

# 數學學習成就之性別差異研究 ——以九年一貫課程七年級數學綱要為例

高雄師範大學教育系。黃國清。博士生

## 摘要

本研究是根據教育部頒定的2005年九年一貫數學領域課程正式綱要，藉由紙筆測驗探討不同性別學生在數學學習成就上表現的差異。本研究以臺南市所有公私立國中七年級的學生為研究對象，每校取各年級的第二個班級為樣本，有效樣本共有870人，以試題反應理論三參數模式估計每位受試者能力值。研究結果發現：1.女生的整體數學成就表現顯然高於男生。2.女生在「代數」及「數與量」兩種課程內容的成就表現皆顯著高於男生。3.針對DIF現象進一步分析題型及認知層次。最後依據本研究結果，提出相關之建議。

關鍵詞：九年一貫課程、性別差異、試題反應理論、試題區別功能

---

# **A Study of the Gender Difference on the Mathematical Achievement Based on the Seventh Grade Guidelines of the Grade 1-9 Curriculum**

University Manuscript received: May 7, 2008; Modified: July 17, 2008; Accepted: July 17, 2008.

Guo-Chin Huang • Doctoral student, Department of Education, National Kaohsiung Normal University

---

## **Abstract**

The purposes of this study were to explore the gender differences in performance of mathematics achievement based on the National Curriculum 2005 (the Grade 1-9 Curriculum). The population of the study consisted of all the seventh graders in Tainan City on Taiwan. The results of this study were: 1. the performance due to gender were significantly different, and the female exceeded the male. 2. both of the performances in the “algebra” and “number” due to gender were significantly different, and the female exceeded the male. 3. DIF was present between the groups of gender. Finally, according to the results and findings of this study, there are some suggestions for the practical use and future research.

**Keywords:** DIF, Gender difference, the Grade 1-9 Curriculum, IRT

## 壹、緒論

### 一、研究動機

教育部在2000年公佈「國民中小學九年一貫課程」，大規模地進行中小學的教育改革，並於2002年開始實施「九年一貫課程暫行綱要」的課程，其中的數學領域課程由於學習內容大幅淺化、能力指標不明確，引起研究數學教育學者及基層教師的一片譁然（張海潮，2003）。因此，教育部為加深數學學習內容而修訂能力指標，在2003年底公佈「國民中小學九年一貫課程數學領域綱要」並訂2005年開始實施。

數學為科學之母，世界各國無不重視數學教育，尤其臺灣社會因為升學考試的因素，更視之為學校教育最重要的學科之一，然而如何得到良好的學校教育效果呢？「性別」在個體學習變因之非智力影響因素中，可說是一個重要的變項，因此為了落實兩性平等教育、促進教育品質的提昇，並作為數學教育改進的參考，探究數學學習之性別差異及影響原因，實有其必要之處，而在教育之推展上，也是一個值得探討的問題。教育心理學者大多認為了解學生的個別差異進而因材施教，是為首要之舉，而且咸認「性別差異」是影響學校教育效果的重要「非智力因素」（張春興，2002；溫世頌，1997），所以數學學習的「性別差異」議題亦經常被引起討論，因此當臺灣實施新訂的國定課程時，若能了解數學學習的「性別差異」現象，對於學校教育的效果、數學課程的修訂都具有重要的參考價值。

世界各國對於數學學習之性別差異的研究為數不少，其中大部分是以「古典測驗」（Classical test theory, CTT）的觀察分數來評斷或考驗性別的變項是否有差異，然而這樣的研究結果可能會因受試者的抽樣因素（受試樣本不同）而有不同的結果（觀察分數）。再者，男女受試者在整體測驗表現（總分）上雖無差異，但是有可能在測驗裡的某單元（或某些內容），例如數學課程中的「代數」、「幾何」……等不同單元中會有差異的表現，甚至於不同測驗的題型（符號運算、應用題……等），對不同性別學習者來說，都可能會有差異的表現。

此外，測驗的試題對於某一性別而言，有可能是特別有利或不利的，亦即這些題目對於不同性別的受試者存在著「試題偏失」（item bias）的情形，因此，在研究性別學習成就差異及發展測驗時，除了控制受試者的取樣因素影響外，如何可以偵測和控制受試者在題目的作答，不致於出現性別偏失情形，是本研究的重要課題。

經由上述，在研究過程中，為排除受樣本取樣所造成的誤差因素，同時又可偵測試題是否有偏失的情形，應可運用「現代測驗理論」（Item Response Theory,

IRT) 補強古典測驗之不足，另外經由現代測驗理論所發展出的DIF (Differential Item Function) 或譯為「試題區別功能」，亦可偵測出「試題偏失」的現象。

## 二、文獻探討

### (一) 現代測驗理論 (IRT) 與古典測驗理論 (CTT) 之差異

在目前的測驗理論裡可大約分為古典測驗理論 (Classical Test Theory, CTT) 及試題反應理論 (Item Response Theory, IRT)，Hambleton & Cook (1977) 認為兩者主要之差別在於：

#### 1. 樣本的影響

古典測驗理論以受試者答對之百分比作為難度指標，以試題分數與測驗總分的相關係數作為鑑別度指標，這些統計量數隨受試者能力水準而變，亦即團體相依 (group-dependent)。而題目反應理論的試題參數，如難度參數 $b$ 及鑑別度參數 $a$ ，雖經過不同群體受試者的施測，試題參數不受樣本影響 (sample-free)。

#### 2. 題目難度的影響

古典測驗理論的觀察分數係依測驗的難度而定 (test-dependent)，此現象造成不同受試者接受不同題目的結果難以直接比較。而在題目反應理論中，受試者能力估計不受題目的影響 (item-free)，因此不同受試者的能力值 (ability) 可以直接比較。另外在IRT的三參數 (難度 $b$ 、鑑別度 $a$ 、猜測度 $c$ ，3PL) 模式下估計受試者之能力值，會考慮受試者答題時的猜測情形，加上每一題的難度不同，因此受試者雖然答對的題數相同，但是因為答題型式 (pattern) 不同而有不同之能力值，因而此種「能力值」也就比「答對題數」更接近受試者的真實分數。

### (二) 現代測驗理論之試題區別功能 (DIF)

DIF (Differential Item Function) 或譯為「試題區別功能」，心理計量學者所謂的DIF的定義為：「來自不同族群，但能力相同的個人，如果在答對某個試題上的機率有所不同的話，則該試題便顯現出DIF的現象」 (Hambleton, Swaminathan & Rogers, 1991)，有了這項定義，試題反應理論 (IRT) 很自然的提供一個研究DIF的架構，因為試題特徵函數正可以說明答對某個試題的機率，是與受試者的潛在能力和試題的潛在特徵有某種關聯存在。因此，根據試題反應理論，Dorans及Holland (1993) 認為DIF應是指在不同群體中，試題答對或答錯的條件機率不相同；亦即來自不同次群體、具有相同能力的受試者對於某一或某些特定題目的答對機率不相等；所以利用試題的DIF可以偵測試題偏失與否，提供性別差異性的統計方法。由上述可知試題反應理論對於受試者的觀察分數較具客觀性，不會受樣本因素影響，且不同群體受的能力值亦可分別計算出，故可作為DIF偵測

功能的比較點。

DIF的定義也可以被寫成下列的操作型定義：「某個試題特徵函數如果對不同的族群而言都不相同的話，則該試題便顯現出DIF；反之，如果跨越不同族群的試題特徵函數都相同的話，則該試題便不具有DIF」（余民寧，2004），根據上述的定義，我們只要比較兩個或多個族群在某個試題特徵函數上的差異，就可以判別該試題是否具有DIF存在。試題反應理論常用來診斷試題偏差的方法有兩種：一為比較試題特徵曲線的參數，另一為比較介於試題特徵曲線間的面積。

### 1. 比較試題特徵曲線的參數

如果兩個試題特徵函數的參數值相同的話，則該試題特徵曲線在線上所有點的功能會相同，答對該試題的正確機率值也會一樣。因此，試題特徵函數的參數均相等的虛無假設。如果我們能夠拒絕某個試題的虛無假設，則顯示該試題具有DIF現象。

### 2. 比較介於試題特徵曲線間的面積

試題參數不受考生能力分佈的影響（亦即具有樣本獨立的估計特性），因此，根據不同族群考生所估計出來的同一個試題參數或試題特徵曲線，在經過銜接或等化之後，這些試題參數應該都已建立在共同的量尺上，其試題特徵曲線（ICC）應該會相同，此時，試題特徵曲線間的面積應該等於零（Rudner, Getson & Knight, 1980；余民寧，2004）；若其面積不是為零的話，則顯示該試題對不同族群考生而言，具有DIF現象。

不同性別受試者在測驗裡的某一個或某些題目可能對不同性別受試者來說是有差異的，也就是說對於某一性別而言是特別有利或不利的，因此除了了解整體的成就表現外，進一步來探討試題的公平性，也是在數學學習上性別差異的重要議題。

## （三）七年級數學課程綱要內容與能力指標

「國民中小學九年一貫課程數學領域綱要」的七年級正式課程綱要內容可分為「數與量」及「代數」兩部分，其中「數與量」是由19個能力指標建構而成，指標流水號由7-n-01到7-n-19，另外一部份為「代數」，分由18個能力指標建構而成，指標流水號由7-a-01到7-a-18（7代表7年級、n為Number之縮寫、a為Algebra之縮寫、01到19為編號）。

本研究將依據上述之「能力指標」與數學領域的國民中學「階段目標」（能理解座標的表示，並熟練代數的運算及數的四則運算）作為雙向細目表，略如表一所示，加以編製一份七年級數學領域成就測驗，以作為本研究之研究工具。

表一 七年級數學領域成就測驗雙向細目表

流水號	能力指標	階段目標		
		座標表示	代數運算	數的四則運算
7-n-01	能以「正、負」表徵生活中相對的量，並認識負數是性質（方向、盈虧）的相反。			
7-n-02	能認識如5及-5在數線上的相對位置。			
7-n-03	能在數線上判別整數的大小。			
.....	.....			
7-n-18	能理解連比和連比例的意義。			
7-n-19	能熟練連比例式的應用，如單位換算、三角形面積與邊長或圓面積與半徑間的變化關係。			
7-a-01	能由命題用、等符號列出生活中的變量，並列成算式。			
7-a-02	能嘗試以代入法或枚舉法求解，並檢驗解的合理性。			
7-a-03	能熟練符號的代數操作。			
7-a-04	能由具體情境中列出一元一次方程式，並理解其解意義。			
.....	.....			
7-a-17	能在直角座標平面上認識二元一次聯立方程式的解。			
7-a-18	能熟練使用消去法解二元一次聯立方程式。			
註：細格中「.....」代表省略7-n-04到7-n-17及7-a-05到7-a-16的能力指標部分。				

**(四) 性別在數學成就表現差異之相關研究**

根據教育心理學家的大規模研究結果發現：不同性別的數學成就表現，在不同年齡階段會有不同的差異情形（張春興、陳李綢，1977；Maccoby & Jacklin, 1974；Linn & Hyde, 1989），故研究對象聚焦在某一學習階段或年齡層，應該比較具有微觀的意義；有關數學學習成之研究論文為數不少，本研究僅將研究對象的焦點放在七年級（約13歲）的學生，其數學學習成就之相關研究整理如表二所示。

由表二的相關研究發現：國中階段或單就七年級的學生而言，在數學學習成就的整體表現方面，有顯示女生優於男生者（Forst, Hyde & Fennema, 1994；Linn & Hyde, 1989），亦有無顯著差異者（張靖宜，2005）。而在學習內容方面，有關代數的部分，男生在優於女生（Fennema et al, 1978；Becker, 1990；陳慧珍，2000），亦有女生高於男生（曾建銘，2006）或者無顯著差異（黃有義，2004），並無一致之結論；而在數量（含分數、比例與連結）的方面，女生優於男生（Fennema & Sherman, 1977；黃財尉、李信宏，1999）或者無顯著（林詠娟，2003；陳亮君，2005；蔡逸勝，2006），但也有研究顯示男生的表現優於女生（Lane, Wang & Magone, 1977）。

表二 國中學生之「代數」及「數與量」學習成就測驗之相關研究

研究者	年份	研究對象	研究結果
Lane, Wang & Magone	1977	七年級	比率題對男生較有利。
Fennema & Sherman	1978	中學時期	1. 女生在計算方面比男生有較高之成功率。 2. 男生在解決情境敘述之文字題的成功率則高於女生。
Linn & Hyde	1989	七年級	女生在數學成就上優於男生
Becker	1990	七到九年級	男生在代數題上表現較佳，算術題的表現並無差異。
Forst, Hyde & Fennema	1994	中學時期	1. 數學整體表現及在計算能力上女生稍優於男生。 2. 在問題解決上女生稍佔優勢。
黃財尉、李信宏	1999	七年級	1. 分數運算對女生較有利，整數加減對男生較有利。 2. 計算題及應用題並未產生顯著的性別DIF結果。
陳慧珍	2000	七年級	1. 在傳統式代數文字題的解題能力沒有顯著差異。 2. 在引導式代數文字題的解題能力，男優於女。
林詠娟	2003	七年級	在分數診斷測驗的表現上沒有顯著差異。
黃有義	2004	八年級	不同性別的學生在代數越級學習成就上並無顯著差異。
張靖宜	2005	八年級	在以代數、比例與計算為內容之測驗，性別無顯著差異。
陳亮君	2005	七年級	在以pattern（數與量之連結）為內容之測驗，測驗結果性別無顯著差異。
蔡逸勝	2006	七年級	在以數型與規律（數與量之連結）為內容之測驗，實驗前之測驗結果，性別無顯著差異。
曾建銘	2006	七年級	男生與女生在成就測驗上的表現有顯著差異，達.01顯著水準，研究顯示女生之表現優於男生。

### 三、研究目的

綜上所述，本研究將以國中七年級（約13歲）為研究對象，並以性別及「九年一貫課程數學領域綱要」的七年級能力指標（「代數」及「數與量」）作為探討內容，經由上述，本研究擬達成之目的為：

- (一) 瞭解男、女學生在「九年一貫課程數學領域綱要」七年級能力指標的整體表現是否有差異。
- (二) 瞭解男、女學生對於「九年一貫課程數學領域綱要」七年級的兩個課程內容（代數、數與量）學習上是否有差異。
- (三) 瞭解數學成就測驗之試題對於男、女學生而言是否有偏失情形。

## 貳、研究方法

### 一、研究程序

本研究所採之研究工具「七年級數學成就測驗」，皆為選擇題之題型，為了考慮受試者答題之猜測因素，故採用IRT的三參數模式下所估計的男、女受試者能力值，用以表示學習成就，應更加接近受試者的真實分數。

本研究程序將分為四部分進行：

- (一) 依照全體受試者在「七年級數學成就測驗」上的作答情形，以試題反應理論的三參數(3PL)模式下估計全體受試者的能力值，再將男、女分為兩不同團體，以統計方法考驗男、女受試者的整體學習成就表現之差異。
- (二) 全體受試者在「七年級數學成就測驗」上的「代數」及「數與量」兩部分的答對率，以統計方法考驗全體受試者在「代數」及「數與量」兩部分的學習成就之差異。
- (三) 以統計方法考驗男、女不同性別的受試者分別在不同學習的課程內容(代數、數與量)上的表現差異。
- (四) 以試題區別功能(DIF)來偵測測驗題目是否有「題目偏失」情形。

### 二、研究對象

本研究為配合學校的行政工作以及受試者的完整學習，故施測對象是臺南市所轄公私立國中七年級的學生為母群體，抽取每一所學校七年級的第二個班級的學生為正式施測樣本，有效樣本共870人。

### 三、研究工具

本研究之測驗為研究者自編之「七年級數學成就測驗」(黃國清, 2004)，該測驗為標準化成就測驗，乃依據教育部在2003年底公佈「國民中小學九年一貫課程數學領域綱要」之能力指標所編製而成：

#### (一) 題型及題數：

為符合施測之實務需求(作答時間45分鐘)，僅編製23題選擇題，因為七年級數學領域共有37個能力指標，但有些能力指標是同屬一個概念，只是深廣度不同，例如7-n-02 04、7-n-09 11、7-a-07 09、7-a-11、14、15以及7-a-03、05、06等，因此一道題目有可能會涵蓋2到3個能力指標，所以本研究工具23題，可具七年級數學學習領域內容之代表性。



**(二) 測驗內容：**

由於九年一貫課程數學領域分為「數與量」、「代數」、「幾何」及「統計與機率」，一般國際大型測驗如TISS、NEAP等的數學內容亦是如此區分，而且「代數」及「數與量」兩主題之能力指標亦包含「座標表示」的概念，例如7-n-02~05、7-a-11、14、15、17等，因此本研究之測驗內容僅分為「代數」及「數與量」兩部份，分別各有12題及11題。

**(三) 信度：**

1. 內部一致性 $\alpha$ 係數及庫李20號(K-R20)信度皆為.85，Rulon折半信度為.86。2. 重測信度亦為.86。

**(四) 效度：**

1. 內容效度方面，此份測驗是由雙向細目表及專家學者審訂題目；2. 效標關聯效度方面，依各校之月考成績為效標，並建立起校標關聯效度（與各校月考成績的相關係數，介於0.69到0.88之間，平均相關係數為0.78，均達.01顯著水準。）；3. 另在建構效度方面，採「內部一致分析法」及「外在效標法」驗證測驗之建構效度（余民寧，2003），其中「內部一致分析法」之「試題內一致性」方面，高分組（272人）、低分組（288人）在每道題目答對率的差異t考驗，t值介於8.89~23.76，均達.01顯著水準；「內部一致分析法」之「試題間一致性」方面，全體受試者在每道題目之點二系列相關係數介於.31~.61，均達.01之顯著水準。

此外本測驗之發展並經預試、題目修正、信效度、常模建立等標準化測驗之編製程序完成。

**四、資料分析**

將870份有效樣本的作答資料分別使用BILOG-MG3.0軟體及SPSS11.0版統計軟體，分兩部分來分析：

**(一) 利用BILOG-MG3.0軟體採三參數模式下**

1. 估計全體及男女受試者的能力值。
2. 利用軟體內建之DIF程式，分別估計男、女受試者在每一題之難度係數，藉以將男、女受試者之能力值經調整在共同量尺下。

**(二) 利用SPSS11.0版統計軟體計算**

1. 以t考驗法檢驗男、女受試者之能力值在95%信心水準下是否有差異。

2. 以  $t$  考驗法檢驗不同性別的受試者在課程內容（代數、數與量）的學習上，在95%信心水準下檢驗是否有差異。

## 參、研究結果

### 一、女生在成就測驗整體表現優於男生

分別以男、女受試者的三參數能力值平均數，作為男、女生在數學成就測驗之整體表現，如表三所示，男生平均能力值為-0.091、女生為0.087，男生比女生的能力值差，差距為0.178，首先利用Levene 法檢驗變異數是否同質性之假設，結果F值為0.543，未達顯著水準，顯示男、女兩個不同性別團體的變異數符合同質性之假設性，再以獨立樣本  $t$  考驗的統計方法檢定，計算其  $t$  值為-2.64，已達.01顯著水準，並顯示女生之能力值在95%的信心水準下，高於男生之能力值，亦即女生在數學成就測驗上的整體表現優於男生。

表三 男、女生在數學成就測驗表現之  $t$  考驗分析表

性別	人數	平均數	標準差	平均差異	差異之標準誤	$t$ 值	差異情形
男生	425	-0.091	1.00	-0.178	0.067	-2.64**	女>男
女生	445	0.087	0.99				

\*\* $p < .01$

### 二、全體受試者在「代數」的學習成就表現優於「數與量」

本測驗依據七年級能力指標的課程內容分為「代數」及「數與量」兩部份，所以可將本測驗的整體表現，細分為在「代數方面」及在「數與量」兩部份的表現，由於在預試完畢（有效預試樣本455份），已分別計算出「代數」及「數與量」兩部份的題目平均難度（答對率）分別為.66及.63，可推論「代數」及「數與量」兩部份的題目難度甚為接近。

所以為了解全體受試者在這兩個課程內容之學習上是否有差異？研究者使用答對率高低的方法來檢定，亦即採全體受試者分別在「數與量」及「代數」兩份測驗的答對率，以相依樣本 $t$ 考驗的統計方法檢定，分析表如表四所示；由表四發現全體受試者在數學成就測驗上，其中「代數」部分平均答對率為.5904，「數與量」部份為.5615，「代數」方面的表現，在95%的信心水準下，其兩者之差異已達顯著，顯然優於「數與量」方面的表現。

表四 不同課程內容學習差異之  $t$  考驗分析表

課程內容	人數	平均答對率	標準差	平均差異	差異之標準誤	$t$ 值	差異情形
數與量	870	.5615	.2348	-.02887	.00619	4.664**	代數 > 數與量
代數		.5904	.2531				

\* $p < .05$  \*\* $p < .01$

### 三、性別在不同課程內容之學習成就有差異

如前述，全體受試者在「代數」及在「數與量」兩部份的表現有所差異，研究者為進一步深入剖析男女受試者分別在「代數」、「數與量」兩部份的表現是否有差異？故將此兩種不同的課程內容，分別與男、女受試者的性別變項作一交叉分析，以獨立樣本 $t$ 考驗分析統計方法檢驗，結果如表五及表六所示：

#### (一) 女生在「數與量」的成就表現優於男生

由表五的 $t$ 考驗分析表中，可以觀察出不同性別在「數與量」部份的答對率，在表五顯示女生在95%信心水準下，其答對率顯著高於男生。

表五 男、女生在「數與量」表現差異之  $t$  考驗分析表

性別	人數	平均答對率	標準差	平均差異	差異之標準誤	$t$ 值	差異情形
男生	425	.5527	.24183	-.03939	.01602	-2.459*	女 > 男
女生	445	.5921	.23074				

\* $p < .05$

#### (二) 女生在「代數」的成就表現優於男生

表六的  $t$  考驗分析表顯示女生在「代數」部份的答對率，在95%信心水準下，是顯著地高於男生的答對率。

表六 男、女生在「代數」表現差異之  $t$  考驗分析表

性別	人數	平均答對率	標準差	平均差異	差異之標準誤	$t$ 值	差異情形
男生	425	.5587	.25770	-0.4292	.01711	-2.508*	女 > 男
女生	445	.6016	.24698				

\* $p < .05$

## 四、性別的試題差別功能（DIF）分析

本研究為了解研究工具（數學成就測驗）的題目是否有「試題偏失」的情形，並探討男女不同性別的兩個團體對於相同一道試題的答對機率，是否會有先天環境或後天條件上的差異，亦即同一道題目對於不同性別之受試者是否有不公平現象，所以本研究採用IRT的「試題區別功能」（DIF）來檢驗每一道試題。

將男生425名、女生455名分為兩個族群，因兩族群都使用同一份試卷施測，且本研究僅在了解兩組群於每道題目能力值（難度 $b$ ）的差異，故採單參數模式，將兩群體的 $a$ 及 $c$ 參數設定相同，分別估計兩群體的難度，其中男生難度為 $b_1$ ，女生為 $b_2$ ，如表八所示，由表七可知男生在IRT試題難度（ $b_1$ ）之平均數為0.259，女生（ $b_2$ ）平均數為0.045，女生試題難度比男生低了0.214。亦即當男生的試題難度參數設為0時，女生的試題難度則為-0.214，故將女生所有題目難度參數（ $b_2$ ）加上0.214，得到已調整之題目難度參數（ $b_3$ ）；進一步算出兩群體所有題目難度參數之差異值（ $D_i$ ）及其標準誤，最後分別考驗男、女生在每一試題之差異是否顯著。

經由  $t$  考驗結果顯示，男、女不同性別的兩個受試群體在第17題的作答情形的差異，達.01的顯著水準，出現DIF現象，在假設男、女生在具有相同能力之下，此題男生答對機率顯然高於女生，亦即女生答對該題的機率小於男生，並非數學學習成就或數學能力低於男生，而是可能其它因素。

從介於試題特徵曲線（ICC）間的面積之觀點來看，依兩群體建立在共同量尺的試題參數，畫出兩群體的試題特徵曲線應該會相等。但將兩組共同量尺的試題參數（ $a$ 、 $b_1$ 、 $c$  及  $a$ 、 $b_3$ 、 $c$ ）畫出兩條特徵曲線，如圖一所示。在能力值-3~3之間，介於兩群體試題特徵曲線的面積並不為零，表示兩群體能力相同但答對機率有差異（Hambleton, Swaminathan & Rogers, 1991）。

另一方面若分別以男、女生兩群體的實際作答資料，來估計第17題之三個參數，如表八，據此畫出兩條由男女兩不同群體，分別所估計出來的第17題的特徵曲線，如圖二所示，分別所估計出的第17題特徵曲線，男生的特徵曲線均在女生特徵曲線之上，呈現「均一變化曲線的DIF」（uniform DIF）（余民寧，2004），由圖可知能力值約在1.91以下的受試者作答第17題時，呈現男生答對機率明顯大於女生的情形。

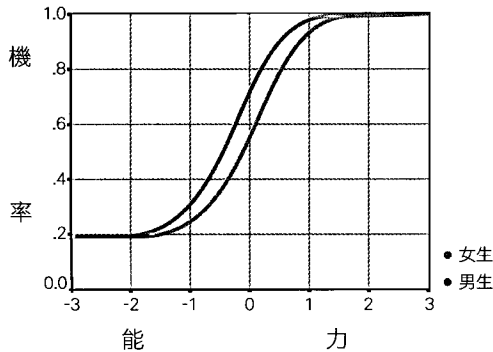
表七 男、女生題目DIF考驗摘要表

題號	鑑別度 (a)	猜測度 (c)	(b1)	(b2)	b1之 標準誤	b3	b3之 標準誤	D1 (b1b3)	D1之 標準誤	D1之 t值
1	2.472	0.364	1.125	1.089	0.089	1.303	0.094	-0.177	0.129	-1.367
2	0.968	0.221	-1.336	-2.057	0.201	-1.843	0.23	0.507	0.305	1.659
3	0.78	0.193	-0.4	-0.314	0.203	-0.1	0.198	-0.299	0.284	-1.054
4	1.774	0.24	-0.143	-0.244	0.097	-0.03	0.098	-0.113	0.138	-0.819
5	1.26	0.15	-0.331	-0.597	0.127	-0.383	0.131	0.053	0.182	0.290
6	0.962	0.233	0.396	0.019	0.151	0.233	0.152	0.163	0.214	0.760
7	1.815	0.307	1.063	1.125	0.096	1.339	0.11	-0.276	0.146	-1.883
8	0.891	0.253	0.57	0.465	0.158	0.679	0.153	-0.109	0.22	-0.495
9	1.553	0.218	0.453	0.193	0.091	0.407	0.09	0.046	0.128	0.359
10	0.985	0.195	0.14	-0.218	0.149	-0.004	0.152	-0.144	0.213	0.676
11	2.2	0.117	1.337	0.983	0.082	1.197	0.066	0.14	0.105	1.330
12	1.714	0.256	-0.306	-0.681	0.122	-0.467	0.139	0.161	0.185	0.870
13	1.284	0.168	0.244	-0.045	0.103	0.169	0.103	0.075	0.146	0.514
14	1.374	0.287	0.465	0.236	0.11	0.45	0.11	-0.016	0.156	0.096
15	1.247	0.18	0.383	0.293	0.103	0.507	0.099	-0.124	0.143	-0.868
16	1.278	0.197	-0.743	-1.054	0.158	-0.84	0.173	0.097	0.235	0.414
17	1.476	0.186	-0.305	-0.111	0.112	0.103	0.102	-0.408	0.151	-2.686**
18	1.011	0.217	-0.903	-1.458	0.191	-1.244	0.21	0.341	0.283	1.187
19	2.312	0.264	1.331	0.954	0.094	1.168	0.075	0.163	0.12	1.363
20	1.694	0.254	1.05	1.017	0.092	1.231	0.094	-0.181	0.132	-1.376
21	1.187	0.224	0.563	0.341	0.113	0.555	0.109	0.008	0.157	0.051
22	0.901	0.216	0.465	0.385	0.151	0.599	0.145	-0.134	0.209	-0.635
23	1.293	0.144	0.839	0.72	0.09	0.934	0.087	-0.095	0.125	-0.758
平均	1.410	0.221	0.259	0.045						

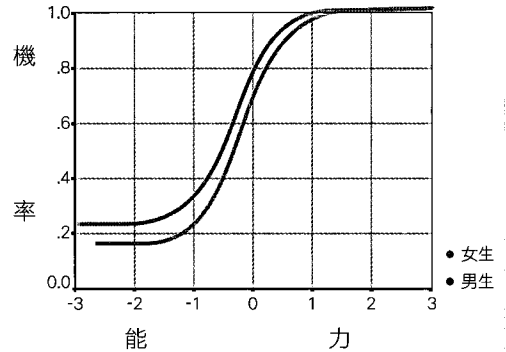
\*\*p<.01

表八 男女兩群體實估之三參數

	鑑別度	難度	猜測度
男生	1.648	-0.269	0.231
女生	1.496	-0.182	0.158



圖一 男女兩群體在共同量尺下第17題之ICC



圖二 男女兩群體實估第17題之ICC

## 肆、研究討論與建議

### 一、在成就測驗整體表現之性別差異

「女生在數學學習成就優於男生」的研究結果，大致符合前述Forst et al. (1994) 及Linn & Hyde (1989) 之文獻結果。造成此一差異之可能因素，其一是傳統文化及教育客觀環境之故(張春興, 2000)，由於臺灣近來的人口結構趨向於「少子化」現象，以及隨著社會的快速變遷，以往的「男尊女卑」、「重男輕女」觀念已不復存在!所以家長對於兒女的教育發展與課業關懷均採取平等的態度；另外亦有研究顯示：在國中前期的女生對於家長所安排的補習、課業輔導的學習態度上較為積極(孫國華, 1993)，所以補習等方式的課外輔導所收到的學習效果，也可能是女生數學學習成就高於男生的因素之一。

### 二、全體受試者在不同課程內容(代數、數與量)之學習差異情形

由表四顯示全體受試者(包括男、女生)在「代數」的答對率，顯然高於在「數與量」的答對率，亦即全體受試者在數學成就測驗上的「代數」方面的表現顯著高於在「數與量」方面的表現。

根據「國際數學與科學成就趨勢調查」(Trends in International Mathematics and Science Study, TIMSS)的結果，臺灣國二學生在2003年在數學科(命題內容大致分為：數與量、代數、幾何、資料呈現與分析等五個領域)的表現，「代數」方面的表現優於「數與量」(題庫發展組, 2007)，因此TIMSS研究的結果亦呼應本研究結果的差異情形。

### 三、不同性別在不同課程內容(代數、數與量)之學習差異情形

本研究的結果，顯示不同性別在「數與量」的學習成就的方面，女生優於男生，此一結果符合 Fennema & Sherman (1977)、黃財尉及李信宏(1999)的論點，另外在「代數」的學習成就部分，與前述之女生高於男生的結論與曾建銘(2006)所作的研究結果一致。

由表五及表六的結果，可以發現不同性別的受試者不管在「代數」或在「數與量」的表現，女生均優於男生，對照表三的研究結果——「女生在數學整體學習成就上的表現優於男生」，可見女生的表現優於男生有其一致性，亦即女生不論在「整體表現」、「數與量」或在「代數」的表現上都比男生來得好。根據研究者在教學現場的觀察，發現女生在學習態度以及參加課後輔導的意願，皆較男生積極，這可能也是上述結果的原因之一。

## 五、試題差別功能 (DIF) 分析

男、女不同性別的兩個受試群體在第17題的作答情形的差異，達.01的顯著水準，出現DIF現象，在假設男、女生在具有相同能力之下，此題男生答對機率顯然高於女生，亦即女生答對該題的機率小於男生，並非數學學習成就或數學能力低於男生，而是可能其它因素，例如對於題目的理解或認知有困難……等。

本研究之研究工具「七年級數學成就測驗」的試題，發現第17題有「DIF」現象的發生，亦即該題被答對的機率有性別上的差異，也就是說該題在題目的內容或是敘述上，對男女受試者而言是有偏誤的。

進一步分析發生「DIF」現象的第17題，該題目為：

已知某星球由星球表面每升高100公尺，氣溫下降0.8度。若星球表面的氣溫為30度，請問下列何者正確？

- (A) 離星球表面上 50 公尺處的氣溫為29.6度
- (B) 離星球表面上100公尺處的氣溫為30.8度
- (C) 離星球表面上150公尺處的氣溫為29.8度
- (D) 離星球表面上200公尺處的氣溫為14度

此題為一元一次方程式的應用題（文字題），題目內容涉及地球科學的相關知識以及戶外活動的情境體驗，就研究者在教學現場及一般生活實務上的經驗，男生應比較常接觸上述之課外知識及情境經驗。

所以造成男生答題機率高於女生的DIF現象，其原因是否可能是男生比較接觸有關地球科學及太空物理等課外知識，故在需應用上述之課外知識及情境體驗的試題上，男生會比女生表現更佳？值得再作探討。另外此題屬「代數」內容之測驗，此一DIF現象亦呼應前點之討論：亦即女生在「代數」之整體學習成就表現優於男生，但在「代數」之細分題型（例如應用題、文字題）上，可能會有相反的結果。

## 六、建議

在認知測驗上，因男女性別而產生的差異表現，一直是教育研究者、心理學家及社會學者關心的重要議題（Feingold,1992），本研究並未進一步探索產生男女生在數學學習上差異現象的主因，另一方面，測驗中有DIF試題也並不代表該試題必須予以剔除，而是可能暗示教學或教材須有所改變（黃財尉、李信宏，1999；Lane et al., 1996；Ryan & Fan, 1996）。

### (一) 就學校教學及教材內容而言

1. 了解「性別差異」亦即「非智力因素的個別差異」，才能因材施教，所以教師應對性別差異有正確認知。
2. 男生在整體數學表現上顯著低於女生，有可能是學習過程的因素，教師應多注意男生在數學教學過程中的學習情況，了解無法專心學習的原因。
3. 教師在講解代數文字題時，對於有關課外知識的數學題目背景，應對女生多做題意的解說；並鼓勵女生多閱讀男生偏愛的課外讀物。
4. 教師在課程內容或者練習題的取材上，應考慮是否兼顧不同性別之認知基礎。

### (二) 就未來研究而言

1. 以本研究結果為基礎，設計「質化與量化並用」的研究方法，進一步探討產生男女生在數學學習上差異現象的主因。
2. 本研究僅針對試題的內容，考驗男女生的試題偏誤（DIF現象）情形，未來研究應可再針對試題的型式、脈絡等變項，進一步探討。
3. 研究發現女生比男生容易產生測驗焦慮（黃美銀，2002；邱俊仁，2003；Hong & Karstenson, 2002），因此本研究產生DIF的題目（第17題）是否因測驗焦慮而影響女生在該題的表現，值得後續研究釐清。

## 參考文獻

- 余民寧（2003）。教育測驗與評量——成就測驗與教學評量。臺北：心理。
- 余民寧（2004）。試題反應理論的介紹（十三）——試題偏差的診斷。2006年4月11日。取自：<http://www.edutest.com.tw/e-irt/index.htm>
- 邱俊仁（2003）。高雄地區國一學生數學焦慮與數學成就之相關研究。國立高雄師範大學數學系碩士論文，未出版。
- 孫國華（1993）。高雄縣市國民中小學學生之課外補習實況與成因之調查研究。教育資料文摘，31（1），135-153。
- 陳亮君（2005）。國中小學生數學胚胎覺察能力發展概況之探討。國立臺南大學測驗統計研究所碩士論文，未出版。
- 陳慧珍（2000）。南投縣國一男女生對文字符號概念與代數文字題之解題研究。國立高雄師範大學數學系碩士論文，未出版。
- 曾建銘（2006）。國中數學題庫之建置——以一元一次方程式單元為例。2007年10月21日，取自<http://www.isst.edu.tw/s44/95/cheng/cheng.htm>
- 黃美銀（2002）。花蓮縣國小六年級學生對班級評量氣氛知覺與學習行為之研究。國立花蓮師範學院國民教育研究所碩士論文，未出版。



- 黃財尉、李信宏（1999）。國中數學成就測驗性別DIF之探討：Poly-SIBTEST的應用與分析。測驗學刊，46（2），45-60。
- 黃有義（2004）。真實情境脈絡的數學教材教法對國二學生代數越級學習影響之研究。明道管理學院教學藝術研究所碩士論文，未出版。
- 黃國清（2004）。南臺南市九年一貫課程七年級數學領域成就測驗之編製與其相關之研究。國立中山大學教育所碩士論文，未出版。
- 張春興（2002）。教育心理學——三化取向的理論與實踐。臺北：東華。
- 張春興、陳李綱（1977）。國小男女生學業成績的性別差異與其教師性別差異的關係。臺灣師範大學教育心理學報，10，21-34。
- 張海潮（2003）。九年一貫數學綱要必須重整。2007年1月1日，取自[http://www.math.ntu.tw/library/math\\_general/article\\_03\\_01\\_16b.htm](http://www.math.ntu.tw/library/math_general/article_03_01_16b.htm)
- 張靖宜（2005）。九年一貫課程數學命題分析研究——「國二段考試題」為例。國立高雄師範大學數學系碩士論文，未出版。
- 張逸勝（2006）。離散數學納入國中小數學課程之研究。國立臺中教育大學數學系碩士論文，未出版。
- 溫世頌（1997）。教育心理學。臺北：三民。
- 題庫發展組（2007）。淺談「國際數學與科學成就趨勢調查」。飛揚，45，15-16。臺北：國民中學學生基本學力測驗推動工作委員會。
- Becker, B. J. (1990). Item characteristics and gender difference on the SAT-M for mathematically able youths, *American Educational Research Journal*, 27(1), 65-87.
- Dorans, N. J., & Holland, P. W. (1993). *DIF detection and description: Mantel-Haenszel and Standardization*. In Holland, P. W., & Wainer, H. (Eds), *Differential Item Functioning*. NJ: LEA.Inc.
- Fennema, E., & Sherman, J. (1978). Sex-related difference in mathematics achievement and related factors: A further study. *Journal for Research in Mathematics Education*, 9, 189-203.
- Forst, U. A., Hyde, J. S., & Fennema, E. (1994). Gender, mathematics performance, and mathematics-related attitudes affect :a meta-analysis. *International Journal of Educational Research*, 21(4), 373-385.
- Hambleton, R. K., & Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement*, 14, 75-96.
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item response theory*. Newbury Park, CA: Sage.
- Hong, E. & Karstenson, L. (2002). Antecedents of state test anxiety. *Contemporary Educational Psychology*, 27(2), 348-367.
- 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.
- Linn, M. C., & Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researches*, 18, 17-22.
- Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex difference*. Stanford, CA : Stanford University Press.
- Rudner, L. M., Getson, P. R., & Knight, D. L. (1980). Biased item detection techniques. *Journal of Educational Statistics*, 5, 213-233.
- 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.