

科學表現的兩性差異縮小了嗎？—國際 科學表現評量資料之探究*

張郁雯

國立台北教育大學
教育學系

林文瑛

中原大學
心理學系暨心理科學
研究中心

王震武

佛光大學
心理學系

本研究分析 TIMSS 1999、2003、2007 以及 PISA 2006 台灣男女學生在科學表現上之性別差異。所有學生依照科學表現之分數分成四個人數相等的能力群組。檢視四組學生在平均值、標準差以及人數比例上的性別差異。結果發現，雖然整體男女生在科學表現的平均值差異不大，然而，男生分數的標準差一致的高於女生。高分組與低分組平均值差異、標準差比值以及人數比例上大部分有顯著差異。在低分組，女生平均分數高於男生，在高分組，則是男生平均分數高於女生。兩組的標準差都呈現男生大於女生，而高分組男生人數明顯多於女生。TIMSS 與 PISA 在兩性差異上的結果並不一致。八年級的兩性差異似乎比四年級稍微大些。未來研究須注意樣本能力水準並採用多元的統計量數。

關鍵詞：性別差異、科學表現、國際數學與科學教育成就趨勢調查、學

生能力國際評量計畫

兩性在數學與科學學習上的差異一直是許多研究者關注的議題。近年的研究顯示兩性在科學能力上的差異縮小了 (Lohman & Lakin, 2009; Strand, Deary, & Smith, 2006)，甚至某些研究得到女生表現得比男性優異之結果 (Martin, Mullis, & Foy, 2008)。表 1 呈現的是國際數學與科學教育成就趨勢調查 (Trends in International Mathematics and Science Study, 以下簡稱 TIMSS) 於 1999、2003 和 2007 之調查結果中，台灣及國際八年級學生在科學量尺的男女生分數。表 1 顯示，三個年度台灣兩性在科學的整體表現均無性別差異。然而，仔細檢視數據會發現，TIMSS 1999 兩性的科學分數差距雖然明顯，卻未達顯著差異。這是因為 TIMSS 1999 在進行兩性差異檢定時運用了 Bonferroni

* 本篇論文通訊作者張郁雯，通訊方式 yuwenc@tea.ntue.edu.tw。本研究承行政院國家科學委員會專題研究計畫 (NSC 96-2522-S-152 -007 ;NSC 97-2511-S-152 -008 -MY2) 經費補助，謹此致謝。

檢驗來控制多重比較的 α 值，因為國家眾多，因而得到極端保守的估計 (Martin, Gregory, & Steamler, 2000; Martin, Mullis, & Chrostowski, 2004)。單從平均分數差異來看，兩性的差距似乎有縮小的趨勢，在國際的平均分數上尤其明顯，2007 年女生在國際量尺上甚至顯著大於男生。雖然性別分數差異的變化，也可能源自於各屆評量參與的國家不同。然而，比較 TIMSS 三屆科學表現報告對性別差異之分析，1999 的表 1.11 (Martin et al., 2000)、2003 的表 1.4 (Martin, Mullis, Gonzalez, & Chrostowski, 2004)、以及 2007 的表 1.5 (Martin et al., 2008)，從 1999 年參與的 38 國家中，16 個國家男生表現顯著優於女生，沒有國家是女生優於男生。到 2007 年參與的 49 個國家 (地區) 中，男、女無顯著差異的有 24 個國家，其餘 25 個國家中，14 個國家是女生表現優於男生，11 個國家是男生表現優於女生。跨年度參與國家，其兩性平均值的差異多半呈現縮小的趨勢。

一、科學領域性別失衡的現象

如果兩性科學能力表現果然拉近了，那麼在科學相關學習領域上，人數的比例是否較先前接近呢？表 2 呈現的是作者根據教育部統計處所公布的 87 學年度與 98 學年度大學生、碩士生以及博士生在學人數的數據 (教育部, 2010)，計算出主修與科學相關科系人數佔該性別人數的百分比。我們可以發現，兩性差異最大的領域為工程學類，在 98 學年度所有的博士班學生中男生有近半數 (48.5%) 攻讀工程學類，而女生只有 14.3%。此外，大學部兩性在科學相關比例差距 11 年來也沒有縮小的趨勢。表 2 清楚顯示此一事實。美國 National Science Foundation (2007) 的統計結果也與此類似：女性比男性更少選擇數學與科學相關職業的現象並沒有隨著兩性科學能力的拉近而產生大的變化。

為什麼兩性的科學能力已經接近，然而女性最終卻不選擇科學相關工作領域呢？Lent、Brown 和 Hackett (1996) 在他們的社會認知生涯理論 (social cognitive career theory) 主張，人們較不願選擇自認有相當障礙之職業領域。例如，若某職業需要相當好的數學能力，而個人認為自己很難達到該項要求，就不會選擇該職業。換句話說，會選擇科學相關工作領域，其科學能力可能須達某種水準。因此，可能兩性的科學成就平均值雖然接近，但分數分佈的情形不同而有了不同的選擇。例如，可能男生的高水準者較多，女生可能中水準者較多。PISA 的研究報告也指出，雖然兩性科學成就平均值差異不大，但是校內間的性別差異仍是值得持續關注的議題 (Organization for Economic Co-operation and Development, 以下簡稱 OECD, 2007)。

表 1 TIMSS 各次調查八年級學生科學成就之性別差異

	性別	1999		2003		2007	
		女	男	女	男	女	男
台灣	平均值	561	578	571	572	559	563
	標準誤	3.9	5.7	3.8	3.8	3.7	4.4
國際	平均值	480	495*	471	477*	469	463*
	標準誤	0.9	0.9	0.7	0.7	0.8	0.7

表格整理自 Martin et al., 2000; Martin et al., 2004; Martin et al., 2008。* $p < .05$

* 表示兩性差異達顯著水準。

表 2 大學以上主修科學各領域男女學生數佔同級、同性別學生數之比例

	87 學年度						98 學年度					
	博士班		碩士班		大學生		博士班		碩士班		大學生	
	男	女	男	女	男	女	男	女	男	女	男	女
自然科學	7.8	5.6	7.0	4.7	4.1	1.4	7.1	4.6	3.5	2.0	2.4	0.8
數學/電算機學	3.0	2.5	4.4	3.8	8.1	5.5	3.1	2.5	5.6	3.1	9.5	5.5
工程學類	50.4	9.5	46.1	9.2	34.7	4.8	48.5	14.3	40.9	8.8	39.9	5.5

二、影響性別差異效果量之因素

雖然 Hedges 與 Nowell (1995) 曾指出，在檢視性別差異的研究上，最大的缺失在於研究資料不具全國代表性。然國內在探討兩性科學表現差異議題時，可能受限於經費與人力，所使用的樣本卻多屬於區域性 (李明昆、江新合, 2000; 熊召弟, 1996); 而在科學能力的界定上，也因研究者之興趣而著重於特定領域或能力。雖然抽樣和樣本選擇性所造成的偏誤也許是小的，但是因為過去研究顯示兩性的平均差異和變異比不大 (Hyde, Fennema, & Lamon, 1990; Willingham & Cole, 1997)，因此，相對於真正的性別差異，這些偏誤可能不容忽視。雖然有些研究利用整合性分析 (meta-analysis) 希望克服樣本不具全國代表性的問題，然而分析眾多不具代表性的研究並無法讓結果就具有代表性。此外，必須注意的是，平均值與變異數些許的不同會導致分配兩端男性與女性極大的差異 (Hedges & Friedman, 1993)。

台灣參與了 TIMSS 國際教育成就評量後，不少學者開始使用 TIMSS 1999 和 2003 的台灣地區資料進行兩性成就表現的探討。不過這方面的探究，都是進行男女生整體在科學整體分數或科學內容領域分數平均值的差異檢定 (如張翠萍, 2007; 陳麗妃, 2006)，忽略了成就表現分配的變異和型態可能有所不同。當分配的變異不同時，兩性平均值無顯著差異，不代表位於分配兩端的兩性也沒有差異。擁有較大的分配變異之群體 (通常是男性) 在極端分數組會佔較大的比例。因此，研究者認為，只分析兩性平均值差異可能是造成無法解釋「兩性科學能力表現無異，但科學相關主修人數女性卻顯著過少」的原因之一。過去研究顯示，兩性的平均成就差異小，但高分組的男生人數多於女生 (Cleary, 1992; Feingold, 1992; Hedges & Nowell, 1995; Lohman & Lakin, 2009; Nowell & Hedges, 1998; Willingham & Cole, 1997; Strand, Deary, & Smith, 2006)。高分組是科學表現成就較好的一群，他們對主修科學相關科系會較有把握，在入學時也比較可能通過科系對科學成就要求的門檻，因此也比較可能選擇科學相關科系。Willingham 與 Cole (1997) 發現不論是在 GRE 的學科測驗、大學進階先修課程 (Advanced Placement Program) 的學科測驗，還是申請大學須考的學業性向測驗，男生在科學的表現均優於女生。值得注意的是，這類測驗的樣本是自我篩選過的，參與考試者有較大的比例是高分群的樣本。也就是說，在性別差異的表現上，研究樣本能力分佈是個重要因素。選擇不同能力樣本，就可能得到不同的性別差異組型。由於過去台灣兩性科學表現差異的研究樣本缺乏全國代表性，本研究擬使用 TIMSS 和 PISA 的代表性樣本，探討不同科學能力水準的兩性差異。科學能力水準的控制是將成就表現以四分位數區隔成四組。換句話說，本研究第一個希望澄清的問題是，藉由代表性樣本以及能力水準的劃分，是否有可能更清楚呈現兩性科學表現的差異？

由於台灣在三屆 TIMSS 的研究中，兩性科學的總分差距均有變動，因此，本研究將進一步探討兩性科學總分的拉近，是否在分數分配上也顯現出較為接近的型態。雖然不同年度的評量所用的評量工具有所不同。然而，三次測試的評量架構、測驗內容與所需認知歷程大抵相似。因此，本研究也將比較三個年度的資料，探討兩性科學表現差異之趨勢變化。

過去的研究顯示，測量工具是影響兩性科學表現差異的因素之一。Willingham 與 Cole (1997) 發現在 12 歲時，男生的自然科學表現大約高於女生 0.25 個標準差。但若測驗內容需要詮釋材料，則女生的表現優於男生。Hyde 等人 (1990) 的研究顯示數學測驗的認知層次不同，男女性別的表現便會有所不同。不少文獻指出，雖然在標準化測驗上兩性表現並無差異，女生的數學學業成績

卻是高於男性 (Ding, Song, & Richardson, 2007; Hedges & Nowell, 1995; Spelke, 2005)。然而, 也有文獻得到相反的結果, 在標準化數學測試上, 如美國教育進展評量 (National Assessment of Educational Progress), 男生表現還是優於女生 (Coley, 2001)。如果從評量工具的角度來檢視台灣學生參與的 TIMSS 與 PISA, 則會發現 TIMSS 主要在研究學生之數學與科學成就, 其測試架構是建立在學校課程內容上 (Mullis et al., 2003)。反觀 PISA, 其測驗內容是以成人生活所需相關知能為主軸, 並不侷限於學校課程內容, 而是更重視習得知識的活用 (OECD, 2006)。因此, 研究者想探究的第二個問題是, 究竟不同測驗架構導向與兩性的科學表現有無關係? 會不會如同 Cleary (1992) 的發現, 男性在複雜問題解決上表現較佳? 本研究將以分析 TIMSS 以及 PISA 的兩性科學表現來探討測驗架構導向可能扮演的角色。

有學者將兩性在數學、科學、工程與科技工作領域比例懸殊, 女性人數過少的現象比喻成有漏洞的管線 (leaky pipeline), 女性在學習階段一路有許多人離開科學相關領域的管道, 到生涯抉擇時, 女性已所剩無幾 (Blickenstaff, 2005), 而科學表現的差異很可能是從中學階段逐漸顯現出來的 (Cleary, 1992; Willingham & Cole, 1997)。由於在 TIMSS 2003 測試時是四年級的學生, 到了 2007 年剛好升到八年級, 因此, 比較 TIMSS 2003 四年級和 TIMSS 2007 八年級, 應該是能探討同一世代在不同年齡時的縱貫發展變化。此外, 由於 TIMSS 2003 和 2007 同時測試四年級與八年級的學生, 也能提供從橫斷角度檢視兩性科學表現隨年級而變化的情形。如此, 分析不同年度、不同年級的資料將能幫助我們釐清第三個問題: 兩性在科學上的表現差異隨著年級而變化的情形。

三、研究目的

綜合文獻的探討, 顯然許多因素都會影響兩性表現差異, 本研究探討的是其中四類變異來源: 統計指標、能力水準、測量工具與年級差異。具體而言, 本研究的目的有三:

(一) 探討在不同科學能力水準下, 兩性的科學表現之平均值、變異數以及人數之差異, 並檢視兩性差異跨年度的趨勢變化。

(二) 比較 TIMSS 和 PISA 根據不同的科學能力建構而設計的不同測量工具, 是否會反映出不同的性別差異組型。

(三) 檢視兩性科學表現隨年級而變化的趨勢。

研究方法

一、資料來源與樣本

本研究的資料來自 TIMSS 1999、2003、2007 和 PISA 2006 台灣地區的資料。分別由 TIMSS 和 PISA 的官方網站下載。TIMSS 1999 只有測試八年級學生, TIMSS 2003、2007 則包含四年級與八年級學生, PISA 是測試 15 歲在學青少年。因此, 本研究共使用六組資料。

TIMSS 和 PISA 的抽樣設計皆採兩階段的分層叢聚取樣 (two-stage stratified cluster sample design)。第一階段, TIMSS 與 PISA 均運用「等比率機率抽樣」(probability proportional to size) 的原則, 先抽取國中和國小學校, 以便將學校層次和班級層次的平均估計之 95% 信賴區間控制在標準差的 16% 以內。他們先將學校所在地理區域分成北、中、南、東四區, 依照各區人數比例和各區平均班級大小決定各區應選校數。對每個抽樣的學校, 同時指定兩個遞補學校, 以便當選取的學校拒絕加入研究時依序遞補。到了第二階段, TIMSS 是從被抽到的學校中隨機抽取一個班級

(Martin et al., 2004)，而 PISA 則是從學生名單中以隨機抽樣方式抽取約 40 名學生，PISA 在學制上涵蓋了高中、高職、國中、五專以及進修學校，詳細抽樣分佈請參見當次測試之台灣國家報告。

依此抽樣方式，TIMSS 1999 共抽取 5772 人。TIMSS 2003 四年級抽取 4661 人，八年級共抽取 5379 人。TIMSS 2007 四年級抽取 4131 人，八年級共抽取 4046 人。PISA 2006 則有 8815 位 15 歲的受測者。考慮抽樣特性，分析時使用資料庫中所提供的樣本加權變項中的 *house weight* 對樣本進行加權計算。依照資料庫使用手冊的建議，因為每位學生的抽樣權值已知，進行分析時必須加權才能估計得到正確的母群估計值，使用 *house weight* 的優點為確保適當加權，但又將樣本數固定在實際抽樣人數，避免因為加權使得人數激增，影響統計顯著性考驗之正確性 (Martin et al., 2004)。

二、研究變項與統計指標

(一) 科學表現

不同年度的 TIMSS 在科學測試架構有些許的變動，但差異不大，主要是透過參與測試國家針對各國科學課程內容發展而得。TIMSS 1999 科學測試的內容包含生命科學、地球科學、環境與資源議題、科學探究與科學本質、化學與物理六個內容領域，在表現期望上則分為理解、問題解決、使用工具與科學歷程、探索自然世界、溝通五個面向 (Martin et al., 2000)。TIMSS 2003 之科學測試架構，由兩個向度所構成，一為內容向度，在四年級分為生命科學、物理和地球科學，在八年級則為生命科學、物理、地球科學、化學和環境科學。另一向度為認知向度，分為事實知識、概念理解和推理與分析三個向度 (Mullis et al., 2003)。TIMSS 2007，在架構上與 TIMSS 2003 相似，四年級的內容領域為生命科學 (45%)、物理 (35%) 和地球科學 (20%)，八年級則更改為生物 (35%)、化學 (20%)、物理 (25%) 與地球科學 (20%) 四個領域，在認知領域上則為知識、運用、推理三個向度 (Mullis et al., 2005)。本研究使用之科學表現為跨各內容領域估計而得的科學總分。

PISA 對科學素養的定義是使用科學知識、形成 (辨識) 科學議題和使用科學證據。而科學能力的表現端視生活情境脈絡的要求與挑戰，以及個人的科學知識與對科學態度之影響。因之，PISA 科學素養的評量著重「在合理與適當的個人、社會以及全球脈絡下，15 歲青年應該知道、重視以及能做的事。」(OECD, 2006)。強調在生活脈絡下運用科學知識，較不強調傳統的學校科學表現。在命題時，以個人、社會以及全球的脈絡加以思考，包含科學內容知識 (knowledge of science, 佔 60-65%) 以及有關科學本質之知識 (knowledge about science, 佔 35-40%)。前者分為理化、生命、地球與太空以及科技四個系統，後者則細分為科學探究和科學解釋兩個層面。PISA 資料檔並未提供個別學生之科學內容領域分數，僅給予科學知識、形成科學議題和使用科學證據三個能力向度以及整體表現分數，本研究使用的是整體科學表現分數。

由於 TIMSS 和 PISA 均使用矩陣抽樣的技術 (matrix sampling) 分派題本，因此，估計學生的成就分數是運用插補技術抽取五個似真值 (plausible values)。本研究每位學生在整體科學量尺上均有五個似真值。研究中表格呈現的統計數據均是用學生的五個似真值進行五次資料分析後，將所得結果加以平均而得 (Martin et al., 2004)。

(二) 統計指標

本研究在比較兩性差異時，採用了平均值、標準差、標準差比值、人數百分比以及人數比值。以下說明標準差比值以及人數比值之計算方式。標準差比值 (ratios of the male standard deviation to the female deviation, 以下簡稱 SDR) 在本研究是以男學生之科學成就分數標準差除以女學生之標

準差。因此，若 SDR 大於 1 表示，男生分數較女生異質。人數比值則是以男生人數除以女性人數，以下以 M/F 表之。 M/F 若大於 1，表示女生人數少於男生。Willingham 和 Cole (1997) 在比較兩性分數分佈時曾採用類似指標。不過，不同於本研究中以女生作為參照團體，他們是以男生為參照團體。為避免混淆，在本文的敘述中，提及該研究之數據時，均是研究者將參照團體轉換為女生後所得。研究顯示，高分群的男生人數多於女生，且分數變異較大。若以女生為參照團體較容易解讀數據，能知道和女生相比較，男生人數或分數變異是女生的幾倍，而且一般在做平均值檢定時，也是以女性作為參照團體。因此，本研究所有統計指標均採用女生為參照團體。

三、資料分析

六組資料之分析方式是相似的。都是經過以下六個步驟：

(一) **資料與變項的選擇**：選取台灣地區資料，並選擇學生識別碼、加權值、Jackknife 法相關變項 (PISA 資料是選取 80 個 replicate weights)、性別以及科學成就五個似真值變項。

(二) **計算四分位數，進行能力分組**：樣本經加權後，計算出四分位數，以四分位數為切截點，區隔出四個能力組。之所以將受測者依能力分成四個人數相等的組別，是考量分組後，若組內人數過少，會影響男女生人數比例數值的穩定性，組數不宜過多。另外，為使計算各項統計量所根據的人數不至於相去過多，故採人數相等的分組。接著，根據受測者的似真值分數，將他分入對應之能力組。由於每個受測者有五個似真值，而每個似真值所計算出來的四分位數並不相同，因此，最後，每位受測者都會產生五個對應的分組變項。

(三) **計算五個似真值之統計量與抽樣變異**：將 TIMSS 的資料匯入 WesVar 5.1 軟體 (Westat, 2007)。研究者之所以選擇 WesVar 軟體，是因為 WesVar 和 IDB Analyzer (IEA, 2006) 相似能夠處理大型教育資料之抽樣與似真值問題。但其使用彈性較大，可以根據評量資料抽樣的性質，自行產生適當的 replicate weights。TIMSS 的資料，以 house weight 加權並運用 JK2 方法，產生 75 個 replicate weights，再分別計算五個似真值在各四分位組男女之平均值、人數、標準差，以及使用迴歸分析，以性別為預測變項，科學表現似真值為依變項，計算男女平均值之差異以及差異分數之抽樣變異。PISA 的資料在抽樣上略不同於 TIMSS，因此，在抽樣變異的估算上是採用 balanced repeated replication (BRR)，並使用 Fay 的修正程序 (OECD, 2009)。本研究將資料匯入 WesVar 5.1 軟體 (Westat, 2007) 後，使用資料庫已產生的 replicate weights 進行分析。

(四) **計算最終統計量和差補變異 (imputation variance)**：研究者再將五個似真值所計算出來的平均值、標準差、人數加以平均，獲得最終統計量估計值。接著，計算五次統計量之變異，即為所謂的差補變異。

(五) **計算統計量之最終變異**：依據 TIMSS 2003 技術報告 (Martin et al., 2004) 所建議之方法將抽樣變異與差補變異合併，計算出統計量之變異。其公式如下：

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

$\text{Var}(T)$ 是統計量的變異， U 是透過 Jackknife 法計算出來的五次抽樣變異之平均值， B_M 從五個似真值計算而得之差補變異， M 是似真值的個數。

(六) **計算效果量、SDR 以及 M/F 值**：效果量的計算是以該能力水準男女平均值差異除以該組整體分數的標準差。SDR 的統計檢定是先計算男女生的變異數比值，再進行 F 檢定，其自由度分子為該組的男生人數減 1，分母為女生人數減 1。

結果與討論

表 3 到表 8 分別呈現六組資料的男女性別在四個不同成就水準之人數、平均值、效果量、 t 值、標準差、SDR、人數百分比和 M/F 值。組別中「1」、「2」、「3」、「4」分別代表成就表現在 25 百分

等級以下、25 百分等級以上（不含 25）到 50 百分等級之間、50 百分等級以上（不含 50）到 75 百分等級之間以及 75 百分等級以上（不含 75）。以下將從這六個表格，仔細檢視我國兩性的科學表現究竟存在著怎樣的差異？

一、不同科學表現水準之兩性差異具跨年度一致性

以男性與女性整個群體來看，六組資料中僅兩組有顯著性別差異，都是男生的科學表現顯著的高於女生，分別是 TIMSS 1999 八年級以及 TIMSS 2003 四年級，效果量分別為 0.2 和 0.1 都屬於小的效果量（Cohen, 1988）。台灣近期的研究，整體男女性別之科學表現平均值無顯著差異。這個結果與國際的趨勢是一致的，以 TIMSS 2007 四年級為例，36 個國家中，有 22 個國家科學成就無性別差異，其餘 14 個有差異的國家中，其中有 6 個是女生的表現優於男生。在八年級的部分則是，參與的 49 個國家中，男、女無顯著差異的有 24 個國家，其餘 25 個國家中，14 個國家是女生表現優於男生，11 個國家是男生表現優於女生。美國在四年級男女生無顯著差異，但在八年級是男生優於女生。

表 3 TIMSS 2003 的四年級樣本能力分組之各項統計數字及性別差異指標

組別	性別	人數	平均值	效果量	t 值	標準差	SDR	百分比	M/F
1	女	585	466.78	-0.21	-2.76**	37.98	1.15***	50.16	.99
	男	581	458.12			43.79		49.84	
2	女	605	531.98	0.04	0.57	13.10	.98	51.87	.93
	男	561	532.51			12.89		48.13	
3	女	566	574.91	0.06	0.64	12.44	1.01	48.53	1.06
	男	600	575.64			12.54		51.47	
4	女	490	633.09	0.14	2.03*	28.55	1.08*	42.08	1.38
	男	675	637.26			30.86		57.92	
全體	女	2245	547.90	0.10	2.70*	64.71	1.11***	48.16	1.08
	男	2416	554.57			71.91		51.84	

人數是經過 house weight 加權所得。所呈現數字是四捨五入到整數位，因此，細格的總和與全體的人數會有些許出入。

* $p < .05$ ，** $p < .01$ ，*** $p < .001$ 。

表 4 TIMSS 2007 的四年級樣本能力分組之統計數字及性別差異指標

組別	性別	人數	平均值	效果量	t 值	標準差	SDR	百分比	M/F
1	女	486	460.67	-0.18	-2.38*	41.94	1.15***	47.05	1.13
	男	547	452.48			48.29		52.95	
2	女	541	534.87	0.05	0.46	14.73	1.00	52.36	.91
	男	492	535.64			14.80		47.64	
3	女	513	583.39	0.03	0.22	14.18	1.01	49.64	1.01
	男	520	583.80			14.34		50.36	
4	女	457	650.58	0.06	0.88	32.60	1.07	44.29	1.26
	男	575	652.69			34.98		55.71	

全體	女	1997	555.74	0.02	0.74	73.17	1.11***	48.34	1.07
	男	2134	557.59						

* $p < .05$, *** $p < .001$.

表 5 TIMSS 1999 的八年級樣本能力分組之各項統計數字及性別差異指標

組別	性別	人數	平均值	效果量	t 值	標準差	SDR	百分比	M/F
1	女	767	456.05	-0.12	-1.43	52.66	1.08*	53.18	.88
	男	675	449.08						
2	女	808	544.97	0.07	0.85	17.29	.99	56.01	.79
	男	635	546.15						
3	女	748	601.06	0.08	0.95	15.95	1.02	51.85	.93
	男	695	602.36						
4	女	574	669.93	0.28	3.16***	33.57	1.19***	39.77	1.51
	男	869	680.56						
全體	女	2898	560.66	0.20	4.24***	82.80	1.13***	50.20	.99
	男	2874	577.56						

人數是經過 house weight 加權所得。所呈現數字是四捨五入到整數位，因此，細格的總和與全體的人數會有些許出入。

* $p < .05$, *** $p < .001$.

表 6 TIMSS 2003 的八年級樣本能力分組之各項統計數字及性別差異指標

組別	性別	人數	平均值	效果量	t 值	標準差	SDR	百分比	M/F
1	女	638	470.70	-0.20	-3.32**	41.23	1.16***	47.41	1.10
	男	707	461.65						
2	女	674	550.05	-0.01	-0.15	16.82	.999	50.10	.97
	男	671	549.81						
3	女	681	602.12	0.00	0.03	14.41	1.00	50.66	.98
	男	663	602.14						
4	女	607	663.18	0.19	2.90**	28.23	1.13***	45.13	1.24
	男	738	669.09						
全體	女	2599	570.63	0.01	0.29	74.49	1.12***	48.33	1.07
	男	2780	571.52						

人數是經過 house weight 加權所得。所呈現數字是四捨五入到整數位，因此，細格的總和與全體的人數會有些許出入。

** $p < .01$, *** $p < .001$.

表 7 TIMSS 2007 的八年級樣本能力分組之各項統計數字及性別差異指標

組別	性別	人數	平均值	效果量	t 值	標準差	SDR	百分比	M/F
1	女	484	446.91	-0.23	-3.45***	49.45	1.16***	47.82	1.09
	男	528	434.25						
2	女	530	540.94	0.00	-0.01	19.19	.97	52.43	.91
	男	481	540.79						
3	女	526	597.17	0.07	0.75	15.20	.99	52.00	.92

	男	486	598.20			15.11		48.00	
4	女	417	661.59	0.18	2.43**	31.85	1.13**	41.20	1.43
	男女	594	667.78			35.86		58.80	
全體	女	1957	558.50	0.05	1.38	82.51	1.15***	48.36	1.07
	男	2089	563.35			95.11		51.64	

** $p < .01$, *** $p < .001$.

表 8 PISA 2006 的 15 歲學生樣本能力分組之各項統計數字及性別差異指標

組別	性別	人數	平均值	效果量	t 值	標準差	SDR	百分比	M/F
1	女	1089	406.95	-0.06	-1.08	43.69	1.03	49.52	1.02
	男	1107	404.06			45.12		50.48	
2	女	1094	504.98	0.00	-0.08	20.98	.99	50.57	1.02
	男	1114	504.90			20.88		49.42	
3	女	1043	570.48	0.05	.78	17.94	1.00	47.19	1.12
	男	1163	571.29			17.97		52.81	
4	女	970	647.21	0.04	.53	34.63	1.04	43.19	1.27
	男	1232	648.77			36.00		56.81	
全體	女	4196	528.72	0.08	.52	93.00	1.03*	47.62	1.10
	男	4616	535.85			95.63		52.38	

* $p < .05$

在分數的變異方面，六組資料的男女整體之 SDR 都大於 1，且均達顯著水準，顯示兩性科學表現分數變異不同，男生的科學成就表現的變異都比女生大。TIMSS 的資料顯示男性的標準差大約大女生 11% 到 15%，而 PISA 的資料兩者的標準差較為接近。Willingham 與 Cole (1997) 綜合七個自然科學測驗所得的 SDR 為 1.14。Lohman 與 Lakin (2009) 檢視 1984、1992 和 2000 三個年度美國具代表性樣本，在數量的認知能力測驗上其 SDR 介於 1.07 到 1.25。若只看四年級到九年級這個年齡層，則 SDR 介於 1.08 到 1.18。Lindberg、Hyde 與 Linn (2010) 整合分析 1990 到 2007 的美國大型資料研究，得到變異數比為 1.09。雖然研究的對象屬於不同國度，蒐集資料的時間也有十幾年的差距但所得的數據似乎頗為穩定。男女生人數抽樣比例，僅有 TIMSS 1999 是女生略多於男生，其他五個資料的男女比例在 1.07 到 1.10 間。

這種在身體與心理特質上男生比女生有較大的變異現象，在文獻上通稱為變異假說 (variability hypothesis)，早於 19 世紀就由達爾文提出，認為透過變異，男性對演化發揮較大的影響力。Havelock Ellis 主張男性變異較大來自於先天能力的差異。因此，男性有較多的天才，智能不足的比例也較高。Thorndike、Jensen 等均持此種觀點。Hollingworth 是第一個系統性蒐集資料駁斥男性變異大乃源自於先天的學者，她認為社會因素，如學習機會以及對女性的歧視與偏見是造成男性有較多傑出表現的因素 (Feingold, 1992)。Wai、Cachio、Putallaz 與 Makel (2010) 認為在優秀族群男女比例的由 13.5 比 1 下降到 4 比 1 左右，顯見社會文化因素扮演重要角色。但近 20 年，雖然平均值拉近，變異數比卻呈現穩定的狀態，其中的緣由值得後續研究加以探究。

下節呈現不同能力群組的性別差異，大體而言，TIMSS 各年度的分析結果相似性大於相異性，具跨年度之穩定性。重要的發現有：高分與低分組兩性的平均值差異顯著，但效果量小；這兩組男女的分數標準差比值大於 1；高分組男生人數多於女生。PISA 的資料呈現不同風貌，將於下一節進一步討論。結果顯示，在檢視兩性科學表現差異時，樣本的科學能力水準是重要的調節變項。

(一) 高分組與低分組性別差異多數達顯著，但效果量小

以 TIMSS 的資料檢視各不同能力群組內男女性別之科學表現差異，得出一個有趣的組型。不論整體男女性別是否有表現差異，高分組和低分組通常會出現性別差異；在低分組時，均是女生的表現優於男生；而高分組則是男生的表現高於女生。在 10 個比較中，8 個達統計顯著水準（表 3，表 6 及表 7 的高分和低分組，表 4 之低分組以及表 5 之高分組），效果量之絕對值介於 0.14 到 0.28，中位數為 0.19，屬於小的效果量。近年的資料似乎有低分組男女差異的效果量大於高分組的趨勢。PISA 的資料雖與 TIMSS 資料有相似的趨勢，但均未達統計顯著水準。

(二) 高分組與低分組的標準差比值 (SDR) 大於 1

在成就表現變異方面，結果很明顯的與平均值差異的組型是相似的。居中間分配的中低與中高兩組，其 SDR 接近 1；而低分和高分兩組其 SDR 皆大於 1，12 組中有 9 組達顯著水準（表 3，表 5，表 6，表 7 之高分及低分組，以及表 4 之低分組），顯示在兩個極端組，男生的分數分佈較為離散。在低分組，SDR 介於 1.03 到 1.16；而高分組則介於 1.04 到 1.19，未達顯著的 3 組有兩組是來是表 8 PISA 2006，另一組是表 4 之 TIMSS 2007 四年級資料組，但該組的資料在顯著邊緣。若以 TIMSS 的資料來看，不同年度間，這個比值並無多大的變動。

(三) 高分組男生人數比例多於女生

在 M/F 指標上，由於每年抽樣的男女比例有些許差距，因此，在解釋各組資料時，應該參照該年度全體男、女性別比例。五組資料的抽樣都是男性略多於女性。在六個資料組中，很一致地發現，高分組男生比例多於女生。在高分組中，男女比例差異最大的資料，出現在表 5 TIMSS 1999 之資料，男生的比例是女生的 1.5 倍，相對於抽樣樣本男女比例為 0.99。也就是說高分組男女人數比為 3 比 2。男女比例差距最小的出現在表 6 TIMSS 2003 八年級的資料。高分組的男生是女生的 1.24 倍，相對於抽樣樣本男女比例為 1.07。而在其他指標都未有明顯差異的 PISA 資料組，高分組也出現男生多於女生的型態。雖然高分組男生比例高於女生的結果，能由男生平均值高於女生，且男生分數變異較女生大推知，但實際 M/F 值仍須視兩性平均值差異以及 SDR 兩者共同作用而定。

(四) 科學能力水準是重要的調節變項

上述的結果顯示，若以傳統採用整體性別平均值差異作為標準，僅有較早期研究 TIMSS 1999 和 TIMSS 2003 四年級是有顯著差異。PISA 2006 和 TIMSS 2007 不論是四年級或是八年級均無性別差異。這和近年來大型評量資料所獲致的結論是一致的 (Lindberg, Hyde, Petersen, & Linn, 2010)。但當考量受試者的能力水準，則 TIMSS 2007 原本整體男女表現無顯著差異的，位於分配兩端的學生，兩性之間仍然存在平均值差異以及標準差比值的差異。平均值差異再加上男生的分數變異大於女生的差異，導致 M/F 指標上顯現出高分組男女性別比例的明顯差異。未來的研究在探討兩性差異時，應該注意研究對象的能力水準。

這個結果意味著進行兩性差異研究時，樣本之能力水準是重要的調節變項，研究對象之能力水準不同，將獲致不同的結論。這可能是過去探究兩性科學表現差異之研究，得到不一致結果的原因之一。如果抽測的是中間能力水準的樣本，容易得到無性別差異的發現。如果是以能力表現較佳者為受測者，則可能得到男性表現優於女性的結果。

根據科學表現水準加以分組，再進行兩性差異比較，所得到的發現能部分解釋何以數理工程相關領域，仍是以男性為主。2003 年國中基測人數約為 32 萬，這屆學生在 97 學年度大學錄取人數約為 11 萬。換言之，接近三分之一的學生會進入大學。也就是說，在 TIMSS 2003 科學表現在平均值以上的群組，是可能進入大學就讀的主要族群。而若以前述社會認知生涯理論來看，人們會往較有把握的領域發展，那麼檢視高分組的男女表現差異，是比檢視全體受測樣本更能看到男女科學成就表現與大學主修科系領域間的關係。本研究分析結果提供一個解釋大學生理工相關科系，兩性人數比例懸殊的一個可能的因素。不過，高分組兩性人數比約為 3 比 2，而大學理工科系兩性人數比例約為 4.4 比 1。顯然，造成兩性在理工相關領域比例如此懸殊尚有其他因素。相關研究顯示動機因素如興趣、科學價值、科學能力信念等與能力因素共同作用影響兩性科系選擇意向 (蔡秀芬、張慧貞, 2006; Su, Rounds, & Armstrong, 2009)。晚近，部份學者主張性別偏好可能是造成科學領域女性比例偏少的重要因素。女性偏好「人」，而男性則偏好「事」(Ceci, William, &

Barnett, 2009; Su et al., 2009)，因而，在科學學門裡，女生最感興趣的學科是生物，而從事物理相關工作的女性較少。

另一個值得重視的發現是近幾年學者特別關注的，台灣學生的低分群之人數比例增加的問題。此外，各組資料的低分群均出現男性表現低於女性的現象，而且從四年級開始就相當明顯。如果這是個跨學科都存在的現象，那麼學校中哪些因子不利於男生的學習，將是個值得進一步探究的重要教育議題。事實上，美國近幾年對於男童是否有學習上的危機已經有不少的討論，有一部份的學者甚至認為現今的學校體制是對男童不友善的環境，並將此問題稱為「新的性別差異 (new gender gap)」，呼籲改變教學措施，主張單一性別的學校（班級）制度（Whitmire, 2010）；然而，相對於此，也有另一部份的學者則認為所謂的男童危機根本不存在（Sadowski, 2010）。因此，台灣顯然需要有更多的研究數據來釐清這個問題。

此外，也應有更多的研究應積極探討造成高分群女生能力落後的原因，並針對此發展有效的教學策略，例如，Anglin、Pirson 與 Langer（2008）以教學實驗證實，採用有意義的學習方式，男女生能達到一樣水準的表現，但是若使用傳統教學方式，則男生表現優於女生。

二、TIMSS 和 PISA 科學表現的兩性差異並不一致

比較 PISA 的結果（表 8）與 TIMSS 的結果（表 3 到 7）的結果，可以明顯的發現，兩個不同評量研究在性別差異的結果上出現不一致的現象。PISA 2006 在各項指標上，均指向兩性無差異，僅有男生的分數標準差略大於女生，以及在高分群男生的比例高於女生。而不同年度 TIMSS 的研究在性別差異的結果上卻是十分一致的，都呈現極端兩組有顯著性別差異的現象。過去的研究指出若問題比較接近教學，女生的表現會比較好，也因此女生通常在校成績較好，而男生則是比較能夠解決較不熟悉，較少練習的問題（Cleary, 1992; Coley, 2001; Ding, Song, & Richardson, 2007; Lohman & Lakin, 2009; Willingham & Cole, 1997）。TIMSS 的題目與學校的科學課程有較為緊密的結合，而 PISA 則是較注重日常生活中的科學問題題目。依照過去研究結果推論應是在 TIMSS 上兩性表現之差異會小於 PISA。然而，台灣 PISA 2006 的研究結果卻得到與過去國外研究不一致的結論，反而是 PISA 的表現無性別差異。

造成此種不一致發現的原因可能是，在內容領域的界定上，TIMSS 是完全以科學學科內容界定；而 PISA 則分為科學內容與科學本質兩類知識。若單就 PISA 科學知識內容的表現來看，如表 9 所呈現，台灣男學生在各個次內容領域均優於女學生。TIMSS 2007 的結果則是八年級男生在物理和地球科學的表現上，顯著優於女生。而女性則在化學領域有較佳的表現，至於在生物領域則是女生與男生無異，或女生有較佳的表現。單就 PISA 科學內容表現來看，則台灣的研究結果仍與國外結果相似。

表 9 PISA 2006 兩性在科學內容領域分數之平均值與標準誤

	地球與太空系統	生命系統	理化系統
男生	537 (3.5)	556 (3.7)	558 (3.6)
女生	520 (4.3)	542 (4.7)	532 (4.5)

整理自 OECD (2007). PISA 2006 Vol 2: Data. () 內為標準誤。三個領域均達顯著差異。

由於科學包含幾個次領域，而男女生各有其擅長的領域，因此，各個次領域的比重不同也可能導致最後的科學總分呈現不同的性別差異現象。所以，在討論兩性在科學成就上的差異時，應該要特別注意科學內涵之界定。

除了上述評量工具問題外，兩個評量研究樣本選擇不同也可能是造成研究結果差異的原因。PISA 的測試年齡為 15 歲，在臺灣是跨國中與高中、高職以及專科等學制。由於進入高中職階段前經過升學篩選，不同學制的學生能力不同，而且進入後接受科學相關課程也有所不同。不同學制的男女生比例可能有所不同，這也可能造成兩性表現差異不同於 TIMSS 的結果。

三、隨著年級上升，兩性差異有微幅增大的趨勢

由於 TIMSS 2003 四年級學生到了 2007 年剛好是八年級，雖然兩個年度所抽的樣本不同，但都是具全國代表性的樣本，可以藉此探討發展趨勢。從表 3 和表 7 的比較可以發現，原本四年級時男女生全體存在著性別差異，到八年級時已不復存在。但若以 SDR 為指標來看，則 SDR 從四年級的 1.11 稍微增加為八年級的 1.15。若從各能力群組加以比較，則可發現兩者的組型並無多大改變，低分組都是男生表現比女生差，高分組都是男生表現比女生好，但八年級的效果量、SDR 呈現微幅擴大，在 M/F 指標上則是低分組和高分組男生的比例都增加了。在效果量和 SDR 上，高分組的變化似乎比低分組來得大一些，但在 M/F 比例上，則是低分組有較大幅度的改變。綜合而言，相較於四年級，八年級低分組男生的平均表現比女生更加落後，且男生所佔的人數也在增加中。八年級高分組男生則其平均表現超越女生的幅度比四年級增加，其分數分佈也更為離散，得到高分的人數也增加了。換句話說，所有的指標均指向男女生的差距似乎在擴大中，有趣的是，整體平均值的方向卻是恰恰相反的。

研究者進一步對照比較了「TIMSS 2003 四年級(表 3)和八年級(表 6)」，以及「TIMSS 2007 四年級(表 4)與八年級(表 7)」，這兩組同一年度抽取的橫斷性樣本之結果。雖然兩個對照樣本除了有年齡上的不同，還有世代間的差異，而且兩個不同年級所測試之科學內容也會有所不同，仍然可以作為檢視發展趨勢的輔助資料。就同一年度的兩個不同年級之性別差異比較而言，大體上，四年級和八年級的相似性多於相異性。TIMSS 2003 兩個不同年級的比較，和過去印象不同的是男女生整體的平均值在四年級有顯著差異，在八年級卻沒有顯著差異。在 SDR 上改變也不大。細看各四分位數分組的統計指標，變化比較明顯的是八年級高分組的效果量增加，SDR 增大，但 M/F 比例卻是下降的。相對而言，八年級低分組的主要變化是 M/F 的比例增加了。

至於 TIMSS 2007 兩個不同年級的比較，則和 TIMSS 2003 四年級(表 3)與 TIMSS 2007 八年級(表 7)以同一世代進行比較的結果較為相似。男女生整體的平均值在兩個年級均不顯著，SDR 從 1.11 增為 1.15。低分組和高分組的效果量、SDR，到八年級都略微增加，在高分組增加更為明顯，平均值差異檢定由四年級的不顯著到八年級達統計顯著。在 M/F 的比較上，則是高分組有較大幅度的增加。整體來看，男女生的差距也是有微幅的擴大。

綜合上述結果，似乎可以看到隨著年級上升，男女生的差距在分配兩端之組別有些許擴大的現象。但男女整體平均值的變化，不是沒有改變，就是和過去的印象相反，在四年級是達顯著差異，到八年級卻沒有明顯差異。觀察年級的變化，無可避免的會涉及不同年級測量工具的可比較性問題，再加上兩性間的差異微小，使得發展趨勢的檢視更為困難。本研究利用全國代表性樣本以不同統計指標對不同能力群組進行細部分析，卻發現高低分組兩性差異隨著年級有微幅增加的趨勢，高中階段的變化情形值得後續研究持續關注。

結論與建議

雖然有不少研究探討性別差異的問題，但是多數研究只進行平均值的差異檢定，而且樣本通常不具全國代表性。本研究使用六組具全國代表性的資料檢視兩性在科學表現的差異。獲致的結論為整體而言，兩性平均值差異是微小的，效果量皆小於 0.20。然而，值得注意的是兩性的分數

分配並不同質，男性分數的標準差約大於女性分數標準差 11-15%，而這個比值呈現跨年度的穩定性。跨年度的資料比較結果，如一般研究所指稱，整體男女兩性的平均值差異在近幾年是縮小的。不過，如果在各能力群組內，以效果量、SDR 以及 M/F 指標來比較兩性差異，則本研究發現，在四個能力群組中，跨年度兩性差異組型是一致的：兩性在低分組以及高分組中在各項統計指標多半出現差異，並無明顯縮小趨勢。此一現象值得研究者以及科學教育工作者繼續探究成因，並尋求改善的可能機制。

「兩性能力的差異如此小，何以大學以上，修習理工相關領域學系的女性僅佔全體女性 15% 以下，與男性介於 40-50% 的比例，有極大的差距？」造成此一現象的原因眾多，本研究結果指出一個可能的原因是男女兩性在高分組的人數比例不等。由於上大學的人是篩選過的族群，在全國樣本上屬於平均值以上的中高分組，再加上個人即使具備修習理工領域所需的能力，個人也會同時考量相對於其他領域，他須付出多少努力方能成功。因此，兩性的平均值差異雖然小，若加上男性分數變異較大，則能導致高分族群男女生人數比例有極大的差距。在我們的資料組中，高分組的男性比例最高可達 1.5 倍。如果只分析分數在前百分之十的族群，則男女兩性的比例將高達 2:1 以上 (Hedges & Friedman, 1993)。Wai 等人的整合分析顯示，過去 20 年來，美國科學推理測驗表現前百分之一的學生，男女生比例穩定地維持在 3 比 1 (Wai et al., 2010)。顯然，當女性的人數只有男性的一半甚至更低時，要達到在理工相關工作領域兩性比例平衡將是困難的工作。我們認為這個分析也能部分解釋「何以大型成就評量顯示兩性在數理成就上的差距極小，但一般大眾仍舊傾向認為男性的理工能力較好？」這個現象。因為社會大眾很可能將焦點放在看得到的工作場域或大學主修領域的男女分佈，而不是科學成就測驗的男女平均值差異。這樣的結果也衍生出一個重要的方法論議題，那就是針對不同性別差異問題，研究者必須注意其研究族群之能力水準。在探討科系選擇之兩性差異時，有必要探究高分族群。當兩性分數分配的變異比值增大時，微小的平均值差異，也能造成在高分族群兩性人數出現極為懸殊的比例。而持續致力於研究開發有效提升女性數理能力之教學策略，以增加高分群女性的比例，進而提高女性就讀理工科系的人數，仍是科學教育界值得努力的方向之一。

本研究針對 TIMSS 和 PISA 兩個不同評量研究資料分析之結果，呈現出不同的兩性差異型態。PISA 2006 在各項指標上，均指向兩性無差異，僅有男生的分數標準差略大於女生，以及在高分群男生的比例高於女生。而 TIMSS 不同年度的研究卻一致的發現，高分與低分兩組有顯著的性別差異。評量工具內容不同很可能是造成不一致結果的主要因素。由於 TIMSS 與 PISA 兩個評量研究之資料庫除了有個別學生之整體科學成就（素養）分數外，也有認知層次能力分數，TIMSS 的資料庫尚有各內容領域的分數，未來研究可以針對這些次內容領域能力以及各認知層次能力的兩性差異再進行深入探究。另外，2012 年 PISA 和 TIMSS 也將同時進行另一輪的教育成就評量，屆時可檢視本研究的發現是否可以類推到下一次的評量資料。

最後，在發展的趨勢上，研究得到一個看似矛盾的結果。那就是在男女整體平均值上，在四年級有顯著差異，到八年級卻沒有差異。然而，若檢視各能力群組的各項統計指標，則可以看到在低分組和高分組兩性的差異有微幅擴大的趨勢。造成兩項結果方向不一致的一個可能因素是，由於兩性的差異即使是達顯著水準，其效果量都不大；而八年級的分數變異比四年級來得大，且 SDR 也增大。若跨年級整體平均值差異量並未改變，但變異數卻變大，自然容易得到不顯著的統計結果。在極端兩組的情形，平均值的增加幅度大於標準差的變化，因此，效果量增加。這個發展的改變趨勢值得未來研究繼續追蹤與探究。

總而言之，兩性科學表現差異是個複雜的議題，無法只是簡單地說，兩性存在著差異或沒有差異。研究樣本的能力水準（受測者居於全國分數分配的那個位置）、受測者的年級、測量工具的評量架構（評量工具與課程的關係密切嗎？）以及所使用的統計指標等多種因素都會影響兩性差異的組型。根據本研究之發現，研究者建議，未來進行兩性差異相關研究時，應特別敘明樣本能力水準，以及採用多元的統計指標。在評量工具上，則應說明各個科學內容領域所佔的比重、認知向度以及出題的情境。

參考文獻

- 李明昆、江新合 (2000)：國民中學學生科學素養之研究—以垃圾焚化爐議題為例。《科學與教育學報》，4，3-20。
- 教育部 (2010)：重要教育統計資訊。取自教育部統計處網站：
http://www.edu.tw/statistics/content.aspx?site_content_sn=8956，2010年9月4日。
- 張翠萍 (2007)：TIMSS 2003 六國國小四年級學生在科學認知領域表現及其相關因素之研究。國立新竹教育大學人資處應用科學系碩士論文。
- 陳麗妃 (2006)：TIMSS 2003 國小四年級學生背景、家庭環境、科學興趣、自信與科學成就關係之比較分析：以七國為例。國立新竹教育大學進修部課程與教學碩士論文。

- 熊召弟 (1996)：台北公立高中（高一）學生科學過程技能和邏輯思考能力之探討研究。台北師院學報，9，545-578。
- 蔡秀芬、張慧貞 (2006)：呼喚女性物理新生代：2005 女性物理人下鄉活動迴響與心得。物理雙月刊，28，610-614。
- Anglin, L. P., Pirson, M., & Langer, E. (2008). Mindful learning: A moderator of gender differences in mathematics performance. *Journal of Adult Development, 15*, 132-139.
- Blickenstaff, J. C. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and Education, 17*, 369-386.
- Ceci, S. J., William, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135*, 218-261.
- Cleary, T. A. (1992). *Gender differences in aptitude and achievement test scores. In sex equity in educational opportunity, achievement, and testing. Proceedings of the 1991 ETS Invitational Conference* (pp. 51-90). Princeton, NJ: Educational Testing Service.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coley, R. J. (2001). Differences in the gender gap: Comparisons across racial/ethnic groups in education and work. Princeton, NJ: Educational Testing Service.
- Ding, C. S., Song, K., & Richardson, L. I. (2007). Do mathematical gender differences continue? A longitudinal study of gender difference and excellence in mathematics performance in the U.S. *Educational Studies, 40*, 279-295.
- Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62*, 61-84.
- Hedges, L. V., & Friedman, L. (1993). Gender differences in variability in intellectual abilities: A reanalysis of Feingold's results. *Review of Educational Research, 63*, 94-105.
- Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*, 41-45.
- Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139-155.
- International Association for the Evaluation of Educational Achievement (2006). IDB analyzer. Retrieved April 3, 2009, from http://www.iea.nl/iea_studies_datasets.html.
- Lent, R. W., Brown, S. D., & Hackett, G. (1996). Career development from a social cognitive perspective. In D. Brown, L. Brooks, & Associates (Eds.), *Career choice and development* (3rd ed., pp. 373-421). San Francisco, CA: Jossey-Bass.

- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*(6), 1123-1135
- Lohman, D. F., & Lakin, J. M. (2009). Consistencies in sex differences on the cognitive abilities test across countries, grades, test forms, and cohorts. *British Journal of Educational Psychology, 79*, 389-407.
- Martin, M. O., Gregory, K. D., & Steamler, S. E. (2000). *TIMSS 1999 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Gregory, K. D., Smith, T. A., Chrostowski, S. J., Garden, R. A., & O' Connor, K. M. (2000). *TIMSS 1999 international science report: Findings from IEA's repeat of the third international mathematics and science study at the eighth grade*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., & Chrostowski, S. J. (2004). *TIMSS 2003 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 international science report: Findings from IEA's repeat of the third international mathematics and science study at the eighth grade*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Martin, M. O., Mullis, I. V. S., & Foy, P. (with Olson, J.F., Erberber, E., Preuschoff, C., & Galia, J.) (2008). *TIMSS 2007 international science report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College, MA.
- Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzalez, E. J., Chrostowski, S. J., & O' Conner, K. M. (2003). *Assessment frameworks and specifications 2003*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora A., & Erberber, E. (2005). *TIMSS 2007 Assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- National Science Foundation (2007). *Women, minorities, and persons with disabilities in science and Engineering*. Arlington, VA: National Science Foundation. Retrieved Sep 4, 2007, from <http://www.nsf.gov/statistics/wmpd/start.htm>
- Nowell, A., & Hedges, L.V. (1998). Trends in gender differences in academic achievement from 1960 to 1994: An analysis of differences in mean, variance and extreme scores. *Sex Roles, 39*, 21-43.
- Organization for Economic Co-operation and Development (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*.

- Organization for Economic Co-operation and Development (2007). *PISA 2006, Vol 2: Data*. Retrieved Dec 24, 2007 from <http://www.pisa.oecd.org>
- Organization for Economic Co-operation and Development (2009). *PISA 2006 Technical Report*. Retrieved March 17, 2009, from <http://www.pisa.oecd.org>
- Sadowski, M. (2010). Putting the “boy crisis” in context. *Harvard Education Letter*, 26(4). Retrieved September 24, 2010, from <http://www.hepg.org/hel/article/473#home>
- Spelke, E. S.(2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*, 60, 950-958.
- Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in cognitive ability test scores: A UK national picture. *British Journal of Educational Psychology*, 76, 463-480.
- Su, R., Rounds, J. & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin*, 135, 859-884.
- Wai, J., Cacchio, M., Putallaz, M., & Makel M. C. (2010). Sex differences in the right tail of cognitive abilities: A 30 year examination. *Intelligence*, 38, 412-423.
- Westat (2007). *WesVar 4.3 User's Guide*. Retrieved September 5, 2009, from http://www.westat.com/westat/statistical_software/WesVar/index.cfm
- Whitmire, R. (2010). *Why boys fail: Saving our sons from an educational system that's leaving them behind*. New York, NY: AMACOM.
- Willingham, W. W., & Cole, N. S. (Eds.) (1997). *Gender and fair assessment*. Hillsdale, NJ: Lawrence Erlbaum Associate.

收稿日期：2011年04月08日

一稿修訂日期：2011年08月10日

二稿修訂日期：2011年09月08日

三稿修訂日期：2011年10月28日

接收刊登日期：2011年10月28日

Bulletin of Educational Psychology, 2013, 44(Special), 459-476
National Taiwan Normal University, Taipei, Taiwan, R.O.C.

Is Gender Gap in Science Performance Closer? Investigating Data from International Science Study

Yuwen Chang

Department of Education
National Taipei University
of Education

Wenyng Lin

Department of Psychology
Research Center for Psychological
Science
Chung-Yuan Christian University

Jennwu Wang

Department of Psychology
Fo-Guang University

TIMSS 1999, 2003, 2007 and PISA 2006 data were used to investigate gender differences in the science performance. Each student was assigned to one quartile group (i.e., Q1 to Q4) based on his or her science performance. For each group, gender differences in mean scores, standard deviations, and numbers were examined. Results indicate that gender differences in mean scores are generally small. However, variances for males are higher than those for females. Most of gender differences in mean, standard deviation ratio, and male-to-female sample ratio from groups below Q1 and above Q3 are significant. Females outperform males in the group below Q1, while males outperform females among high-scoring group. In those two groups, the scores of male have larger variance. Males outnumber females substantially among high performance group. The patterns for gender differences from TIMSS and PISA are not the same. Gender gap in science performance score is slightly larger in grade 8 than in grade 4. It is suggested that in future research. The sample be better specified in terms of ability levels and gender differences be examined by multiple statistical indicators.

KEY WORDS: gender difference, PISA, science performance, TIMSS