單位申請後取得，而其整體的概念性抽樣架構大致如圖1所示。首先，運用分層叢集隨機抽樣自台灣25縣市所組成之TASA樣本，從中精確界定是屬於新移民子女者（即來自大陸地區（含港澳）及東南亞地區之移民女性，與具中華民國國籍之台灣地區出生男性所共同孕育之子女），分別獨立出來以組成新移民子女樣本，而同時為有效回答本研究之各項待答問題，另依據TASA抽樣架構中之各分層學校內之班級，分別從中隨機抽取相對應人數之本土學生為配對樣本，以形成本研究分析新移民子女學習成長表現之樣本。

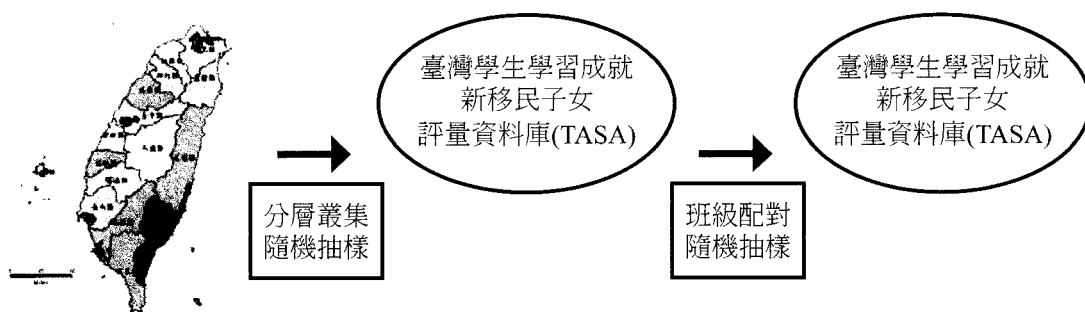


圖1 本研究分析TASA2007新移民子女學習成長資料之抽樣概念架構圖

TASA2007主要是針對國小四年級、六年級為目標母群（target population），採二階段隨機抽樣設計（臺灣學生學習成就評量資料庫網站，2007）。由於TASA資料庫是以抽取具全國代表性樣本為目的，並非特立以新移民子女為標的，因此，本研究分析之新移民子女樣本，是為從中獨立出組成之新移民子女樣本，則不能完全視之等同於採用二階段隨機抽樣於新移民族群者。但有鑑於該抽樣設計概念乃著眼於依據各縣市人口規模、學校規模

的差異性，有效的調整各區域實際應抽取之樣本比率，同樣促使本研究分析樣本具備分層抽取的特性，使新移民子女樣本能有效的涵蓋全國25縣市，減少可能因城鄉差距、區域限制等因素干擾分析結果，以能反映出新移民子女族群的特性。

此外，研究者為分析本土與新移民子女之數學成就表現差異，本研究另依TASA資料庫第一階段之分層隨機抽樣抽取出之樣本學校，分別計算學校內各班級之新移民子女數，再根據此訊息，隨機於該班級內抽取相對應之本土子女人數作為配對樣本，而若該班級內並未有充足本土子女作為對照，則於配對完畢後，隨機於該學校剩餘本土子女樣本中隨機抽取相對應之人數，此舉乃是希望於分析本研究之差異比較問題時，能減少無關干擾因素的影響。如此，本研究是以TASA資料庫中國小四年級、六年級學生為標的，分別抽取新移民子女與本土配對樣本，以進行分析與比較新移民子女數學學習表現。經分析，為因應本研究之分析目的，樣本型態可分為以下幾類：

（一）TASA2007數學科全部分析樣本

鑑於本研究所運用於估計學生數學能力之技術，乃是採試題反應理論之單向度Rasch模式，而為求其估計之準確性，是需要大量樣本作為分析之基礎，因此，本研究所採用於後續分析之第一種樣本，稱之為TASA2007數學科全部分析樣本，是以有接受TASA2007數學科之全部分析樣本，作為數學科成就測驗各試題之原始參數值的估計，進而，是以此作為定錨試題（即固定試題參數值），以進行新移民子女能力（即下述之數學成就配對樣本）之估計。此外，對於單向度的檢測、模式適配度之分析，亦是以此樣本為基準，而再經刪除未作答者，經分析，可得國小四、六年級有效樣本，分別是8218人、8215人。

（二）DIF分析樣本

在進行如上述數學成就測驗試題之原始參數值估計前，為避免試題

內之差別試題功能影響本土與新移民子女之能力比較結果，本研究預先採用DIF分析樣本進行差別試題功能的分析，而此樣本類型的組成，主要是選擇有參與數學科成就測驗之TASA2007數學科全部分析樣本，進而將之分為新移民子女與本土子女兩大族群，而排除其它非屬此兩族群者。經分析後，國小四年級可得有效本土子女樣本6479人、新移民子女278人；國小六年級可得有效本土子女樣本7373人與新移民子女184人。

（三）數學成就配對樣本

為有效比較本土與新移民子女之數學成就表現，本研究運用之第三類樣本，稱之為數學成就配對樣本，是以全部參與TASA2007評量之國小四年級、六年級，分別各17154名、21227人中，再經前述班級配對抽樣選取性別一致，所得之本土與新移民子女配對樣本，接續，再排除未接受數學成就測驗之學生後，所得之本土與新移民子女數學成就配對樣本，經分析後，國小四年級可得有效本土子女樣本271名，新移民子女273名；國小六年級可得有效本土子女樣本180名與新移民子女181名。是以作為比較新移民與本土子女之數學成就表現。

二、資料分析

資料分析方面，茲分幾個面向，分別是單向度檢測、模式適配度、差別試題功能與多元潛在迴歸分析，陳述如下：

（一）單向度檢測

鑑於TASA2007對於數學科試題之組卷，如採等化連結設計，意即全部受試者僅接受部份的試題，其間是利用預先作為共同定錨之試題進行能力等化，如此會產生出許多作答反應是屬研究設計造成的完全隨機缺失值（missing completely at random），而此現象所影響的則是投入上述方法進行

分析之四分或PHI相關的計算、正定（positive definite）問題與後續特徵值抽取的不穩定等（Beaton, 1985; Hulin, Drasgow, & Parsons, 1983）。鑑於此，有別於相關係數，本研究乃是採以針對試題反應型態（item response patterns）進行分析之充分訊息因素分析法（full information factor analysis）為考量，在運用TESTFACT4.0向度分析軟體（Wood, Wilson, Gibbons, Schilling, Muraki, & Bock, 2003），採邊際最大概似估計法（marginal maximum likelihood estimation）與E-M演算法進行因素負荷量、特徵值等之估計，以增強分析結果之可信度。此外，並搭配著Lord（1980:21）的建議：第一特徵值比第二特徵值大很多，且第二特徵值並未比其它特徵值大很多的判斷標準，以進行單向度的檢驗。

（二）單向度Rasch模式、模式適配度與多元潛在迴歸分析

在上述TASA2007受試者數學成就資料若能符合單向度假設前提下，在參考過去教育資料庫（如TIMSS2003）曾採用過之分析模式與考量本研究研究目的、研究軟體取得等因素下，進而預先決議是以單向度Rasch模式或稱單參數試題反應模式以執行本土與新移民子女數學能力之估計，並以ConQuest2.0電腦軟體（Wu, Adams, & Wilson, 2007）進行統計分析。在上述試題反應模式與母群模式兩相結合後，本研究對於本土與新移民子女之數學成就比較探討，即可運用受試者的試題反應資料與背景特質，直接以ConQuest軟體執行多元潛在迴歸模式（multiple latent regression model），同時，再搭配著邊際最大概似估計法與E-M演算法後，以針對各參數值進行估計（Wu, Adams, & Wilson, 2007）。

其次，在模式適配度分析上，根據Linacre（2006）認為當MNSQ（mean squares）值是可接受時，則可忽略t值，而其建議是認為當MNSQ大於2時，表示該試題將會扭曲或破壞測量系統；MNSQ介於1.5~2之間時，表示該試題雖對測量的建構不具生產性，但也不具破壞性；MNSQ在0.5~1.5之間時，是為該試題對測量具生產性；MNSQ小於0.5時，表示該試題對測量有較少生

產性。概括而言，是以MNSQ介於0.5與2範圍內時，是為本研究中可接受之適配現況。

（三）差別試題功能分析

在檢定向度的假設與資料一模式適配程度後，本研究對於TASA2007數學科成就資料關心的另一項議題，就是差別試題功能（Differential Item Functioning, DIF）分析，意含著試題對於不同的群體發揮了不同的功用（Holland & Thayer, 1988），而其定義能為目前大多數學者所接受的則是：「來自不同的團體或族群，但能力相同的個人，在某試題的作答機率上卻有所不同，則該試題便具有DIF現象。」（余民寧，2009）。相對本研究目的，則是將新移民子女稱為焦點團體（focal group），以作為研究中所感興趣的族群，而本土子女則稱之為參照團體（reference group），以視為對照、比較的族群，進而，在針對兩族群能力進行均等控制後，以檢定族群間的作答機率是否有差異，此則為DIF分析的粗略概念，而其目的則是在排除後續於族群間進行能力比較時，可能因試題產生的干擾因素。

本研究主要是IRT取向的Rasch模式概率比法（likelihood ratio test）（王文中、張智宏，1998），其概念主要是承續於Thissen, Steinberg與Gerrard（1986）概率比法，但有別於過去的研究，更著重的是直接估計DIF效果，並採逐題檢定增大模式與精簡模式中資料的概率差異，其作法主要是比較增大模式（augment model）與精簡模式（compact model）間之概率統計量的差異，其中，增大模式代表的是假設研究中每一試題皆存在DIF，而精簡模式則是逐題限制DIF參數為0（即逐題假設欲檢定之試題不存在DIF），接續，再透過兩模式之差異比較，以進行顯著性考驗，公式如下所示：

$$\Delta G^2 = G_{compact}^2 - G_{augment}^2 \quad (\text{公式1})$$

其中， $G_{compact}^2$ 、 $G_{augment}^2$ 分別代表精簡模式與增大模式之概率統計量，而其卡方差異值 ΔG^2 ，會近似於卡方分配，若精確說明時，由於研究中是採逐題限制，亦即會等同於自由度為1之卡方分配，如此，再利用其臨界值（3.84，

df=1) 進行顯著性考驗。此外，為避免DIF的效果是來自於兩群體能力差異(即impact影響)，在採用Rasch模式概率比法進行分析時，研究者就直接獨立出參數來代表團體間的平均能力差異，以排除此干擾(王文中、張智宏)。

(四) 多元潛在迴歸分析

有關本研究運用多元潛在迴歸技術的目的，在比較本土與新移民子女在數學成就上之表現差異。實際執行上，在考量研究中進行分析之數學成就配對樣本人數並不多，可能會影響Rasch模式能力估計效果下，乃是預先投入TASA2007數學科全部分析樣本以求得各試題之參數值，進而，在排除具有差別試題功能之試題後，運用剩餘測量指標作為定錨試題(固定其試題參數值)，以進行後續之分析。

為達成上述前者的目的而言，本研究則是在採SPSS13.0統計軟體將原始資料之本土子女虛擬鍵檔為1、新移民子女虛擬鍵檔為0後，搭配數學成就配對樣本與上述定錨參數值，輸入ConQuest2.0軟體，以進行多元潛在迴歸分析，如此，再就估計之迴歸參數值，對照其標準誤，以確立顯著的水準。

參、結果與討論

本研究目的在分析在搭配TASA2007現行既有評量架構下，研究中是針對國小（四、六年級）新移民子女之數學成就進行分析，而在詮釋其影響因素與相關研究方法後，以進行本研究議題之分析，茲將相關結果與討論，陳述如下：

一、TASA2007數學成就測驗之向度檢測與模式適配度分析

在進行數學成就資料分析方面，一如前述，TASA2007數學科雖然所測量的是學生的數學成就，其下子領域是包含有數與量、幾何、代數、統計與機率與連結等五大主題，因此，為確保資料分析的準確度，則需於正式分析前預先確立本研究單向度模式的假設，進而再確立資料與Rasch模式之適配程度。

在單向度假設分析上，再投入TASA2007數學科全部分析樣本，經採TESTFACT向度分析軟體，運用充分訊息因素分析法（full information factor analysis）估計後，其檢測結果如表1所示，不論是TASA2007國小四年級或六年級之數學成就測驗，其最大特徵值乃分別為第二大特徵值的15.61與18.24倍，而第二大特徵值分別為第三大特徵值的1.51與1.66倍，大致符合Lord（1980, p.21）所認為：第一特徵值（latent root）比第二特徵值大很多，且第二特徵值並未比其它特徵值大很多的單向度考驗標準，顯示TASA2007國小數學科測驗資料多能符合單向度之考驗，且在此單向度下則分別能解釋整份測驗之24.244%與30.866%之解釋變異量。

表1 TASA2007國民小學數學成就測驗之單向度檢測一覽表

主成份	年級	
	國小四年級	國小六年級
第一特徵值	24.715	31.532
第二特徵值	1.583	1.729
第三特徵值	1.050	1.043
第四特徵值	0.987	0.710
第五特徵值	0.803	0.636
第六特徵值	0.750	0.580

在確立分析資料之單向度假設後，接續針對資料與Rasch模式之適配程度進行檢測，同樣再投入TASA2007數學科全部分析樣本，採ConQuest軟體，經估計後，可發現在檢測國小四年級數學科成就測驗適度方面，不論是未加權（unweighted）或加權（weighted）的適配指標MNSQ（mean squares）值，再經轉換後求得之檢定t值分佈差異十分大，最小可近似到0（第63題），其MNSQ值為0.98，最大則為25.7（第1題），其MNSQ值為1.46。而國小六年級分析結果，最小同樣可近似到0（第87題），其MNSQ值為1.26，最大則為22.6（第7題），其MNSQ值為1.39。其中，雖然兩者都或多或少摻雜某些試題之轉換後t值是大於一般顯著標準2以上，但若在根據Linacre（2006）對MNSQ合理範圍的建議，認為MNSQ在0.5與2內時，該試題多半不會產生嚴重扭曲測量的影響力，發現全部試題幾乎均能符合此準則，顯示TASA2007國小數學科測驗資料，在某程度以上是符合Rasch模式。綜合上述，在單向度假設成立基礎下，後續分析中，乃以採用單向度Rasch模式進行數學科成就測驗之分析。

二、TASA2007國小數學科成就測驗之差別試題功能分析

在運用單向度Rasch模式進行TASA2007國小數學科成就測驗資料分

析前，為確保在比較新移民與本土子女差異表現分析之準確性，本研究乃事先檢定各試題內是否存在差別試題功能，意即若兩族群間存在顯著數學成就差異時，可排除並非肇因於試題本身所產生具備較有利於某一族群的因素干擾。而在研究中，對於差別試題功能之分析，主要採用以IRT為基礎之分析方法，首先是採用DIF參數Z檢定法，接續則是以修訂Thissen, Steinberg與Gerrard（1986）的Rasch模式概率比法（王文中、張智宏，1998）進行分析。研究者在投入DIF分析樣本，運用ConQuest軟體進行估計，並經由兩族群之平均數差異參數估計以反映能力分佈的平均差異後，TASA2007數學科國小四與六年級分析結果分別如表2、表3所示，在Z檢定法方面，國小四年級中第10、25、28、46、50、60、74題與國小六年級中第15、42、63、76、81等試題，皆大於0.05顯著水準下之絕對值1.96臨界值，但由於本研究進行DIF分析之新移民子女樣本並不多，且本土與新移民子女樣本數差異較大，多會影響DIF參數估計之穩定性，如出現在表2中國小四年級第66題之DIF參數為4.479，而標準誤為2.806等皆過大的結果。有鑑於此，研究中則另搭配概率比法以期有效避免如焦點組與參照組樣本比率差異大之因素干擾（Lei, Chen, & Yu, 2006）。

在執行Rasch模式概率比法方面，經分析後，國小四年級與六年級原始模式之概率比統計量分別為175641.81與200016.27，兩者各共估計出193個參數，而在研究過程中，則將此模式視為增大模式，接續，再逐一限制各試題之DIF參數為0，並同時開放估計其餘各試題之DIF參數，以形成精簡模式，例如以表2中國小四年級第1試題之分析結果為例，則是將該試題之DIF參數設定為0，同時仍開放估計其它試題之DIF參數，因此，在總共估計192個參數下，此精簡模式之概率比統計量則為175641.88，進而，與上述國小四年級增大模式之概率比統計量兩相比較後，可得 $\Delta G^2 = 0.07$ 或開根號後之 $\sqrt{\Delta G^2} = 0.26$ ，而在對照0.05顯著水準下，自由度相差（ Δdf ）為1的卡方分配值，其 ΔG^2 臨界值為3.84，或經開根號後所反應與Z值雷同的 $\sqrt{\Delta G^2}$ 臨界值1.96，皆可發現第1試題是屬未具備顯著的差別試題功能。

本研究在經逐題分析後，綜整結果如表2、3所示，可發現主要是在國小

四年級第18、46、50、60、74題與國小六年級之第15、42、63、76、81等試題，皆展現出顯著的差別試題功能，若與上述Z檢定法分析結果兩相比較，於國小六年級方面所檢定出之結果則屬完全雷同，但於國小四年級中，則有少數試題略顯差異，多數是於Z檢定法中顯著，但於概率比法中卻是位於臨界附近以下（如第10、25、28題）或是標準誤估計過大致使結果不顯著（如第18題）等，而本研究在考量各方法對於受樣本因素影響之優勢，研究中仍是以Rasch模式概率比法之分析結果為準則。此外，在檢視上述顯著試題對於族群之偏好現況，可發現所有DIF試題均如圖2國小四年級數學測驗之第18題差別試題功能所示，圖中之上方平滑線是屬新移民子女的作答ICC曲線，下方則是屬本土子女作答曲線，可明顯看出是屬於一致性（uniform）DIF試題，亦即各個試題均是一致性的偏好某一族群，若以第18題為例，該試題則是顯著偏好新移民子女（DIF參數=-0.872），顯示對於同等能力之新移民與本土子女而言，該試題之難度對新移民子女，是顯著較本土子女簡單。

表2 TASA2007 國民小學四年級數學成就測驗之差別試題功能分析一覽表

題號	DIF	標準誤	Z值	ΔG^2	$\sqrt{\Delta G^2}$	題號	DIF	標準誤	Z值	ΔG^2	$\sqrt{\Delta G^2}$
1	0.007	0.122	0.06	0.07	0.26	49	0.347	0.262	1.32	1.91	1.38
2	-0.050	0.070	-0.71	0.57	0.75	50	-0.983	0.400	-2.46*	8.89	2.98*
3	0.116	0.070	1.66	2.41	1.55	51	0.155	0.247	0.63	0.48	0.69
4	-0.133	0.069	-1.93	2.07	1.44	52	-0.420	0.224	-1.88	3.42	1.85
5	0.142	0.074	1.92	3.14	1.77	53	0.021	0.219	0.10	0.02	0.14
6	-0.041	0.070	-0.59	0.38	0.62	54	-0.184	0.245	-0.75	0.50	0.71
7	0.040	0.071	0.56	0.34	0.58	55	0.090	0.230	0.39	0.20	0.45
8	0.089	0.076	1.17	1.19	1.09	56	0.066	0.230	0.29	0.16	0.40
9	0.124	0.075	1.65	2.34	1.53	57	0.332	0.255	1.30	1.67	1.29
10	-0.172	0.074	-2.32*	3.21	1.79	58	0.062	0.210	0.30	0.14	0.37
11	-0.141	0.084	-1.68	1.60	1.26	59	-0.037	0.246	-0.15	0.09	0.30
12	-0.069	0.099	-0.70	0.57	0.75	60	0.449	0.206	2.18*	4.65	2.16*
13	-0.088	0.071	-1.24	0.47	0.69	61	0.039	0.216	0.18	0.04	0.20
14	-0.012	0.070	-0.17	0.16	0.40	62	0.094	0.204	0.46	0.23	0.48
15	-0.095	0.072	-1.32	0.63	0.79	63	-0.006	0.215	-0.03	0.07	0.26
16	-0.083	0.082	-1.01	1.10	1.05	64	-0.333	0.203	-1.64	2.55	1.60

17	0.009	0.208	0.04	0.04	0.20	65	0.157	0.255	0.62	0.45	0.67
18	-0.872	0.491	-1.78	5.15	2.27*	66	4.479	2.806	1.60	0.52	0.72
19	0.122	0.241	0.51	0.24	0.49	67	0.072	0.234	0.31	0.11	0.33
20	0.286	0.296	0.97	0.74	0.86	68	0.054	0.253	0.21	0.07	0.26
21	0.133	0.270	0.49	0.22	0.47	69	-0.110	0.233	-0.47	0.25	0.50
22	-0.523	0.386	-1.35	2.81	1.68	70	-0.030	0.264	-0.11	0.07	0.26
23	0.168	0.325	0.52	0.28	0.53	71	-0.001	0.279	0.00	0.04	0.20
24	-0.092	0.333	-0.28	0.12	0.35	72	-0.197	0.236	-0.83	0.72	0.85
25	-0.506	0.239	-2.12*	3.74	1.93	73	-0.207	0.274	-0.76	0.64	0.80
26	0.082	0.255	0.32	0.27	0.52	74	-0.546	0.273	-2.00*	4.87	2.21*
27	-0.263	0.225	-1.17	1.15	1.07	75	-0.170	0.224	-0.76	0.63	0.79
28	-0.441	0.224	-1.97*	3.56	1.89	76	-0.191	0.225	-0.85	0.86	0.93
29	0.213	0.225	0.95	1.27	1.13	77	0.040	0.275	0.15	0.01	0.10
30	-0.126	0.236	-0.53	0.20	0.45	78	-0.041	0.223	-0.18	0.11	0.33
31	-0.205	0.223	-0.92	0.64	0.80	79	0.041	0.233	0.18	0.02	0.14
32	-0.080	0.227	-0.35	0.05	0.22	80	-0.268	0.224	-1.20	1.55	1.24
33	0.084	0.351	0.24	0.18	0.42	81	-0.143	0.237	-0.60	0.50	0.71
34	-0.260	0.230	-1.13	1.03	1.01	82	-0.191	0.238	-0.80	0.84	0.92
35	-0.072	0.275	-0.26	0.08	0.28	83	-0.181	0.247	-0.73	0.74	0.86
36	0.261	0.206	1.27	2.12	1.46	84	-0.182	0.264	-0.69	0.68	0.82
37	0.125	0.223	0.56	0.58	0.76	85	0.111	0.248	0.45	0.19	0.44
38	-0.374	0.204	-1.83	3.06	1.75	86	-0.286	0.278	-1.03	1.45	1.20
39	-0.182	0.197	-0.92	0.66	0.81	87	0.206	0.266	0.77	0.43	0.66
40	0.004	0.199	0.02	0.13	0.36	88	0.223	0.239	0.93	0.71	0.84
41	-0.271	0.214	-1.27	1.41	1.19	89	0.268	0.257	1.04	0.87	0.93
42	-0.015	0.274	-0.05	0.00	0.00	90	0.603	0.407	1.48	2.36	1.54
43	-0.079	0.209	-0.38	0.10	0.32	91	0.014	0.219	0.06	0.03	0.17
44	0.269	0.208	1.29	1.99	1.41	92	-0.224	0.218	-1.03	1.36	1.17
45	-0.154	0.207	-0.74	0.48	0.69	93	0.109	0.219	0.50	0.12	0.35
46	-0.473	0.231	-2.05*	4.15	2.04*	94	0.394	0.230	1.71	2.64	1.62
47	-0.411	0.241	-1.71	2.81	1.68	95	0.140	0.217	0.65	0.24	0.49
48	0.038	0.208	0.18	0.09	0.30	96	0.342	0.258	1.33	1.52	1.23

註：表中最後1題乃作為定義量尺之試題，以解決參數的不定性，而該題之對應值乃是以倒數第2題作為量尺定義試題，重新計算之結果。

* $p < .05$ 。

表3 TASA2007 國民小學六年級數學成就測驗之差別試題功能分析一覽表

題號	DIF	標準誤	Z值	ΔG^2	$\sqrt{\Delta G^2}$	題號	DIF	標準誤	Z值	ΔG^2	$\sqrt{\Delta G^2}$
1	-0.046	0.088	-0.52	0.31	0.56	49	-0.223	0.266	-0.84	0.61	0.78
2	-0.136	0.099	-1.37	2.25	1.50	50	0.231	0.254	0.91	0.80	0.89
3	0.007	0.088	0.08	-0.22	0.47	51	0.146	0.265	0.55	0.03	0.17
4	-0.086	0.113	-0.76	0.64	0.80	52	-0.148	0.270	-0.55	0.14	0.37
5	0.008	0.090	0.09	-0.19	0.44	53	-0.144	0.265	-0.54	0.06	0.24
6	0.112	0.093	1.20	0.71	0.84	54	-0.095	0.303	-0.31	-0.04	0.20
7	-0.099	0.093	-1.06	1.38	1.17	55	-0.081	0.261	-0.31	-0.15	0.39
8	-0.043	0.096	-0.45	0.17	0.41	56	-0.248	0.391	-0.63	0.20	0.45
9	-0.004	0.090	-0.04	-0.15	0.39	57	0.420	0.273	1.54	2.18	1.48
10	-0.039	0.088	-0.44	0.23	0.48	58	0.101	0.273	0.37	-0.06	0.24
11	0.075	0.088	0.85	0.23	0.48	59	0.155	0.268	0.58	0.19	0.44
12	-0.059	0.088	-0.67	0.49	0.70	60	0.008	0.296	-0.03	-0.04	0.20
13	0.003	0.090	0.03	-0.14	0.37	61	-0.513	0.314	-1.63	2.64	1.62
14	-0.008	0.089	-0.09	-0.07	0.26	62	-0.103	0.565	-0.18	-0.20	0.45
15	-0.197	0.089	-2.21*	5.62	2.37*	63	0.783	0.268	2.92*	8.89	2.98*
16	0.151	0.095	1.59	1.71	1.31	64	0.229	0.268	0.85	0.82	0.91
17	0.362	0.253	1.43	1.43	1.20	65	-0.435	0.573	-0.76	0.00	0.00
18	0.108	0.309	0.35	-0.20	0.45	66	-0.037	0.251	-0.15	-0.15	0.39
19	-0.146	0.344	-0.42	-0.05	0.22	67	-0.584	0.408	-1.43	1.94	1.39
20	-0.365	0.257	-1.42	1.74	1.32	68	-0.117	0.250	-0.47	-0.10	0.32
21	-0.519	0.275	-1.89	3.71	1.93	69	-0.053	0.251	-0.21	-0.17	0.41
22	0.262	0.259	1.01	0.85	0.92	70	0.029	0.266	0.11	-0.17	0.41
23	-0.331	0.266	-1.24	1.29	1.14	71	-0.170	0.257	-0.66	0.16	0.40
24	0.149	0.249	0.60	0.27	0.52	72	-0.083	0.291	-0.29	-0.18	0.42
25	-0.287	0.303	-0.95	0.32	0.57	73	0.197	0.298	0.66	0.31	0.56
26	-0.411	0.320	-1.28	0.88	0.94	74	0.032	0.260	0.12	-0.01	0.10
27	-0.189	0.342	-0.55	-0.17	0.41	75	0.151	0.256	0.59	0.19	0.44
28	-0.229	0.303	-0.76	0.04	0.20	76	-0.749	0.260	-2.88*	7.30	2.70*
29	-0.167	0.313	-0.53	-0.05	0.22	77	-0.062	0.259	-0.24	-0.09	0.30
30	0.358	0.338	1.06	2.31	1.52	78	0.087	0.256	0.34	-0.04	0.20
31	-0.242	0.381	-0.64	-0.14	0.37	79	0.398	0.257	1.55	2.11	1.45

32	-0.038	0.303	-0.13	-0.12	0.35	80	-0.401	0.256	-1.57	2.41	1.55
33	-0.210	0.296	-0.71	0.13	0.36	81	0.928	0.384	2.42*	5.83	2.41*
34	0.449	0.301	1.49	3.15	1.77	82	0.518	0.309	1.68	1.86	1.36
35	0.142	0.296	0.48	0.43	0.66	83	-0.330	0.309	-1.07	2.29	1.51
36	0.463	0.318	1.46	3.14	1.77	84	-0.206	0.325	-0.63	1.27	1.13
37	0.441	0.287	1.54	2.95	1.72	85	0.285	0.313	0.91	0.14	0.37
38	-0.209	0.287	-0.73	0.08	0.28	86	0.468	0.310	1.51	1.06	1.03
39	-0.043	0.287	-0.15	-0.09	0.30	87	-0.326	0.318	-1.03	2.40	1.55
40	0.145	0.321	0.45	0.20	0.45	88	0.191	0.318	0.60	0.01	0.10
41	0.426	0.294	1.45	2.19	1.48	89	-0.081	0.277	-0.29	0.52	0.72
42	-0.827	0.296	-2.79*	7.33	2.71*	90	-0.078	0.308	-0.25	0.55	0.74
43	0.195	0.294	0.66	0.24	0.49	91	0.593	0.630	0.94	2.06	1.44
44	-0.063	0.320	-0.20	-0.11	0.33	92	-0.391	0.282	-1.39	2.95	1.72
45	0.099	0.313	0.32	-0.12	0.35	93	-0.291	0.282	-1.03	1.95	1.40
46	0.088	0.294	0.30	-0.09	0.30	94	0.045	0.284	0.16	-0.23	0.48
47	0.184	0.294	0.63	0.06	0.24	95	0.325	0.277	1.17	0.57	0.75
48	0.245	0.296	0.83	0.41	0.64	96	0.140	0.290	0.48	-0.44	0.66

註：表中最後1題乃作為定義量尺之試題，以解決參數的不定性，而該題之對應值乃是以倒數第2題作為量尺定義試題，重新計算之結果。

* $p < .05$

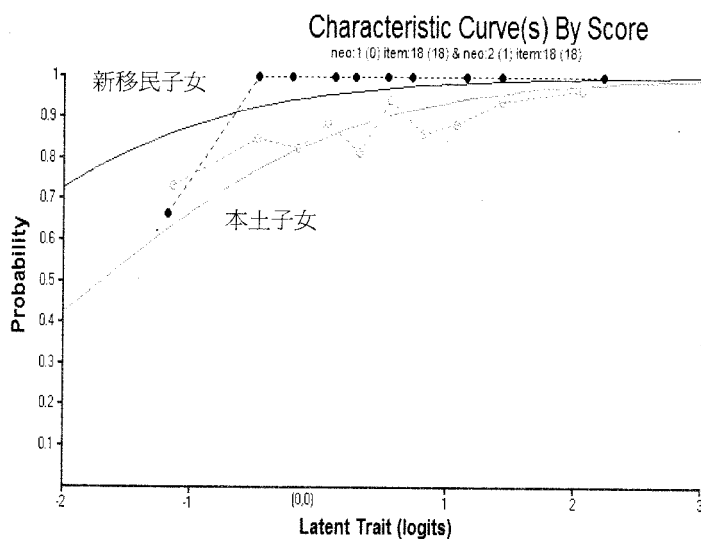


圖2 TASA2007國小四年級數學科測驗第18題差別試題功能示意圖

三、TASA2007新移民子女數學表現現況及與本土子女之差異比較

在分析TASA2007新移民子女數學成就方面，在考量本研究進行比較之本土與新移民子女數學成就配對樣本並不多，為避免此因素影響單向度Rasch模式執行能力估計的穩定性下，本研究是投入TASA2007國民小學數學科全部分析樣本，採用ConQuest軟體，以預先估計出各試題參數值，接續，刪除前述各年級中具顯著差別試題功能之試題，將剩餘試題之估計參數值另作為定錨試題（即固定參數值），以運用於本土與新移民子女數學成就配對分析樣本中進行Rasch能力的估計，而經分析後，就產出之EAP能力值進行描述統計，其分析結果是如表4所示，可發現不論是國小四或六年級，本土子女之平均表現多是較新移民子女為佳（0.5873大於0.2804；0.3719大於0.2499），而分數變動程度則約相當，但此差異是否達到顯著水準，則是接續檢定如下。

表4 TASA2007 國小四、六年級本土、新移民子女於數學科成就EAP能力值之描述統計一覽表

年級	比較	N	數學科成就測驗			
			平均數	標準差	偏態	峰度
國小四年級	本土子女	271	0.5873	0.7269	-0.066	-0.495
	新移民子女	273	0.2804	0.7397	0.039	-0.327
國小六年級	本土子女	180	0.3719	1.0048	-0.009	-0.235
	新移民子女	181	0.2499	1.0964	0.375	-0.676

在進行新移民與本土子女數學成就表現之差異比較檢定前，本研究是預先採用EAP數學能力值，針對來自不同國家之新移民子女，進行Levene變異數同質性考驗，經分析後，可發現國小四與六年級之檢定值分別是0.813（ $p=.592$ ）、1.169（ $p=.325$ ），皆未達顯著的水準，代表著來自不同國家之新移民子女於數學成就表現的變異程度並未有太大歧異性，因此，研究中是適合將其整體合併為單一新移民族群，藉以與本土子女作比較。接續，為

期能獲得較穩定、準確的比較分析結果，研究中則並非是藉由EAP數學能力值，再運用其它統計軟體（如SPSS）進行獨立樣本t考驗等二階段的方式進行差異分析（Wu, Adams, & Wilson, 2007），而是一貫的運用ConQuest軟體，預先固定試題參數值，以投入數學成就配對樣本進行潛在迴歸模式分析，其分析結果如表5所示，國小四年級所估計出之迴歸係數是為0.307，若相對於其標準誤，可計算出其檢定t值為4.21（ $p < .01$ ），是達顯著差異水準，對照上表4，本土子女數學成就表現平均是顯著大於新移民子女0.307分，其效果量為0.42，進而，在根據Cohen（1992）認為d等於0.2、0.5與0.8時，分別是小、中、大效果量的建議，此結果則是屬介於中至小的差異效果量。此外，在國小六年級的族群檢定方面，則是出現不同於上述的結論，是屬未達顯著的差異。

整體而言，在對照過去的研究，分析結果大致是雷同於柯淑慧（2005）、張慧貞（2005）在國小數學成就研究上所發現之顯著差異結論，但本研究猜測其中亦可能存在著年級的因素，亦即隨著年紀的上升，學童身心亦會愈趨成熟，致使於國小六年級時的表現愈趨穩定，而展現出無顯著差異的成就表現，但此論點是否成立，仍需待未來長期縱貫資料之佐證。最後，就小學四年級分析結果而言，雖然研究中顯示出中等至小的顯著差異效果量，雖然本研究已採配對抽樣方式，以避免可能無關干擾，但此差異亦可能是伴隨著其它因素之影響，而非完全是先天消極的族群因素所致，有待未來研究加以探討。

**表5 TASA2007國小四、六年級新移民與本土子女數學成就比較之
潛在迴歸分析**

比較	常數項	迴歸係數 (標準誤)	t值	p值
國小四年級 本土子女－ 新移民子女	0.281	0.307 (0.073)	4.21	<.01
國小六年級	0.248	0.122 (0.121)	1.01	.314

接續，在探討控制本土、新移民學生家庭社經地位時，族群間數學成就差異，研究者同樣是在投入數學成就配對樣本，並就產出之家庭社經地位主成份因素分數以進行多元潛在迴歸分析，首先，在預先檢定不同族群與社經地位之交互作用，可發現國小四年級之交互作用迴歸參數檢定結果為0.157 (0.097)，並未達顯著水準，因此，進而可單純分析其主要效果，以檢定族群與家庭社經地位同時對數學成就之影響。分析結果如表6所示，可發現在排除社經地位的影響後，國小四年級本土與新移民子女間之數學表現差異，由原先表5的顯著效果 ($t = 4.21, p < .01$) 轉變為表6之不顯著結果 ($t = 1.79, p = .07$)，顯示除了族群的因素外，家庭社經地位的不平等也許更是干擾族群間學生表現歧異的主因。

整體而言，在過去的研究（王振世、蔡清中，2008；張芳全主編，2007；陳玉娟，2009），都認為家庭社經地位是會影響新移民子女學習成就，但本研究更進一步指出在控制此因素的干擾下，新移民子女與本土子女的數學成就表現，是完全不具差異性，顯示家庭的弱勢可能才是影響新移民子女學習的主要因素。

表6 TASA2007國小四年級族群、家庭社經地位與數學成就表現關聯之多元潛在迴歸分析

年級	變項	常數項	迴歸係數(標準誤)	t值	p值
國小四年級	本土-新移民	0.467	0.174(0.097)	1.79	.07
	家庭社經地位		0.196(0.049)	4.00	<.01

肆、結論與建議

一、研究發現

首先，在向度假設檢定方面，不論是國小四或六年級樣本，分析結果皆能充分展現出單向度之特徵。進而，再針對資料與研究中採行之Rasch模式進行適配度估計時，各試題分析結果亦均能符合Linacre（2006）對MNSQ建議之合理範圍內，顯示TASA2007國小數學科成就測驗是適用於單向度Rasch模式。接續，為避免評量試題中，因歸屬族群之不同而產生差異的試題功能等，以影響能力估計之準確，本研究是採Rasch模式概率比法（王文中、張智宏，1998）進行DIF的檢定，顯示是以國小四年級第18、46、50、60、74題與國小六年級之第15、42、63、76、81題等指標，出現顯著的同能力者但卻因屬不同族群而產生的正確作答差異，進一步檢視時，多是屬於一致性的偏好某一族群之評量試題。

鑑於此，於後續分析，是將此類試題全部刪除，以進行族群間之數學成就表現比較，經採多元潛在迴歸分析，結果顯示在國小六年級內，並未存在顯著族群差異，但是在國小四年級中，是以本土子女的表現平均優於新移民子女，差異效果為0.42，約屬Cohen（1992）認為中等至小的效果量，但在伴隨分析是否存在其它因素影響時，是顯示在先天消極的族群因素外，族群間家庭社經地位的不均衡可能才是真正影響之主因。

二、建議

在過去新移民子女的研究中，因礙於經費、人力限制之影響，其抽取之樣本多具過度集中於某區域、過少或缺乏對照組等缺點，致使在結果推論上，較易受到質疑。鑑於此，本研究是以分析TASA2007資料庫為標的，以期能有效的涵蓋全國各區域之樣本，但由於此資料庫並非是以新移民子女為抽樣之目標母群，因而，在後續進行有關數學成就之分析時，會產生人數過少之現象，肇使如國中二年級之新移民子女樣本僅有78名，而無法納入本研究分析。因此，在考量節省人力、財力資源下，建議能在既有TASA2007資料

庫之抽樣架構，擴大對國小四、六與國中二年級新移民子女之抽樣，是可在決定抽取之學校數與學生數後，於執行原訂之簡單隨機抽樣外，另要求被選取學校之全部新移民子女亦需參與評量，亦或是事先要求學校提供每位學生是否為新移民子女之資訊，以預先納入樣本的選取。如此，不僅能保持樣本有效涵蓋全國各區域之特性，亦能增加樣本數，以使後續分析更為穩定。接續，建議可搭配本研究採行之班級配對抽樣設計，以進行本土與新移民子女之各項重要議題探討。

其次，TASA國資料庫的成就測驗，應逐年針對試題是否有DIF的問題進行檢定，並用不同的方法檢定，將有DIF的試題找出，並進一步請學科專家進行DIF試題內容的檢定，若到最後DIF試題被檢定為偏誤，就應將偏誤題刪除並統計歸類，避免不公情形的出現，並瞭解試題的內容與性質及可能導致DIF的原因，提供作為資料庫試題編擬的修正，以維持測驗的穩定性，確保測驗的公正性與測驗的價值性。

再者，研究發現，國小四年級中，本土子女的表現平均優於新移民子女表現，顯示族群間家庭社經地位的不均衡可能會造成學習成就的差異，但國小六年級兩族之間卻沒有差異。因此除了在試題編製上減少DIF問題，教育單位應透過各種不同的補救方式及輔導策略，協助新移民子女，將每位孩子都帶上來，如此當其就讀越高年級，其身心發展越趨成熟，懂得如何運用有效策略強化學業，將可減少家庭社經背景的問題，破除負面的刻板印象，成為一位優秀的國民。

DIF 檢視的最終目的，探討區域（族群）間可能的學習成就、文化社經背景等影響因素，但仍限於量方面證據，實質面向上仍有待研究。因此，建議未來研究者或實務工作者，可採實驗方法，理清分組變項間可能真正影響因素，以進行實質面的探討。

參考文獻

- 王文中、張智宏（1998）。Rasch模式概率比法的差別試題功能分析。《中華心理學刊》，40（1），15-32。
- 王世英、溫明麗、謝雅惠、黃乃熒、黃嘉莉、陳玉娟、陳烘玉（2006）。我國新移民子女學習成就現況之研究。《教育資料與研究》，68，137-170。
- 何美瑤（2007）。外籍配偶子女學業成就之研究：以文化資本的觀點探析。高雄師範大學教育研究所博士論文，未出版，高雄市。
- 余民寧（2009）。試題反應理論（IRT）及其應用。臺北市：心理。
- 呂念慈（2007）。六位東南亞籍新移民母親子女教養經驗之研究。實踐大學家庭研究與兒童發展研究所碩士論文，未出版，臺北縣。
- 李嘉莉（2008）。新移民女性母職覺知與實踐之研究：以一個越南籍母親為例。屏東教育大學教育行政研究所博士論文，未出版，屏東縣。
- 那昇華（2007）。新移民子女的學業成就。載於張芳全（主編），新移民子女的教育（頁47-70）。臺北市：心理。
- 周秀潔（2005）。外籍配偶子女的學校適應－教師的覺知與應對。臺灣師範大學教育學系碩士論文，未出版，臺北市。
- 周炳言（2009）。東南亞裔新移民女性之子女學業成就三年追蹤及其與配偶親職角色支持狀況之關聯。臺北教育大學教育心理與輔導學系碩士班碩士論文，未出版，臺北市。
- 林佩蓉（2008）。國小四年級新移民學童之家庭背景數學自信與數學興趣對數學成就的影響－TIMSS 2003次層次資料分析。臺中教育大學數學教育學系碩士班碩士論文，未出版，臺中市。
- 柯淑慧（2005）。外籍母親與本籍母親之子女學業成就之比較研究？以基隆市國小一年級學生為例。臺北師範學院幼兒教育學系碩士班碩士論文，未出版，臺北市。

- 高玉靜（2005）。新移民家庭子女教養問題及因應。載於玄奘大學教育人力資源與發展學系（主編），*外籍配偶與社區學習*（頁79-90）。臺北市：師大書苑。
- 國家教育研究院籌備處（2007a）。*臺灣學生學習成就評量資料庫建置計劃-數學領域：96年命題研習手冊*（小四、小六、國二、高中二、高職二），未出版，臺北縣。
- 國家教育研究院籌備處（2007b年，3月）。*2007年臺灣學生學習成就評量資料庫：試題研發暨統計分析組業務諮詢會議紀錄*（發文字號：教研研字第0960001510號）。未出版，臺北縣。
- 張芳全主編（2007）。*新移民子女的教育*。臺北市：心理。
- 張敏華（2005）。*新台灣之子的媒體形象：外籍配偶子女之新聞框架研究*。中正大學電訊傳播研究所碩士論文，未出版，嘉義縣。
- 張慧貞（2005）。*外籍配偶子女數感表現之研究：以高雄市國小四年級為例*。屏東教育大學數理教育研究所碩士論文，未出版，屏東縣。
- 教育部（2005）。*外籍配偶就讀國小子女學習及生活意向調查結果*。2009年12月23日，取自 http://www.edu.tw/EDU_WEB/EDU_MGT/STATISTICS/EDU7220001/publish/foreign_parent_student_in_basic/brief.pdf。
- 教育部（2008）。*2007國際數學與科學教育成就趨勢調查（TIMSS）臺灣學生表現亮眼 展現我對數學及科學教育努力成果，肯定九年一貫課程改革方向*。2009年12月28日，取自 http://www.edu.tw/news.aspx?news_sn=2241&pages=0
- 教育部（2010a）。*臺閩地區近年嬰兒出生數按生母國籍分*。2009年12月28日，取自 www.edu.tw/files/site_content/B0013/overview77.xls
- 教育部（2010b）。*外籍配偶子女就讀國中小男女學生人數*。2009年12月28日，取自 http://www.edu.tw/files/site_content/b0013/fomas.xls
- 教育部統計處（2007）。*教育部全球資訊網*。2007年6月24日，取自 <http://>

www.edu.tw/

- 許崇憲（2002）。家庭背景因素與子女學業成就之關係：臺灣樣本的後設分析。*中正教育研究*，1（2），25-62。
- 陳玉娟（2009）。新移民子女教育問題、成因及其策略之研究：以國民教育階段為例。*初等教育學刊*，33，33-57。
- 陳柏霖、陳書農（2007）。從多元文化教育觀點論述學校、政府層面對外籍配偶及其子女教育策略。載於玄奘大學社會科學院、教育人力資源與發展學系（主編），*弱勢族群議題省思與對策*（頁189-214）。臺北市：師大書苑。
- 黃森泉、張雯雁（2003）。外籍新娘適應與子女教養問題之探討。*社會科教育研究*，8，135-169。
- 黃琬玲（2005）。*東南亞外籍配偶子女學習適應狀況之研究*。臺北師範學院國民教育研究所碩士論文，未出版，臺北市。
- 楊孟麗、譚康榮、黃敏雄（2001）。*心理計量報告：TEPS 2001分析能力測驗*。中央研究院臺灣教育長期追蹤資料庫。
- 溫明麗（2007）。臺灣人口結構變遷之外籍配偶子女補救教學模式建構。*教育資料與研究*，74，73-96。
- 臺灣學生學習成就評量資料庫網站（2007）。*臺灣學生學習成就評量資料庫*。2007年6月24日，取自<http://tasa.naer.edu.tw/brief.htm>
- 趙善如、鍾鳳嬌、江玉娟（2007）。影響外籍暨大陸配偶學童子女學業成績關鍵因素：以個人人口特徵、學習行為能力、家庭環境探討之。*教育心理學報*，39（1），129-147。
- 劉秀燕（2003）。*跨文化衝擊下外籍新娘家庭環境及其子女行為表現之研究*。中正大學犯罪防治研究所碩士論文，未出版，嘉義縣。
- 蕭彩琴（2005）。*臺中縣外籍配偶之國小低年級子女同儕社會地位與學業成*

- 就之研究。朝陽科技大學幼兒保育科碩士論文，未出版，臺中縣。
- 賴翠媛（2006）。新移民子女就讀國民中小學之學業表現及接受學校相關輔導措施之調查研究。*特教論壇*，1，12-24。
- 鍾文悌（2005）。外籍配偶子女學業表現與生活適應之相關研究。屏東師範學院教育行政研究所碩士論文，未出版，屏東縣。
- Beaton, A. E. (1985). *NAEP analysis procedures and methodology*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.
- Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Harker, K. (2001). Immigrant generation, assimilation, and adolescent psychological well-being. *Social Forces*, 79, 969-1004.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp.129-145). Hillsdale NJ: Lawrence Erlbaum Associates.
- Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Application to psychological measurement*. Homewood, IL: Dow Jones-Irwin.
- Lei, P-W., Chen, S-Y., & Yu, L. (2006). Comparing methods of assessing differential item functioning in a computerized adaptive testing environment. *Journal of Educational Measurement*, 43(3), 245-264.
- Linacre, J. M. (2006). *A user's guide to FACETS: Rasch-model computer program*. Retrieved July 30, 2007 from the world wide web: <http://www.winsteps.com>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453.
- Thissen, D., Steinberg, L., & Gerrard, M. (1986). Beyond group mean differences: The concept of item bias. *Psychological Bulletin, 99*, 118-128.
- Wood, R., Wilson, D., Gibbons, R., Schilling, S., Muraki, E., & Bock, D. (2003). *TESTFACT 4.0*. Chicago: Scientific Software International.
- Wu, M. L., Adams, R.J., & Wilson, M. R. (2007). *ConQuest2: Multi-Aspect Test Software [computer program and computer program manual]*. Camberwell: Australian Council for Education Research.

5A Panel discussion V

親子關係、憂鬱、偏差行為和幸福感結構關係模式之驗證：
以TEPS資料庫資料為例

— 薛凱方

親子關係、憂鬱、偏差行為和幸福結構關係模式之驗證： 以TEPS資料庫資料為例

薛凱方

國立政治大學教育研究所博士候選人
95152502@nccu.edu.tw

摘要

許多研究發現，父母親對孩子表達出較多的情感慰藉、接納及支持，孩子會表現出較高的自尊、快樂與幸福，表現出較少的沮喪和焦慮，因此，良好且具支持性的親子關係，不僅能減少青少年的負向行為，也可增進其身心健康和幸福感。本研究希望能進一步提出模式來驗證相關研究所提出之親子關係對於青少年憂鬱、偏差行為及幸福感的影響情形。模式驗證的樣本取自台灣教育長期追蹤資料庫(TEPS)第一波調查中的國中學生共5,300人，運用結構方程模式統計方法，對本研究提出的模式加以驗證。結果發現(1)親子關係對於個體的憂鬱及偏差行為有負向影響效果；(2)親子關係除了會直接影響幸福感外，也會透過「減緩憂鬱」而間接正向影響幸福感；(3)憂鬱對個體的偏差行為有正面影響效果；(4)憂鬱對個體的幸福感有負向影響效果。另外，本研究也針對研究結果和方法論提出建議，以作為未來研究的參考。

關鍵字：親子關係、憂鬱、偏差行為、幸福感

Confirmation of structural equation model of parent-child relationship 、 depression 、 deviant behavior and well-being

Kai-Fang Hsueh

Doctor candidate, Department of Education, National Cheng-Chi University
95152502@nccu.edu.tw

Abstract

A number of studies have imply that concordant parents-child relation influence teenagers external behaviors and wellbeing. To further confirm this suggestion, researcher built up a model of parent-child relationship, depression, deviant behavior and well-being. By using data from Taiwan Educational Panel Survey(TEPS), 5300 students were selected as samples. Results show that the empirical data support the hypothetic model in this study and hypotheses are almost confirmed: (1) Parent-child relationship correlates negatively with depression and deviant behavior. (2) Not only is significant direct effect of Parent-child relationship on the well-being, but also the indirect effect through Depression. (3) Depression correlates positively with deviant behavior. (4) Depression correlates negatively with well-being. Furthermore, this study proposes some useful suggestions to stakeholders and methods for the future study.

Keyword: depression, deviant behavior, parent-child relationship, well-being

壹、研究背景與目的

青少年時期是人生全程發展（life-span development）中介於兒童期（childhood）與成人期（adulthood）之間的一段生命時期（郭靜晃，2006）。學者Erikson（1969）的心理社會發展理論指出，青少年的發展任務是追求對自我角色的認定，如果沒有在這個人生階段如期完成，將造成自我角色混淆的危機，嚴重的甚至會產生偏差、犯罪行為等。因此，這段時期是身心發展急遽的風暴期，青少年們正面臨了從「依賴兒童」到「獨立成人」的轉變。此外，他們也必須面對許多個人的挑戰，包括生理的轉變、學習與家人和同儕相處、對未來的生涯規劃、發展價值觀等（魏弘貞，2006）。

Hurlock（1974）認為：1.個人在一生中生活在家庭裡的時間比在任何其他場所多，而在行為、態度及價值觀上，也深受家庭的影響，2.父母對子女行為的控制，比其他任何人或團體所施加的控制時間都來得長，3.雖然兒童及青少年會與教師或其他同儕團體建立起情緒接觸，但此種情緒接觸並不像與家人所建立起的這麼持久，4.在家庭中，個人能自由的表達其喜悅及悲傷，家人間能彼此分享其歡樂，也能撫慰其創傷，5.家庭是個人的避風港，此種安全感對兒童及青少年來說，別具重要性。由此可知，家庭的支持力與父母的影響力在青少年的生活裡扮演了一個非常重要的角色。

在相關研究部分，許多研究發現，家庭支持有助於抑制青少年偏差、犯罪行為與減少藥物濫用的機會。學者Printz（1991）的研究發現，家庭支持可以減緩生活壓力，其中家庭成員的強烈支持可減少身心疾病症狀；Martinez（1995）針對82位15歲到19歲的青少年研究發現，較多的家庭支持、同儕支持和教師支持可以減少青少年的越軌行為；Sheeber, Hops, Alpert, Divis & Andrews（1997）針對231位青少年女生和189位青少年男生及其母親所做的一項研究發現，感受到較少家庭支持、較多家庭衝突的青少年，在未來一年中，會表現出較多的沮喪情緒，因此，研究最後指出，從家庭互動品質、家庭支持可以了解青少年沮喪症狀的發展；國內研究部分，王者欣（1995）的

研究結果顯示，若家庭對青少年提供支持越多、越廣，則青少年在生活事件中的個人壓力、家庭壓力、同儕壓力、學校壓力、社會壓力感與整體壓力感都有相對減少的趨勢；鄭照順（1996）的研究也得到相同的結果，感受到較高家庭支持的青少年，會表現出較少的沮喪情緒。除了負面行為的減輕外，也有研究發現，家庭支持對青少年生活適應、身心健康及情緒發展有正向的影響，Martinez（1995）的研究指出，較多的家庭支持、同儕支持和教師的支持可以增進青少年的身心健康和幸福感；Kostelecky & Lempers（1998）發現，強烈且正向的家庭支持會讓青少年心理更健康，比較快樂，在生活和未來上有比較多正向的表現，李素菁（2002）也發現，家庭支持的重要程度直接影響青少年之幸福感受。

綜合上述研究發現可知，當父母親對孩子表達出較多的支持行為，如情感慰藉、接納及支持，孩子會表現出較多的正向情緒與幸福感受，表現出較少的沮喪情緒和焦慮，因此，父母與家庭的支持有助於青少年的生活適應，也可以減少壓力危機的發生，更是影響青少年身心健康的主要因素，因此，良好且具支持性的親子關係，不但可能減少青少年的越軌行為，也可增進其身心健康和幸福感。因此，本研究之目的便是瞭解家庭的親子關係，青少年憂鬱情緒、偏差行為及青少年幸福感之間的關係，進而提出模式來驗證現今相關研究所提出之親子關係對於青少年憂鬱、偏差行為及幸福感的影響情形。

一、親子關係與憂鬱、偏差行為之關係及相關研究

所謂親子關係，是指在家庭中父母與子女互動所構成的一種人際關係，此關係包含關愛、情感、溝通三個向度的程度（陳春秀，2001），在互動過程中含有孩子與父母的需求，父母重視子女的態度，彼此親子間的包容力與親密度。

Mitchell（1983）的研究指出，青少年的家庭系統內代罪羔羊傾向愈明顯，其適應問題也就愈多，Stutman（1985）的研究指出，當家庭系統為

失功能、形成三角關係時，對13~18 歲青少年的適應影響愈大，Sabatelli & Anderson（1991）的研究則指出，家庭系統分化、同儕支持與青少年晚期焦慮與沮喪憂鬱有密切的關係。

Goldstein（1990）、Seydlitz與Jenkins（1998）提出家庭因素中有幾項對青少年犯罪行為影響甚鉅，分別是，缺乏父母親的愛、父母親拒絕、父母親管教因素(如苛嚴之管教、虐待、管教鬆散、根本不管)、家庭成員偏差行為或犯罪行為之學習感染、家庭失和與家庭暴力、父母親的關係疏離；國內黃富源（1986）經實證比較研究後發現，親子關係較差之少年，其犯罪傾向較嚴重，並指出由於親子關係不良，影響家庭氣氛，造成家庭成員之凝聚力較差，影響到子女之行為反應和表現，促成了偏差行為之發生，此外，吳齊殷（1996）也指出，家庭因素對青少年之負面影響有內化症狀（即憂鬱症狀）與外化症狀（即行為問題），更嚴重的甚至會有兩者併發症。

綜合上述可知，當父母與青少年的關係越親密，父母對孩子愈接納，使得青少年對自己滿意，那麼家庭對青少年所呈現的是一種隱形的拉力，暗示著青少年所獲得的父母支持應該越高，而這樣的支持將有助於減少孩子產生憂鬱、偏差行為的情況。

二、親子關係與幸福感之關係及相關研究

Van, Bogt and Raai（2002）探討親子關係及其相關因素對子女幸福感的影響，其研究發現，父母參與和子女之幸福感有關，而親子關係亦直接影響子女的幸福感，其提出青少年及前成人期子女幸福感之概念化模式，其以子女的性別、年齡為背景變項，親子關係（parental bond）、友伴關係，以及最要好的朋友為歷程變項，進行幸福感探討之研究發現，親子關係之穩定與否與子女幸福感程度有關；另外，Kostelecky and Lempers（1998）針對133 位高中生所做的一個家庭支持與幸福感的相關研究發現，強烈且正向的家庭支持會減少青少年的苦惱，且讓心理更健康，他們會比較快樂，在生活和未來上有比較多正向的表現；在國內研究部份，郭怡伶（1995）研究發現母子溝

通品質對青少年及母親的心理社會幸福感皆呈顯著正相關，李素菁（2002）研究青少年家庭支持與幸福感之關係發現，家庭支持對於青少年而言，是一項非常重要的支持來源，而其重要程度更直接影響了青少年之幸福感。因此，當父母親與青少年的互動良好、並且讓孩子感受到被接納，那麼將有助於提升孩子的幸福感受。

三、憂鬱與偏差行為之關係及相關研究

心理學與精神病理學研究一致地發現，在整個少年時期與進入成年期的反社會行為與憂鬱症狀間有穩定的同步關聯性（翁國峰，2008），當少年無法避免痛苦或是處於逆境時，可能會產生一系列的負向情緒，包含憂鬱，而遭受憂鬱侵蝕的少年，可能會從事偏差行為以舒緩其苦悶（Beyers and Loeber, 2003; Kluegel and Smith, 1986）。

Angold & Costello（1993）整理少年併發性偏差行為研究，發現有四種不同的觀點假設，分別為：1.穩定（stability）觀點；2.相互影響（mutual influence）觀點；3.表現出來（acting out）觀點，以及4.挫敗（failure）觀點（吳齊殷、李文傑，2003），這四個觀點可說是統合目前探討有關憂鬱-偏差行為實證研究發現最完整的分類形式，且這四種觀點各有其實證上之支持與反對者，例如，近年社會學與犯罪學者運用全國性與大型樣本進行研究，發現犯行與憂鬱間確實有正向關聯，而其研究通常是聚焦在二者間併發的現象（Hagan and Foster, 2003），而本研究則欲以表現觀點來探討憂鬱與偏差行為間的關係，也就是認為問題行為是被表現出來的內化症狀的一部分，少年憂鬱情緒或感覺會以戴面具方式輾轉表達出來，換言之，我們所察知的外化症狀（偏差行為），乃是內化症狀（憂鬱）發展的後端結果（吳齊殷、李文傑，2003），因此，憂鬱會先於外化的偏差行為，並可用來預測偏差行為的發生。

四、偏差行為、憂鬱與幸福感之關係及相關研究

如何定義一個具有「幸福感」的人呢? Schmeidler (1998) 指出, 美國兒童福利機構以經濟安全、身體健康狀況、教育水準、行為表現, 以及社會環境優劣等五個指標, 評定兒童幸福感程度之高低; 而學者Csikszentmihalyi 和 Wong (1991) 則指出, 具幸福感者有以下特徵: 正向情緒、喜歡參與活動、善於社交、人際關係良好、喜愛自己從事的活動、善於控制自己的活動類型, 以及具有避免自己受到不幸外在事件影響之外控能力。由上述學者對於幸福感的認定指標可知, 一個擁有幸福感的人, 是具有正向情緒感受、且能控制自己的行為表現、避免受到不幸事件影響的能力, 而這似乎也暗示著偏差行為、憂鬱與幸福感之間的負向關係。

Katja, Paivi, Maria-Terttu, & Pekka (2002) 研究青少年主觀幸福感、健康行為及學校滿意度發現, 吸煙者比非吸煙者呈現較多生理上的抱怨、不健康 (ill-being) 的感受, 酗酒的行為和不健康也有高度的相關, 然上述研究並未指出偏差行為 (抽煙、喝酒) 與主觀幸福感的明確關係, 因此, 本研究欲進一步探討兩者的關係; 另Day和Maltby (2003) 指出, 具有樂觀的性格、少有憂鬱傾向、自尊心高者, 其幸福感程度愈高, 陳艷玲 (2007) 研究國小學童父母教養方式、情緒能力、幸福感與憂鬱傾向之關係發現, 國小學童幸福感與憂鬱傾向間有顯著負相關, 因此, 根據上述相關理論與研究之發現, 本研究推論偏差行為及憂鬱對幸福感有負向的影響效果。

綜合上述文獻內容, 本研究所提出的假設如下:

- H1: 親子關係對憂鬱情況有負向影響效果。
- H2: 親子關係對偏差行為有負向影響效果。
- H3: 親子關係對幸福感有正向影響效果。
- H4: 憂鬱情況對偏差行為有正向影響效果。
- H5: 偏差行為對幸福感有負向影響效果。
- H6: 憂鬱情況對幸福感有負向影響效果。

貳、研究方法

一、理論架構與研究目的

本研究架構主要以探討親子關係、憂鬱、偏差行為與幸福感的關係模式，結合相關實證研究，建構出以親子關係為前置變項，憂鬱與偏差行為為中介變項，幸福感為結果變項的結構方程模式，以探討青少年的「親子關係」對其行為及心理感受的因果機制。研究主要分析架構如圖1所示。

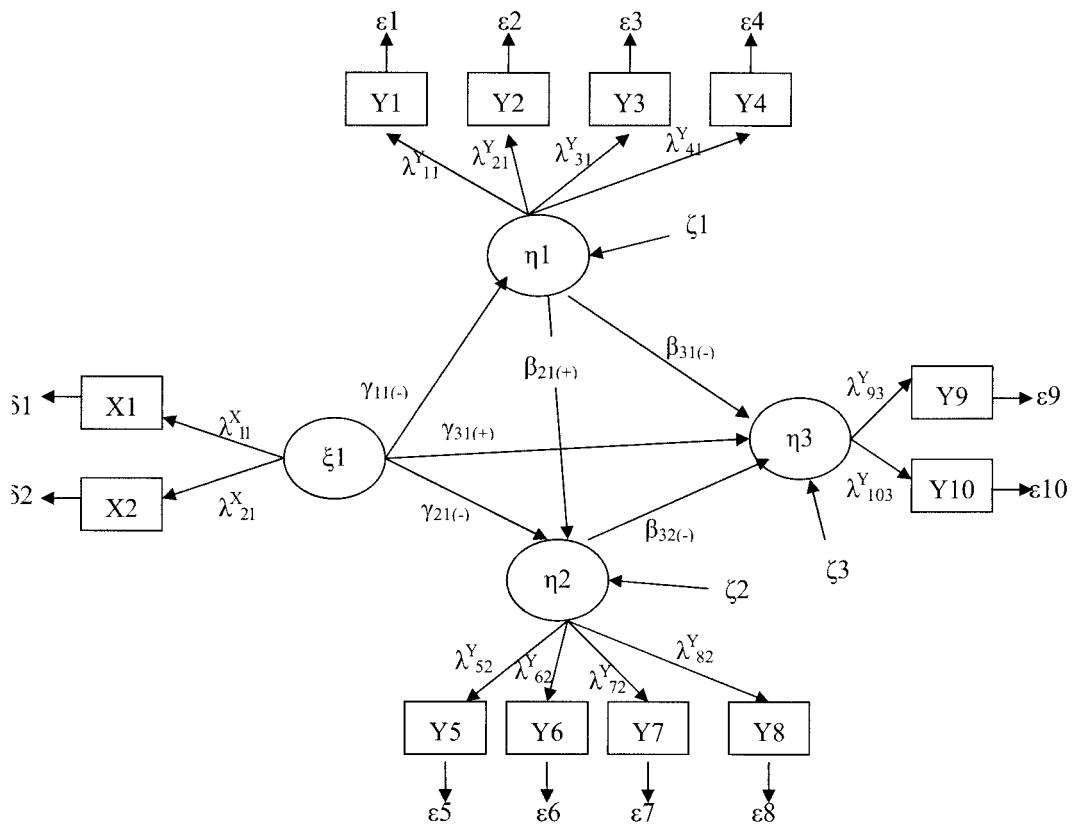
潛在自變項分別為(1)「親子關係 ($\xi 1$)」，主要是以親子互動(X1)、父母接納(X2)等兩個指標變項來表示。潛在依變項分別為(1)「憂鬱 ($\eta 1$)」，主要以Y1至Y4四個指標變項來表示；(2)「偏差行為 ($\eta 2$)」，主要以Y5至Y8四個指標變項來表示；(3)「幸福感 ($\eta 3$)」，主要以正向自我評價(Y9)及快樂感(Y10)等兩個指標變項來表示。

基於所提出的模式架構，本研究的研究目的如下：

- 1.以TEPS所得的實證資料，可以支持本研究提出的理論模式
- 2.在上述所提出的理論模式下，H1~H6的研究假設都能獲得證實。
- 3.在跨樣本的交叉驗證上，本研究所提出的理論模式具有模式穩定性。

二、研究對象

本研究對象取自台灣教育長期追蹤資料庫(TEPS)第一波調查中的國中學生問卷，該研究抽樣係根據公私立及城鄉，採分層隨機抽樣方式，先抽出學校，依學校規模分別抽出4~6班，再從班級中隨機抽出15名學生，在第一波的調查中，國中部分抽出的學校總數為338所，班級數為1,244班，學生為20,004人，因研究者所取得的資料庫為公共使用版，內容為實際受訪學生70%的樣本資料，共13,978人。



ξ_1 ：親子關係

X_1 ：親子互動

X_2 ：父母接納

η_1 ：憂鬱

Y_1 ：問卷題目 W1s517

Y_2 ：問卷題目 W1s518

Y_3 ：問卷題目 W1s522

Y_4 ：問卷題目 W1s523

η_2 ：偏差行為

Y_5 ：問卷題目 W1s509

Y_6 ：問卷題目 W1s512

Y_7 ：問卷題目 W1s513

Y_8 ：問卷題目 W1s514

η_3 ：幸福感

Y_9 ：正向自我評價

Y_{10} ：快樂感

圖一 親子關係、憂鬱、偏差行為和幸福感結構關係模式圖

三、研究變項的測量

(一) 親子關係

本研究的親子關係，包含下列兩個測量指標來源，茲分別說明如下：

1. 親子互動

以學生問卷中的「爸爸會不會和你談升學或就業的事情」、「爸爸會不會聽你講內心的話」、「爸爸會不會看你的作業或考卷、瞭解你的學習情況」、「媽媽會不會和你談升學或就業的事情」、「媽媽會不會聽你講內心的話」、「媽媽會不會看你的作業或考卷、瞭解你的學習情況」等六個題項加以測量，題項計分方式皆為四點量表。本研究以上述六個題項的題平均數作為親子互動測量指標，分數越高代表親子互動越高。

2. 父母接納

以學生問卷中的「爸爸很能接納我」、「媽媽很能接納我」等兩個題項加以測量，題項計分方式皆為四點量表，本研究以兩題項的題平均數作為父母接納的測量指標，分數越高代表父母接納越高。

(二) 憂鬱

以學生問卷中「這學期曾鬱卒」、「這學期曾焦慮擔心」、「這學期曾感覺到孤單」、「這學期曾感覺到無助，沒有人可以依靠」等四個題項加以測量，題項計分方式皆為四點量表。本研究以此四個題項的題平均數作為「憂鬱」的測量指標，分數越高代表憂鬱程度越高。

(三) 偏差行為

以學生問卷中「這學期有逃學或翹課」、「這學期有抽菸、喝酒、嚼檳榔」、「這學期有逃家」、「這學期有偷竊或破壞他人物品」等四個題項加以測量，題項計分方式皆為四點量表。本研究以此四個題項的題平均數作為「偏差行為」的測量指標，分數越高代表偏差行為的程度越高。

(四) 幸福感

本研究的幸福感，包含兩個測量指標來源：

1.正向自我評價

以學生問卷中的「要發表意見、報告或說明時，我都表現的很好」、「對在活動中如何分工，我頗有概念」、「不論任何事，我都能有條不紊的規劃清楚」、「和任何人都能合作很好」、「生活中沒有問題可難倒我」等五個題項加以測量，題項計分方式皆為四點量表。本研究以上述五個題項的題平均數作為正向自我評價的測量指標，分數越高代表正向自我評價越高。

2.快樂感

以學生問卷中的「整體來說，你現在生活快樂嗎」作為快樂感的測量指標，分數越高，代表快樂的程度越高。

四、資料處理

本研究以SPSS15.0版求取變項間之相關矩陣及平均數和標準差，接著再以LISREL8.80版電腦統計套裝軟體進行線性結構方程模式的參數估計、適配度考驗與模式確認，並以.05作為統計顯著水準。

有鑒於樣本人數高達13,978人，可算是大樣本資料，為了評估假設模型的穩定性，研究者特別將樣本切割為兩個次級樣本，隨機選取50%的測定樣本，作為模式建立之用，剩餘50%為效度樣本，作為模式交叉驗證之用。至於對於樣本資料缺失值的處理，研究者採用整列剔除法方式加以處理；因此，LISREL分析的實際測定樣本數為5,300人，效度樣本數為5,774人，合計11,074人。

參、研究結果

首先以測定樣本，進行理論模式與觀察資料的適配度考驗，並以最大似估計法（Maximum Likelihood, ML）進行參數估計。表1為各觀察變項的平均數、標準差和相關係數一覽表，對角線以下之三角形矩陣為測定樣本相關係數矩陣、平均數及標準差；對角線以上之上三角形矩陣為效度樣本相關係數矩陣、平均數及標準差。由4個潛在變項、12個測量指標的相關係數矩陣得知，除了9個係數外，其餘相關係數皆達.01的顯著性水準，顯示12個觀察變項之間有著密切的關係存在。

表1 各觀察變項的平均數、標準差及相關係數

變項	X1	X2	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	M	SD
親子關係 ($\xi 1$)														
親子互動(X1)	—	.406*	-.028*	.002	-.077*	-.085*	-.076*	-.070*	-.046*	-.054*	.210*	.145*	3.61	.72
親子接納(X2)	.343*	—	-.103*	-.074*	-.122*	-.138*	-.093*	-.058*	-.077*	-.064*	.175*	.208*	4.19	.74
憂鬱 ($\eta 1$)														
Y1	-.027	-.102*	—	.697*	.495*	.477*	.061*	.057*	.067*	.065*	-.161*	-.263*	1.70	.84
Y2	.019	-.081*	.694*	—	.467*	.451*	.022	.048*	.053*	.048*	-.149*	-.244*	1.74	.83
Y3	-.058*	-.098*	.494*	.460*	—	.735*	.113*	.097*	.118*	.121*	-.152*	-.298*	1.47	.78
Y4	-.052*	-.107*	.479*	.434*	.737*	—	.083*	.081*	.099*	.107*	-.132*	-.289*	1.40	.74
偏差行為 ($\eta 2$)														
Y5	-.054*	-.065*	.062*	.003	.067*	.091*	—	.524*	.625*	.458*	-.062*	-.053*	1.04	.27
Y6	-.049*	-.065*	.071*	.026	.077*	.074*	.464*	—	.558*	.441*	-.042*	-.048*	1.05	.29
Y7	-.010	-.058*	.087*	.016	.074*	.086*	.551*	.545*	—	.570*	-.046*	-.053*	1.03	.23
Y8	-.034*	-.042*	.095*	.052*	.106*	.103*	.450*	.457*	.543*	—	-.062*	-.050*	1.04	.25
幸福感 ($\eta 3$)														
自我評價(Y9)	.228*	.191*	-.182*	-.163*	-.154*	-.140*	-.048*	-.048*	-.051*	-.062*	—	.230*	2.75	.59
快樂感(Y10)	.167*	.235*	-.279*	-.263*	-.292*	-.287*	-.050*	-.040*	-.048*	-.055*	.230*	—	3.18	.69
平均數(M)	2.72	3.31	1.68	1.74	1.44	1.39	1.04	1.04	1.03	1.04	2.76	3.19		
標準差(SD)	.67	.60	.82	.82	.75	.73	.26	.27	.23	.25	.58	.68		

註1:對角線以下之三角形為測定樣本相關係數矩陣 (N=5300)，對角線以上之三角形為效度樣本相關係數矩陣 (N=5774)

註2：* p < .01

一、理論模式之適配度考驗

依Bagozzi與Yi (1988) 建議，本研究從基本適配度、整體適配度與內在結構適配度三方面進行結構方程模式之評鑑，說明如下：

(一) 模式的基本適配度考驗

模式評鑑前，須確立所估計的參數並未違反統計所能接受的範圍。若發生不適當的解就是一種違犯估計(offending estimate)，表示模式有問題存在，必須先行處理。一般常發生的違犯估計有以下兩項條件：1.有負的誤差變異數存在，或是在任何建構中存在著無意義的變異誤，2.有太大的標準誤。而本研究假設模式的參數估計並沒有負的誤差變異，也沒有過大的標準誤，符合上述條件，表示所獲得的模式在理論上是合適的，可進一步看整體模式適配度及模式內在結構適配度。

(二) 模式的整體適配度考驗

表2為本研究模式的整體適配度考驗結果，由表2得知，理論模式與觀察資料整體適配度卡方考驗 $\chi^2_{(48)}=2156.21$ ，N=5300，p<.001，達顯著水準，表示觀察所得之共變數矩陣與理論上的共變數矩陣均等的假設必須予以拒絕，亦即理論模式與觀察資料並不適配，顯示本模式需進行模式修正。本研究主要以卡方差異檢定和Lagrange多重檢定，作為模式修正方法，並依據LISREL所提供的修正指標進行修正，每次修正只限於一個修正指標，以避免「為改善適配度而修正」，或修改後的參數理論無法解釋。依據理論模式分析所提供的修正指標訊息，研究者把相關測量指標的測量誤差設相關，進行模式修

正，修正後卡方考驗結果得 $\chi^2_{(46)} = 255.71$ ， $N=5300$ ， $p < .001$ ，仍達顯著水準，顯示修正後的理論模式與觀察資料仍然無法適配。然而，卡方檢定對樣本數相當敏感，並且資料是否呈現多變量常態分配有較嚴格的要求，一旦樣本過大或資料偏離多變量常態分配會造成卡方統計量急遽上升而導致拒絕虛無假設（李敦仁，余民寧，2005），也因為卡方值容易受樣本大小所影響，因此學者認為可以不必太在乎這個指標（林志哲，2007），尚須參考其他重要適配指標作為評鑑之依據。

在其他重要的適配度指標或替代性指標方面，GFI、AGFI、PGFI、IFI、NFI、NNFI、CFI及RFI皆大於.90以上之要求，而RMSEA則小於.05以下之標準；殘差分析指標，如SRMR也小於.05以下之標準，這些結果都顯示修正後的模式與觀察資料的整體適配度相當理想。

在模式精簡度方面，本模式之AIC指數為319.71，雖然小於獨立模式之指數23108.16，但大於飽和模式之指數156.00，未能完全符合「理論模式的AIC必須小於飽和模式與獨立模式的AIC」標準。其次， χ^2 比率（ χ^2/df ）為5.56，也不符合「 χ^2 比率應小於1~3之間」的標準，這些結果顯示修正後的理論模式並非是一個精簡的模式。

整體而言，本研究所建構的親子關係、憂鬱、偏差行為與幸福感的關係模式，在整體適配度的考驗上，各項適配指標都顯示修正後的理論模式與觀察資料有相當的適配度，即修正後的理論模式應可以用來解釋實際的觀察資料。

表2 初始理論模式與修正後理論模式之整體適配度考驗指數比較
(N=5300)

指標		初始理論模式	修正後理論模式
df		48	46
卡方考驗	χ^2	2156.21	255.71
	χ^2/df	44.9210	5.5589

適配度指標	GFI	.94	.99
	AGFI	.90	.99
	PGFI	.58	.59
	IFI	.91	.99
	NFI	.91	.99
	NNFI	.88	.99
	替代性指標	Independence AIC	23108.16
Model AIC		2216.21	319.71
Saturated AIC		156.00	156.00
PNFI		.66	.69
CFI		.91	.99
RFI		.87	.98
RMSEA		.091	.029
殘差分析	RMR	.043	.023
	SRMR	.043	.023

(三) 模式之內在結構適配度考驗

依Hair、Anderson、Tatham和Black等人（1998）主張應包括測量模式及結構模式適配度兩方面進行模式內在結構適配度評鑑。就測量模式的適配度而言，表3顯示所有估計的因素負荷量皆達顯著水準， $t=-19.80\sim 48.08$ ， $p < .05$ ，大致符合「因素負荷量應達顯著水準」的評鑑標準。

其次，根據表4，本研究假設模式中12個測量指標的個別指標信度，有10個低於.50的標準，其中有5個指標大於.45，應屬可接受範圍，然有4個指標信度低於.40，依序是親子互動、父母接納、正向自我評價及快樂感，此一結果顯示這4個觀察變項的測量誤差不小。再者，表4顯示本研究假設模式中，4個潛在變項親子關係、憂鬱、偏差行為與幸福感的組成信度依序為.515、.777、.803、.384，除親子關係、幸福感的組成信度較低外，全部都達「.60以上」的評鑑標準。最後，在變異抽取量方面，本研究假設模式4個潛在變項的變異抽取量依序是.348、.466、.505、.241，其中，只有偏差行為

的變異抽取量達.05以上之標準，憂鬱則是大於.45，應屬可接受範圍，而親子關係與幸福感的變異抽取量則是偏低。

在結構模式的適配度方面，Hair等人（1998）認為應從結構參數（表3中的 γ 及 β 值）的顯著性考驗、潛在變項的 R^2 值和潛在變項的相關係數等三方面來評量，根據表3，除了 β_{32} 外，所有估計的結構參數都達.05的顯著水準（ $t=-16.64\sim 16.76, p<.05$ ），其次，根據表3殘差變異量（ ξ ）計算所得的潛在依變項 R^2 值，憂鬱、偏差行為及幸福感3個潛在依變項依序為.03、.03、.81，其中，只有幸福感的 R^2 值高於.50，其餘2個潛在依變項明顯偏低。最後，表5顯示本研究假設模式4個潛在變項的交互相關係數介於-.17~.69之間，合乎「潛在變項的相關係數應低於.90」的評鑑標準。

表3 修正理論模式之參數顯著性考驗及完全標準化估計值

參數	標準誤	t值	標準化估計值	參數	標準誤	t值	標準化估計值
λ_{11}^X	.021	26.15*	.55	δ_1	.023	30.80*	.70
λ_{21}^X	.023	27.60*	.63	δ_2	.027	22.75*	.61
λ_{11}^Y	----	----	.70	ε_1	.022	23.37*	.51
λ_{21}^Y	.013	48.08*	.64	ε_2	.021	28.44*	.59
λ_{31}^Y	.029	24.05*	.71	ε_3	.022	23.01*	.50
λ_{41}^Y	.029	23.70*	.68	ε_4	.021	25.27*	.54
λ_{52}^Y	----	----	.68	ε_5	.013	40.73*	.54
λ_{62}^Y	.017	40.74*	.68	ε_6	.013	40.83*	.54
λ_{72}^Y	.018	44.67*	.81	ε_7	.012	28.64*	.35
λ_{82}^Y	.017	40.39*	.67	ε_8	.013	41.29*	.55
λ_{93}^Y	----	----	.41	ε_9	.019	44.12*	.83
λ_{104}^Y	.028	19.80*	.56	ε_{10}	.022	31.00*	.69
γ_{11}	.023	-7.48*	-.17	ξ_1	.051	19.21*	.97
γ_{21}	.022	-4.21*	-.091	ξ_2	.039	25.20*	.97
γ_{31}	.035	16.76*	.59	ξ_3	.056	3.45*	.19
β_{21}	.019	6.71*	.13				
β_{31}	.035	-16.64*	-.58				
β_{32}	.023	.06	.014				

註1：----為限制性條件，因此未進行估計。

註2：* $p < .05$

表4 修正理論模式之個別指標信度及潛在變項的組成信度與平均變異抽取量

變項	個別指標信度	潛在變項的組成信度	平均變異抽取量
親子關係 ($\xi 1$)	----	.515	.348
親子互動 (X1)	.30	----	----
父母接納 (X2)	.39	----	----
憂鬱 ($\eta 1$)	----	.777	.466
Y1	.49	----	----
Y2	.41	----	----
Y3	.50	----	----
Y4	.46	----	----
偏差行為 ($\eta 2$)	----	.803	.505
Y5	.46	----	----
Y6	.46	----	----
Y7	.65	----	----
Y8	.45	----	----
幸福感 ($\eta 3$)	----	.382	.241
正向自我評價 (Y9)	.17	----	----
快樂感 (Y10)	.31	----	----

表5 幸福感影響關係模式四個潛在變項的交互相關係數

潛在變項	1	2	3	4
1.親子關係 ($\xi 1$)	1.00			
2.憂鬱 ($\eta 1$)	-.17	1.00		
3.偏差行為 ($\eta 2$)	-.11	.14	1.00	
4.幸福感 ($\eta 3$)	.69	-.68	-.14	1.00

綜合而言，本研究假設模式的內在品質評鑑方面，有4個測量指標的指標信度略低，親子關係、幸福感的變異抽取量不到.50的標準，而憂鬱及偏差行為的 R^2 值也稍低。其餘的評鑑指標都顯示本研究假設模式的內在品質不錯。

二、模式中潛在變項之間的效果

除了整體模式適配度的考驗與模式內在品質的評鑑之外，尚需進一步比較各潛在變項之間的效果，才能了解變項間的線性結構關係，而潛在變項間的效果包括直接效果、間接效果和全體效果三方面，如圖2與表6所示。

（一）就親子關係效果方面

從圖2與表6的理論模式分析結果發現，親子關係除了對憂鬱（ $\gamma_{11}=-.17$ ）、偏差行為（ $\gamma_{21}=-.09$ ）與幸福感（ $\gamma_{31}=.59$ ）具有直接影響效果，亦會透過憂鬱對幸福感產生間接效果（.098），故親子關係對幸福感的全體效果為.692。研究結果支持本研究所提的研究假設H1、H2、H3，即親子關係愈好，則個體的憂鬱情緒愈少，個體偏差行為愈少，幸福感愈高。

（二）就憂鬱效果方面

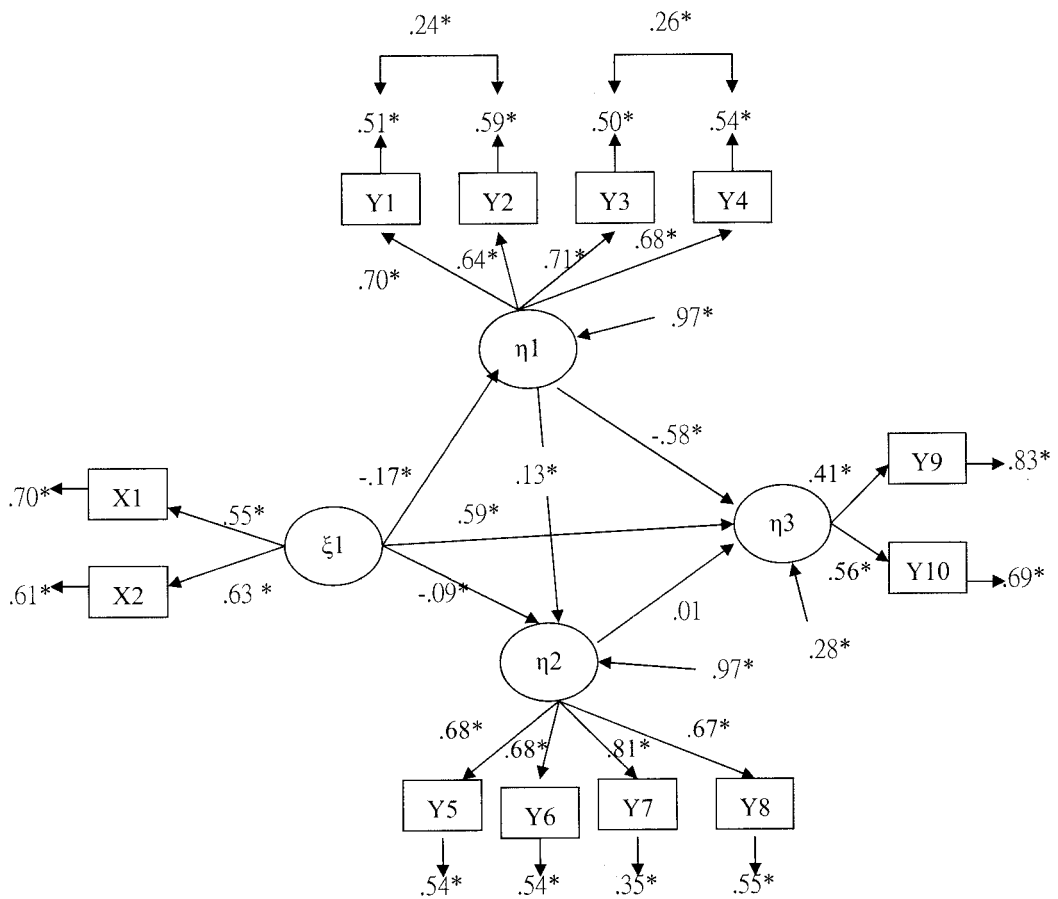
理論模式分析結果也發現，憂鬱對偏差行為（ $\beta_{21}=.13$ ）具有直接正向的影響效果，且對幸福感（ $\beta_{31}=-.58$ ）具有直接負向的影響效果，此研究結果支持研究假設H4、H6，即個體的憂鬱情緒愈高，則偏差行為愈高，幸福感愈低。

（三）就偏差行為

理論模式分析結果發現：偏差行為對幸福感（ $\beta_{32}=.01$ ）並沒有顯著直

接的影響效果，此結果並沒有支持假設H5。

綜合上述，親子關係對個體的憂鬱情緒、偏差行為及幸福感皆有直接的影響力，另外，憂鬱對偏差行為有直接影響力，顯示憂鬱與偏差行為間的表現觀點獲得支持，即憂鬱會先於外化的偏差行為，並可用來預測偏差行為的發生。而就親子關係對幸福感的影響路徑來說，親子關係對幸福感的直接影響力最大（.59），透過憂鬱對幸福感的間接影響力（.099）次之，透過偏差行為對幸福感則沒有顯著的影響力。



ξ1：親子關係

X1：親子互動

X2：父母接納

η1：憂鬱

Y1：問卷題目 W1s517

Y2：問卷題目 W1s518

Y3：問卷題目 W1s522

Y4：問卷題目 W1s523

η2：偏差行為

Y5：問卷題目 W1s509

Y6：問卷題目 W1s512

Y7：問卷題目 W1s513

Y8：問卷題目 W1s514

η3：幸福感

Y9：正向自我評價

Y10：快樂感

圖二 修正後之標準化參數係數圖

表6 四個潛在變項效果值的顯著性考驗及完全標準化效果值

自變項	依變項(內衍潛在變項)			
	η 1憂鬱	η 2偏差行為	η 3幸福感	
ξ 1親子關係				
外衍變項	直接效果	-.172 (.023)	-.091 (.022)	.590 (.056)
	間接效果	-7.480*	-4.210*	3.450*
	全體效果	-.172 (.023)	-.112 (.021)	.692 (.038)
		-7.482*	-5.317*	18.087*
η 1憂鬱				
內衍變項	直接效果		.127 (.019)	-.580 (.035)
	間接效果		6.712*	-16.713*
	全體效果		.127 (.019)	-.580 (.035)
			6.712*	-16.713*
η 2偏差行為				
	直接效果			.014 (.023)
	間接效果			.060
	全體效果			.014 (.023)
				.060

註：表中的第一排為完全標準化效果值，()為標準誤，()下方為t值。

* $p < .05$

三、模式的交叉驗證

考量在研究樣本上雖有良好的適配度，但不代表在其他樣本也具有相當的適配度。因此，模型的交叉驗證即是在檢定一個理想模型在不同樣本上重複出現的程度（Diamantopoulos & Sigauw, 2000）。本研究接著以隨機選取剩餘的5774名學生為效度樣本，進行以相同母群檢驗單一模式的交叉驗證形式，以檢定修正後所建立的模式是否具有模式穩定性。

為了檢定模式的穩定性，本研究採用較嚴謹的交叉驗證模式，指兩個樣本之間具有完全相等的模型設定，同時參數的數值也完全相等。當所有參數均設定為等同的跨樣本結構方程式估計結果顯示（詳見表7），整體模式自由度為121，卡方值為623.50 ($p=.00$)，效度樣本加入後的卡方貢獻統計量為327.734，佔整體模型卡方值的51.70%（卡方貢獻百分比）；相對下，測定樣本模型估計所產生的卡方值為48.30%。

從嚴謹策略發現，效度樣本的所有參數均套用了測定樣本導出的參數值，此時效度樣本的MFF卡方值為327.73，較溫和策略的MFF卡方值(302.63)增加了25.10。但由於嚴謹策略的效度樣本參數估計值，較單一樣本的模式少了14個參數，因此，若將此卡方增量 $\Delta \chi^2$ 進行卡方差異檢定，以14為自由度，顯著水準為.05時，臨界值為23.685。由於卡方差異檢定25.10大於臨界值23.685，表示卡方增量不是一種隨機變化量，但如果以倒置程序來進行複核效化的檢驗，效度樣本的卡方增量則未達.05顯著水準；另外，從ECVI指標來看，越小的ECVI值表示模型效化越高，根據表7，由溫和策略調整到嚴謹策略時，ECVI指標也增加，表示複核效化變得較不理想，但兩個ECVI值落在彼此的95%信賴區間內，表示ECVI值的變化在隨機變化的範圍內，兩種複核效化估計程序的ECVI值不能說有顯著的不同（邱皓政，2003）。這樣的結果顯示，從測定樣本套用到效度樣本的模式穩定度尚可。

表7 交叉驗證策略的適配度評估摘要表

整體模式適配度			效度樣本		
MF χ^2 (df)	WLS χ^2	ECVI		MF χ^2	%
溫和策略			原始	302.63	51.54
587.15(107)	578.22	0.061	倒置	284.52	48.46
嚴謹策略			原始	327.73	51.70
633.94(121)	623.50	0.063	$\Delta \chi^2=25.10^b$, $\Delta df=14$, $p < .05$		
$\Delta \chi^2=45.28^a$, $\Delta df=14$, $p < .01$			倒置	306.21	48.30
			$\Delta \chi^2=21.69^b$, $\Delta df=14$, $p > .05$		

註：a差異值的計算是以嚴謹策略減去寬鬆策略的整體WLS χ^2 。

b差異值的計算是以嚴謹策略減去寬鬆策略的效度樣本的MF χ^2 值。

肆、綜合討論

本研究根據相關理論及實證研究提出假設模式，加以探討親子關係、憂鬱、偏差行為與幸福感的因果模式，並以TEPS國一學生為觀察對象進行模式適配度考驗，考驗結果顯示，修正後的理論模式具有不錯的整體適配度，各項整體適配度指標大致都指出本研究所建構的理論模式可以用來解釋實際的觀察資料。本研究結果支持了親子關係會直接影響個體的憂鬱、偏差行為及幸福感，同時，憂鬱也會直接影響個體的偏差行為及幸福感，此外，親子關係透過憂鬱對個體的幸福感有間接的正向影響力。

就理論模式的內在適配度而言，雖然多數評鑑標準都顯示理論模式具有良好的內在品質，但是仍有一些值得探討之處，例如，部分觀察指標的信度偏低，如親子關係二個觀察指標（親子互動、父母接納）及幸福感的二個觀察指標（正向自我評價、快樂感），這四個觀察指標的個別指標信度略微偏低，而這樣的結果有可能是本研究的所有觀察指標皆是以Likert量尺加以測量，嚴格來講Likert量尺算是一種次序量尺，把次序量尺當成等距量尺處理，因而可能造成測量不夠精確之慮。此外，親子關係及幸福感這兩個潛在變項的組成信度分別為.515、.384，未達「潛在變項的組成信度.60以上」的標準，且變異抽取量只有.348、.241，未達「潛在變項的變異抽取量至少在.50以上」的標準。由於個別指標信度、組成信度、變異抽取量這三項都是判斷觀察指標是否適當的重要依據，顯示本研究取自資料庫來源的研究變項上，存有指標界定不明確、測量不精確等問題，未來仍有很大的改進空間。

就本研究理論模式四個潛在變項間的效果觀之，幾項重要的發現討論如下：

第一，理論模式考驗結果發現，親子關係對於個體憂鬱及偏差行為有負向直接效果，其標準化效果值分別為-.17及-.09，此結果支持親子關係愈好，則個體的憂鬱狀況愈緩和，偏差行為愈少，此結果與相關研究如Martinez（1995）、Printz（1991）、Sabatelli & Anderson（1991）、李娟慧

(1999)、Sheeber等人(1997)、Winde(1992)、鄭照順(1996)等文獻相互呼應。

第二，模式考驗結果顯示，親子關係對於幸福感有強烈的正向直接影響效果，其標準化效果值為.59，此結果支持親子關係愈好，則個體的幸福感愈高，這與許多研究結果呼應，如Van等人(2002)、Kostelecky et.al(1998)、郭怡伶(1995)、李素菁(2002)等；此外，親子關係透過憂鬱對幸福感也有正向間接影響效果，其標準化效果值.099 ($\gamma_{11} * \beta_{31}$)，顯示良好的親子關係透過減緩個體的憂鬱狀況，仍能間接提升個體的幸福感受。

第三，模式考驗結果顯示憂鬱對個體的幸福感有直接負向影響效果，顯示個體的憂鬱狀況愈嚴重，則其幸福感也會較低，此結果與相關研究如Day和Maltby(2003)、陳艷玲(2007)、Aneshensel and Frerichs(1982)、Kessler and McLeod(1985)呼應。

第四，模式考驗結果顯示，憂鬱對個體的偏差行為有正面影響效果，其標準化效果值為.13，此結果顯示個體的憂鬱狀況愈嚴重，則其偏差行為也會較多，顯示遭受憂鬱侵蝕的少年，可能會從事偏差行為以舒緩其苦悶(Beyers and Loeber, 2003; Kluegel and Smith, 1986)，另外，此結果也支持憂鬱與偏差行為間的表現觀點，即外化症狀(偏差行為)乃是內化症狀(憂鬱)發展的後端結果；此外，親子關係透過憂鬱對偏差行為也有負向間接影響效果，其標準化效果值為-.022 ($\gamma_{11} * \beta_{21}$)，顯示良好的親子關係透過減緩個體的憂鬱狀況，仍能間接減少個體的偏差行為。

綜合整個模式的因果路徑來看，親子關係良好不但能減緩個體的憂鬱狀況，減少個體的偏差行為，提升個體的幸福感受；亦可以透過減緩個體的憂鬱狀況而間接減少個體的偏差行為、提升個體的幸福感受；意即，促進親子間良好的互動關係及紓解個體的憂鬱情緒，對於減少個體的偏差行為及提升個體的幸福感受是有效果的。

伍、結論與建議

一、結論

本研究根據相關理論及實證研究，提出一個結構方程模式，以探討親子關係、憂鬱、偏差行為及幸福感間的因果機制，理論模式考驗及相關假設驗證結果發現如下：

(一) 從適配度考驗結果來看，修正後的理論模式可以與實證觀察資料適配

整體適配度上，除卡方考驗外，其他幾個重要的適配度指標都顯示修正後的理論模式可用來解釋觀察資料，另外，模式交叉驗證結果也顯示，修正後之理論模式的穩定性尚可，表示本研究所建構的理論模式可以獲得實徵觀察資料的支持。

(二) 就潛在變項間的效果，五項研究假設獲得支持

就親子關係的效果而言，親子關係會直接影響憂鬱、偏差行為及幸福感，支持研究假設H1、H2、H3；另外，就憂鬱效果而言，憂鬱會直接影響偏差行為及幸福感，與研究假設H4、H6相符，然偏差行為並沒有直接影響幸福感，研究假設H5未獲得支持。

二、建議

最後，本研究針對研究結果與研究方法提出幾點反思及建議：

(一) 就研究結果方面

1. 增進家長與青少年的良性互動關係

研究結果顯示親子關係對於憂鬱、偏差行為有負向關係，而與幸福感具有正向影響，因此，在學校方面可以辦理親職教育講座，並透過班親會等親師交流來提供家長如何增進親子關係的策略，藉以協助家長增進與

青少年的親子關係，另外，也可藉由課程，教導青少年如何與父母有效溝通、表達心中的想法，減低青少年與父母的衝突機會。

2. 納入不同階段的問卷資料

本研究的模式建構與驗證樣本來源為TEPS資料庫第一波調查的國中學生，未來可使用多波的追蹤資料來加以檢驗親子關係、憂鬱、偏差行為和幸福感四個變項間更細部的因果關係。

(二) 就研究方法論方面

1. 以單一理論強化完整性

研究結果大致合乎研究假設，但並未能以單一理論將親子關係、憂鬱、偏差行為和幸福感串連，因此在彼此的交互關係上常以個別的理论敘述與解釋，較缺乏完整性；未來研究者可以尋求或建立更為完整可以涵蓋此四個項目的理論一以貫之。

2. 採取自行命題反應變項內涵

由於透過TEPS資料庫現有資料的緣故，因此在變項的選擇上無法全面反應各個變項（親子關係、憂鬱、偏差行為和幸福感）的完整內涵，未來研究者若要確實整合與提出更具信效度的模式，可以採用自行命題的方法，以涵蓋所有變項應有的面貌。

3. 競爭模型的檢驗比較

四個變項（親子關係、憂鬱、偏差行為和幸福感）之間的方向可以多加驗證，尋找是否還有較佳的組合可能，另外可以尋找四個變項間是否存在其他的中介或調節變項（如青少年的壓力、青少年社會支持系統、家庭溝通等）。

參考文獻

- 王者欣(1995)。家庭支持與國中生主觀壓力的相關性研究-以高雄市為例。國立中山大學中山學術研究所碩士論文，未出版，高雄市。
- 李娟慧(1999)。青少年自尊、社會支持與社會適應之相關研究。國立彰化師範大學輔導研究所碩士論文，未出版，彰化縣。
- 李敦仁、余民寧(2005)。社經地位、手足數目、家庭教育資源與教育成就結構關係模式之驗證。臺灣教育社會學研究，5(2)，1-48。
- 林志哲(2007)。以結構方成模式驗證期望、價值與數學成就的關係。教育學刊，29，103-127。
- 李素菁(2002)。青少年家庭支持與幸福感之相關研究--以台中市立國中生為例。私立靜宜大學青少年兒童福利學系未出版碩士論文，未出版，台中縣。
- 吳齊殷、李文傑(2003)。青少年憂鬱症狀與偏差行為併發之關係機制。台灣社會學，6，119-175。
- 吳齊殷(1996)。教養實施與青少年併發性偏差行為。解構青春少年時。一九九六年台灣青少年犯罪與矯治問題研討會，台北市立師範學院。
- 邱皓政(2003)。結構方程模式-LISREL的理論、技術與應用。台北：雙葉書廊。
- 翁國峰(2008)。非行同儕、負向情緒與效能影響少年偏差行為發展之研究。國立中正大學犯罪防治研究所博士論文，未出版，嘉義縣。
- 郭怡伶(1995)。青少年母子互動與婦女及其子女心理社會幸福感比較研究。私立文化大學兒童福利學系未出版碩士論文，未出版，台北市。
- 郭靜晃(2006)。青少年心理學。台北：洪葉。
- 陳春秀(2001)。國小高年級學童親子溝通、家庭氣氛與親子關係之研究。嘉

- 義大學教育研究所碩士論文，未出版，嘉義縣。
- 陳艷玲(2007)。國小學童父母教養方式、情緒能力、幸福感與憂鬱傾向之相關研究。國立臺南大學教育學系輔導教學碩士班碩士論文，未出版，台南市。
- 黃富源(1986)。台灣北部地區犯罪少年與一般少年家庭背景之比較研究。國立中央警官大學碩士論文，未出版，桃園縣。
- 鄭照順(1996)。高壓力青少年所知覺的家庭、社會支持及其因應效能之研究。國立政治大學教育研究所博士論文，未出版，臺北市。
- 魏弘貞(2006)。影響癲癇青少年生活壓力因應策略相關因素之研究。社區發展季刊，115，363-379。
- Aneshensel, C. S., & Frerichs, R. R. (1982). Stress, support and depression: A longitudinal causal model. *Journal of Community Psychology*, 10, 363-376.
- Angold, A., & Costello, E. J. (1993). Depressive co morbidity in children and adolescents: empirical, theoretical, and methodological issues. *American Journal of Psychiatry*, 150, 1779-1791.
- Bagozzi, R. P. & Yi, Y.(1988). "On the Evaluation of Structural Equation Models." *Academic of Marketing Science*, 16, 76-94.
- Beyers, Jennifer M., and Rolf Loeber. (2003). "Untangling Development Relations Between Depressed mood and Delinquency in Male Adolescents." *Journal of Abnormal Child Psychology*, 31, 247-266.
- Csikszentmihalyi, M., & Wong, M. M. (1991). The situational and personal correlates of happiness: a cross-national comparison. In Fritz Strack, Michael Argyle Norbert Schwarz (Eds.), *Subjective Well-Being* (pp.193-212). England: Regimen Press Plc.
- Day, L., & Maltby, J. (2003). Belief in good luck and psychological well-being: The mediating role of optimism and irrational beliefs. *The Journal of Psychology*,

137, 99-110.

- Diamantopoulos, A. & Siguaw, J. A. (2000). *Introducing LISREL: A Guide for the Uninitiated*. Thousand Oaks, CA: Sage.
- Erikson, E. (1969). *Psycho-social theory of human development*. New York : Norton.
- Goldstein, A. P. (1990). *Delinquents on delinquency*. Champaign, IL : Research .
- Hagan, John, and Holly Foster. (2003). "S/He's a Rebel: Toward a Sequential Stress Theory of Delinquency and Gendered Pathways to Disadvantage in Emerging Adulthood." *Social Forces*, 82, 53-86.
- Hair, J. F. Jr., Anderson, R. E., Tatham, M. L., & Black, W. C. (1998). *Multivariate Data Analysis*(5th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Hurlock,E.B. (1974). *Adolescent Development*. New York : McGraw- Hill.
- Martinez, L. T. (1995). *The relationship between stress and delinquency among Hispanic youth living in a rural environment*. Unpublished doctoral dissertation. Chicago University. Chicago, IL.
- Katja, R., Paivi, A .K., Maria-Terttu, T., & Pekka, L. (2002). Relationships among adolescent subjective well-being,Health behavior, and school satisfaction. *The Journal of School Health*, 72, 243-249.
- Kluegel, J. R., & E. R. Smith (1986). Beliefs about Inequality : *America View Of What is and What Ought to Be*. NY : Aldine de gruyter.
- Kostelecky, Kyle L., & Lempers, Jacques D. (1998). Stress, family social support, distress, and well-being in high school seniors. *Family and Consumer Sciences Research Journal*,27(2), 125-145.
- Kessler, R. C., & McLeod, J.(1985). Social Support and Mental Health in Community Samples. In: S. Cohen & S. Syme (Eds). *Social Support and*

Health, 219-240.

- Mitchell,Arnold (1983). *The Nine American Lifestyles :Who We Are and Where We Going*. NewYork: Warner Books.
- Printz, B. L. (1991). *Adolescent stress and coping: A path analytic investigation of the hypothesized stress buffering roles of social support and social problem solving*. Unpublished doctoral dissertation. University of Texas, Austin, TX.
- Sabatelli, R., & Anderson, A.(1991). Family system dynamics, peer relationships, and adolescents" psychological adjustment. *Family Relation*,40, 363-369.
- Schmeidler, E. (1998). Activity limitations, behavior problems, and children' s well-being. *Journal of Visual Impairment & Blindness*, 92, 698-690.
- Seydlitz,R.,and Jenkins,P.(1998).*The influence of families, friends, schools, and community on delinquent behavior*. In T. P. Gullotta, G. R. Adams , R.
- Sheeber, L. Hops, H. Alpert, A. Davis, B., & Andrews, J. (1997). Family support and conflict: Prospective relations to adolescent depression. *Journal of Abnormal Child Psychology*,25(4), 333-344.
- Stutman, S. S. (1985). *Exploration of critical developmental phase in family life*. Paper presented at the Annual Convention of the American Psychological Association. (ERIC Document Reproduction Service, No. ED 262351).
- Van Wel, F., ter Bogt, T., & Raaijmakers, Q. (2002). Changes in the parental bond and the well-being of adolescents and young adults. *Adolescence*, 37, 316-333.
- Winde, M. (1992). A longitudinal study of stress buffering for adolescent problem behaviors. *Developmental Psychology*, 28, 522-530.

國家圖書館出版品預行編目(CIP)資料

大型教育資料庫建置及相關議題研討會論文集 / 廖

革為等著. -- 新北市：國家教育研究院，

2011.03

面；公分

ISBN 978-986-02-7367-0(平裝)

1. 教育 2. 資料庫 3. 學習評量 4. 文集

520.29

100004433

大型教育資料庫建置及 相關議題研討會論文集

發行人：吳清山

主編：陳清溪

執行編輯：曾建銘、周慧玲

著者：廖革為、郭俊賢、曾建銘、郭伯臣、張意宗、
謝進昌、陳柏霖、余民寧、薛凱方

出版單位：國家教育研究院

出版年月：100年3月初版

地址：新北市三峽區三樹路2號

電話：(02) 8671-5157

傳真：(02) 8671-1478

網址：<http://www.naer.edu.tw>

印刷：領航協會 曦望美工設計社

展售：政府出版品展售中心

五南文化廣場：台中市中山6號

電話：04-22260330；傳真：04-22258234

國家書店松江門市：台北市松江路209號1樓

電話：02-25180207；傳真：02-25180778

網址：<http://www.govbooks.com.tw/>

G P N：1010000550

I S B N：978-986-02-7367-0

定價：新台幣 150元

版權所有·翻印必究



定價新台幣 150 元 / 本

ISBN 986027367-7

00150



9 789860 273670

GPN:1010000550