

# 「第三次國際數學與科學教育成就研究 後續調查」之抽樣設計

羅珮華

國立臺灣師範大學 科學教育中心

## 摘要

第三次國際數學與科學教育成就研究後續調查(Third International Mathematics and Science Study-Repeat, 簡稱 TIMSS-R) 採用三階分層群集抽樣設計(three-stage stratified cluster sample design) 來抽選各國的受測學生, 這三個階層包括學校、班級、以及班級內的抽樣, 同時在抽樣的過程中必須考量各種必要的排除條件, 使符合國際調查抽樣的正確性和調查資料品質的要求。

關鍵字: TIMSS-R、抽樣設計、第三次國際數學與科學教育成就研究後續調查

## 前言

國際教育成就調查委員會(The International Association for the Evaluation Achievement, IEA) 在 1993-1995 年舉辦第三次國際數學與科學教育成就研究(Third International Mathematics and Science Study, 簡稱 TIMSS)之後, 有許多國家紛紛表示希望能有後續的調查, 除了要取得有關學生在數學和科學科目中的知識和能力的資訊, 以及有關文化環境、教學實施狀況、課程目標和教育上的安排等所伴隨學生成就的資訊外, 希望瞭解當年在 TIMSS 研究中九歲群(國小四年級)的學生, 經過四年升上八年級(國中二年級)之後的學習成就, 更可以比較這兩群相隔四年的八年級學生的學習成就間是否有差異, 於是 IEA 在 1997 年展開第三次國際數學與科學教育成就研究後續調查(Third International Mathematics and Science Study-Repeat, 簡稱 TIMSS-R 或稱為 TIMSS 1999)。對於之前沒有參加 TIMSS 而新加入後續調查的國家而言, 雖然無法對其國內學生進行趨勢比較, 但仍可以與其他國家進行比較(Foy, 1998)。後續調查研究計畫由美國 Boston College 主持, 抽樣工作由加拿大國家統計局(Statistics Canada)負責。

我國國立臺灣師範大學科學教育中心曾於 1987 年經由 IEA 同意, 引用「第二次國際科學學習成就調查研究」(The Second International Science Study, 簡稱 SISS)的調查工具進行測驗, 希望將我國國小、國中及高中學生數學及科學成就與其他國家學生成就做比較。由於此次測驗並不是與各國同時進行, 且不在國際監控範圍, 因此國際上並沒有將我國此次測驗的結果納入討論與比較(魏明通, 1988)。1990 年 IEA 開始推動進行 TIMSS 研究時, 我國未能參與; 直到 TIMSS-R 展開後, 為了解我國學生數學與科學教育學習成就, 並與世界主要國

家互相觀摩溝通，在國科會積極努力下，使我國得以正式參加此一國際性的成就調查研究。

國立臺灣師範大學科學教育中心使用「第二次國際科學學習成就調查研究」的調查工具進行測試時，採用二維矩陣之「分層隨機抽樣」來選取受測的班級學生（魏明通，1988）。TIMSS-R 的抽樣設計仍是分層抽樣，但不全是隨機抽樣，而是借重 TIMSS 的經驗稍微改良的三階分層群集抽樣設計（three-stage stratified cluster sample design），使其符合分析與實際運作的需求(Foy & Joncas, 1999)。由於實地參與我國 TIMSS-R 的研究調查工作，個人希望將我國進行的 TIMSS-R 抽樣過程提供給有興趣者參考。

### TIMSS-R 受測對象

此次調查研究是延續 1995 年的 TIMSS 研究調查，為了使同時參加 TIMSS 和 TIMSS-R 的國家可以得到一個具有比較性的成就資料趨勢，特別選定研究的學生母群為 TIMSS 中的「Population 2」，因此 TIMSS-R 受測學生的年級界定與 TIMSS 相同，即是「在測驗時，包含最多 13 歲群的在學學生的年級」。決定受測年級需要靠各國的國家聯絡人(National Research Coordinator, NRC) 結合各國教育體系的訊息來判斷，各國的受測學生應該是哪一個或哪二個年級。

我國在決定受測學生年級時，是依照我國對兒童入學年齡之規定，以當年度九月一日前滿六足歲者必須就讀國小一年級來計算，因此我國包含最多 13 歲群的在學學生的年級是國中二年級，即相當於其他國家所謂的八年級。而在測驗結果中，國際上的所有受測學生平均年齡是 14.4 歲，我國所有受測學生的平均年齡是 14.2 歲，相當符合 TIMSS-R 的調查對象之要求。

TIMSS-R 調查研究的對象除了學生之外，尚有受測學生的數學和科學課程教師（我國包括生物和理化教師），以及學校校長或相關人員，他們分別需要填寫教師問卷和學校問卷，提供用來了解不同的教學安排與學生成就相關的訊息。

### TIMSS-R 抽樣設計

#### 一、抽樣單位與抽樣人數

一個研究調查結果的正確程度全賴抽樣設計的品質與抽樣實施的過程，由於 TIMSS-R 的分析焦點在於學生的累積學習和影響學習的教學特性，因此探討的對象層級包含了學生、班級、學校等，而這些層級就成了抽樣單位(sampling unit)。TIMSS-R 的抽樣設計要求各階層抽樣時，都必須符合抽樣的正確性和調查資料品質的要求(Foy & Joncas, 1999)。

調查資料的信度來自於製作可信的學生特質評估，因此 TIMSS-R 希望每一個國家都能選

出足夠且具有代表性的受測學生，同時對樣本母群參數和抽樣變數提出正確的加權評估。在 TIMSS-R 抽樣精確性的標準中，要求每一個探討的變數至少要有 400 個有效的學生作答資料，這樣才能使的抽樣誤差不會大於從 400 個學生簡單隨機抽樣所產生的誤差。TIMSS-R 預計要分析學校和班級層次的資料，原則上以班級作為抽樣單位時，因此會有班級內的群集效應(clustering effect)存在，為了避免抽樣的精確性下降，TIMSS-R 藉由以前 TIMSS 的經驗，對於影響群集效應的班級大小與班級內的相關(intraclass correlation)大小進行評估。包括以各國各年級的班級平均人數做為群集大小，並選定班級內的相關是 0.3，根據抽樣設計表(表一)決定抽樣的學校數量。最後決定各國都從受測學生母群中選取 150 所學校，每校一個班級接受測驗調查，這樣可以在學校和班級層次產生 95%的信賴界限(confidence limits)，而且標準差在  $\pm 16\%$  以內(Foy & Joncas, 1999)。

表一 抽樣設計表：在 95%的信賴界限下，預估的學生人數  $\pm 0.1\%$  的標準差。最低學校樣本需求：150 校。(Foy & Joncas, 1999)

每校受測人數*		班級內的相關( intraclass correlation)								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
5	a	150	150	189	221	253	285	317	349	381
	n	750	785	945	1105	1265	1425	1585	1745	1905
10	a	150	150	155	191	227	263	299	335	371
	n	1500	1500	1550	1910	2270	2630	2990	3350	3710
15	a	150	150	150	180	218	255	292	330	367
	n	2250	2250	2250	2700	3270	3825	4380	4950	5505
20	a	150	150	150	175	213	251	289	327	365
	n	3000	3000	3000	3500	4260	5020	5780	6540	7300
25	a	150	150	150	172	211	249	287	326	364
	n	3750	3750	3750	4300	5275	6225	7175	8150	9100
30	a	150	150	150	170	209	248	286	325	364
	n	4500	4500	4500	5100	6270	7440	8580	9750	10920
35	a	150	150	150	169	208	246	285	324	363
	n	5250	5250	5250	5915	7280	8610	9975	11340	12705
40	a	150	150	150	168	207	246	285	324	363
	n	6000	6000	6000	6720	8280	9840	11400	12960	14520
45	a	150	150	150	167	206	245	284	323	362
	n	6750	6750	6750	7515	9270	11025	12780	14535	16290
50	a	150	150	150	166	205	245	284	323	362
	n	7500	7500	7500	8300	10250	12250	14200	16150	18100

\* 每校受測人數通常用平均班級人數來代替。

a 抽取的學校數

= 抽取的學生人數

## 二、分層抽樣

分層抽樣的設計通常是為了提高抽樣設計效率或信度，使抽樣結果能夠兼顧各種族群的分布，而更具有代表性。TIMSS-R 的分層變數可以是外顯的(explicit)或是內隱的(implicit)，例如地理上分區是一個外顯的分層變數，用來考量學校的不均勻分布情形，因此將所有學校名

「第三次國際數學與科學教育成就研究後續調查」之抽樣設計

單分成不同區，再進行抽樣；學校分布較均勻的情況下，無須特別分區抽樣，此分層變數即為內隱的。

TIMSS-R 全體抽樣設計採用三階分層群集抽樣設計，第一階抽樣單位是學校，由各國自行決定外顯或內隱的分層變數，然後利用機率等比例(probabilities-proportional-to-size, 簡稱 PPS) 系統抽樣方法來抽取受測學校（表二是 PPS 系統抽樣的範例）。這種系統抽樣方式的優點在於操作方便，且容易監控執行過程是否恰當。如果被抽到的受測學校無法施測時，改由候補學校受測；因此抽取受測學校時，同時也將每個受測學校的第一順位和第二順位候補學校名單抽取出來。這個階段是由加拿大國家統計局根據各國提供的資料來抽取受測學校。

第二階抽樣單位是受測學校內的班級。每一個國家都是根據加拿大國家統計局提供的受測學校名單進行班級抽樣工作，這個階段是採用亂數隨機抽樣，通常是在每一個受測學校抽出一個受測班級，但是有些國家也許會因為一個班級學生人數太少而需要抽出一班以上的學生來受測。有些國家的學生並不是所有課程都在同一班一起上課，因此 TIMSS-R 界定班級時，是以所有在同一個數學課程班級上課的學生稱為一個班級。如果這個數學班級學生成員來自不同年級時，只有屬於受測年級的學生需要接受測試。如果受測學校無法界定出班級時，只好從受測年級中的所有學生隨機抽樣抽出受測學生。所有教過受測班級或受測學生的數學和科學課程的教師，必須填答教師問卷。

第三階抽樣單位是受測班級內的學生。基本上被抽到的受測班級內的所有學生都應接受測驗，但是有些國家的受測班級內的人數過多，必需再從該班進行部分抽樣，以亂數隨機選取適當的人數參加測驗。所以本階段的抽樣並不是每一個國家都需要使用，我國 TIMSS-R 抽樣過程只進行了前面二個階段，即已經完成學生受測名單。

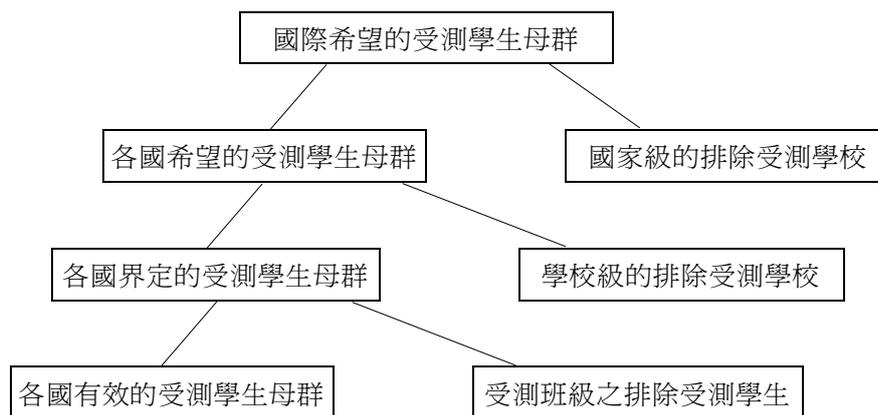
表二 這是利用一個虛構的抽樣架構來說明 PPS 系統抽樣方法的運用，其中條件是：抽取 150 所學校，所有學校學生總人數是 392154 人，抽樣區間是 2614.3600，起始亂數是 1135.1551。(Foy & Joncas, 1999)

學校代碼	各校受測年級學生總人數	累進所有受測年級學生人數	抽樣
172989	532	532	
976181	517	1049	
564880	487	1536	受測學校
387970	461	1997	第一順位候補
483231	459	2456	第二順位候補
550766	437	2893	
228699	406	3299	
60318	385	3684	
201035	350	4034	受測學校
107346	341	4375	第一順位候補

294968	328	4703	第二順位候補
677048	311	5014	
967590	299	5313	
644562	275	5588	
32562	266	5854	
194290	247	6101	
129135	215	6316	
1633	195	6511	受測學校
256393	174	6685	第一順位候補
754196	152	6837	第二順位候補
750793	133	6970	
757843	121	7091	
743500	107	7198	
84930	103	7301	
410355	97	7398	

### 抽樣過程與排除受測的條件

由於每一個國家可以依據未來正式測驗實施時的費用、複雜度、困難度、地理因素與特小型學校等各種相關考量，而在不同的抽樣階段決定排除受測的學校、班級、或學生名單，各階段的排除受測名單與受測學生母群的關係如圖（一）(Foy& Joncas, 1999)。但是每一個國家排除受測的比例不可以超過該受測年級學生總人數的 5%。



圖（一） 不同階段的排除受測名單與受測學生母群的關係(Foy& Joncas, 1999)。

依照國際的時程規定，各國要正確估算排除學校的比例與班級的學生平均人數，並將各種抽樣相關表格寄至加拿大國家統計局，再由加拿大國家統計局依照 PPS 方式，將試測(Field Test)與實測(Main Survey)受測學校與候補學校名單抽選出來。我國依照國際排除受測的規

定，將啓聰、啓明、啓智等特殊學校列於排除受測名單；由於實施調查時需抽樣訪視，基於天氣、交通、經費等因素考量下，我國將離島學校列於排除受測名單。扣除排除學校名單之後的所有公私立國民中學、完全中學、附設國中部學校，即是我國學校抽樣的母群。而在第一階段分層取樣時，我國提供兩項條件作為分層依據，一是依照地理位置分為四區：新竹以北，含宜蘭為北區；苗栗至雲林為中區；嘉義以南為南區；臺東和花蓮為東區。另外是以學校受測年級班級數區分為大、中、小型學校，其中 13 班（含）以上為大型學校，6-12 班為中型學校，5 班（含）以下為小型學校。

由於我國當年度教育統計資料尚未出爐，於是科學教育中心發函至所有國中，調查各學校一、二年級的班級男女學生分布、班級人數、以及各校可排除受測的附設特殊班級如啓聰班、啓明班、啓智班等訊息。統計結果，各校一年級可排除受測班級的人數總和約為當時我國所有國中一、二年級各年級總人數的 0.8%，低於國際不可以超過 5% 的規定。當時我國國中一、二年級最高的班級平均人數是 60 人，最低的班級平均人數是 8 人，各年級班級總平均人數皆約為 39 人，加拿大國家統計局後來將我國學校班級平均人數低於 20 人的 15 所學校列於排除抽樣名單，目的希望能夠有足夠的學生人數接受調查。

在確立試測與實測的受測學校與候補學校之後，各國再根據本計畫所提供的抽樣手冊和電腦軟體 W3S(Within-School Sampling)進行班級或學生的抽樣。我國國中各班課程基本上都是以班為單位，且同一個班級學生都在一起上課，因此我國 TIMSS-R 抽樣是以班級為單位。在抽樣受測學生時，必需先向已經被選定的受測學校確認排除受測班級，然後依照電腦 Within-School Sampling 程式進行亂數隨機抽樣。受測班級挑選出來之後，首先通知各校並再向各個學校收集受測學生個人相關資料，以便製作測驗紀錄表格，並請學校轉知學生即將接受此一調查。雖然校內抽樣過程由我國自己進行，但仍接受國際監控，需於抽樣完成時將抽樣電腦程式寄回總部確認所有抽樣過程是否符合規定。

學校內的排除受測學生的條件主要有三項：(1)學習遲緩的學生：此類學生意指經由校長或其它合格的教職員工的“專業”意見認定為學習遲緩的學生，或者是經由心理醫師認定為學習遲緩的學生。此類學生包括了情緒性或心智上不能照著即使最一般的測驗說明去做的學生。學業成績不良或管教上有問題的學生不可以此原因將之排除在外。(2)肢體殘障學生：此類學生意指為永久性的身體殘障，導致無法參加 TIMSS-R 測驗。但肢體雖然殘障，可以回答測驗問題者仍須被包含在測驗之內。(3)母語有困難者：學生實質上無法讀或說本測驗的語言，而且在考試狀態下無法克服語言障礙。尤其是學習本測驗所使用的語言的時間未逾一年者；其他的狀況者均應被包含於本測驗之中。

我國在校內抽樣結果後，同時調查學生受測名單與排除受測學生名單，結果需接受實測調查的 5889 位學生中有 37 位可以排除受測，詳見表一。

表一 我國 TIMSS-R 受測學生名單中之各類排除受測學生的人數統計

排除因素	學習遲緩的學生	肢體殘障學生	母語有困難者	總計
人數	13	6	18	37

### 誌謝

感謝國科會補助 TIMSS-R 調查研究經費（NSC 87-2511-S-003-058 和 NSC 89-2511-S-003-007）、IEA 同意使用測驗的相關資料、所有受測學校的配合與協助；同時也感謝國立臺灣師範大學理學院各系教授們與科學教育中心全體同仁協助 TIMSS-R 調查工作，使調查計畫順利完成。

### 參考資料

1. 魏明通、許榮富、沈青嵩（1988）：國際科學學習成就調查研究報告（國內報告）。國立臺灣師範大學科學教育中心。
2. Pierre Foy (1998). *Third international mathematics and science study – repeat (TIMSS-R) School Sampling Manual - Version 2*. Chestnut Hill, MA: Boston College.
3. Pierre Foy & Marc Joncas (1999). *TIMSS-R technical report: sample design (draft)*. Chestnut Hill, MA: Boston College.