



TEPS-B 技術報告第 2 號

**建構 TEPS-B Panel 1 SH 樣本
2010 年面訪調查完訪成功樣本的機率權數**

第一版

關秉寅 詹傑勝

民 國 一 〇 四 年 十 一 月

目錄

一、前言.....	1
二、TEPS-B 2010 年面訪調查抽樣設計.....	2
三、以不同機率權數分別加權推估之比較.....	4
四、結論與討論.....	13

一、 前言

「『台灣教育長期追蹤資料庫』後續調查」(簡稱 TEPS-B)於 2010 年時以面訪追蹤方式調查部份曾於 2001 年及 2003 年接受 TEPS 追蹤調查的高中職五專樣本(簡稱為 Panel 1 SH 樣本)。TEPS-B 曾於 2009 年時針對同一樣本進行全面電話追蹤調查。此電話追蹤調查的結果即成為 2010 年面訪調查抽樣設計的重要參考。雖然 2009 年的正式電話調查追蹤到 10,546 人,¹但蒐集到的資料有限,因此需藉由 2010 年面訪調查來深入瞭解這群年輕人自高中職五專三年級以後就學就業的情況。2010 年面訪問卷的內容包括了詳細的教育史、工作史及居住史,也有與社會網絡、全球化行為、非認知能力、性別與婚姻態度、身體與心理健康等相關的題目。²

TEPS-B 2010 年面訪調查共完訪了 3,977 人,初步整理出來的資料,目前已經由中央研究院調查研究專題中心之學術調查研究資料庫(簡稱 SRDA)釋出,提供學界使用。³由於此面訪調查之樣本為由原先 TEPS 樣本抽樣取得,且並非所有預定接受調查者均接觸到或完訪,因此面訪調查的完訪樣本可能有樣本偏誤的問題,從而無法代表原先的樣本或原樣本欲推估的母群體。本技術報告的目的即在評估此偏誤的情況,並說明使用此面訪調查完訪樣本時,可如何建構適當機率權數,以作為推估之用。

¹ 請參考 TEPS-B 技術報告第 1 號 (http://tepsb.nccu.edu.tw/?page_id=1199) 的說明。

² TEPS-B 2010 年面訪調查問卷詳細內容可於 http://tepsb.nccu.edu.tw/?page_id=660 瀏覽或下載。

³ 此項資料可至中央研究院調查專題研究中心學術調查資料庫網站 <https://srda.sinica.edu.tw/group/scigview/2/27> 瀏覽說明,並申請下載。

二、 TEPS-B 2010 年面訪調查抽樣設計

TEPS-B 於 2010 年面訪調查的 Panel 1 SH 樣本為由 TEPS 於 2001 年及 2003 年追蹤調查之高中職五專樣本中抽取的機率樣本。TEPS 有完整可追蹤紀錄之樣本數為 19,051 人。這群年輕受訪者於 2010 年時是約 25 至 26 歲，大多已經進入勞力市場，且散佈在台灣各地，甚至在國外求學或就業。因此，TEPS-B 依經費、人力、未來長期追蹤需求，以及與 TEPS 另一較年輕世代樣本（即當時之國中樣本）之面訪追蹤結果的比較等因素考量，將預定完訪數訂為 4,200 人。此外，參考了 2009 年電訪 Panel 1 SH 樣本居住地區分佈的情況後，2010 年面訪調查之抽樣設計的步驟如下：

- 1、先以 TEPS 當年調查之學校為初級抽樣單位（primary sampling units），抽取百分之六十的高中職五專學校，如此選出的潛在受訪人數約為一萬三千人。依此抽樣方式選出之一萬三千人經分析後發現與原先二萬人之樣本在學習表現的預測分數之分配上並無顯著差異。
- 2、為考量執行面訪時需能有效率的運用訪員人力，本計畫進一步將所有這些取出之樣本所在的地區建立抽樣區域，然後再以一階段按規模大小成比例（PPS）的抽樣方法，從中抽出分散在一百二十個抽樣區域的六千多人為預定面訪對象（其中包括 2009 年電訪未能接觸者）。經此多階段抽樣設計，可以計算出每位預定面訪對象的抽樣機率。目前已經釋出之 2010 年面訪調查資料中提供了每位受訪者的機率權重，變項名稱為 sh10wt。

2010 年面訪調查最後完訪數為 3,977 人，雖然相當接近原訂完訪數的目標，但仍非所有預定調查到的樣本都接觸到或完訪，因此完訪樣本有偏誤的可能。為評估此可能偏誤的情況，以下使用 SRDA 釋出之 TEPS 公共版資料，並依 TEPS 提供之原來樣本的權數來推估一些重要背景變項的平均數或比例，以及依據標準誤而估計出的 95% 信賴區間，作為 TEPS-B 追蹤樣本之權數推估結果的比較基準。本報告將分別以釋出之 Panel 1 SH 完訪樣本資料提供之 sh10wt 來推估同樣變項的平均數或比例及其 95% 信賴區間。然後，也用原先 TEPS 提供之權數與 TEPS-B Panel 1 SH 資料提供之權數相乘，並用新建構的權數來估計同樣變項的平均數或比例，以評估以此簡易方式建構之新權數是否會比只用 sh10wt 加權的推估來得更接近用以比較的基準。

三、 以不同機率權數分別加權推估之比較

(一) TEPS 第二波 w2stwt2 加權與 TEPS-B 2010 年 sh10wt 加權的比較

表一呈現以 2003 年 TEPS 第二波資料提供的機率權數 w2stwt2 加權估計受訪者在 2001 年第一波及 2003 年第二波一些背景變項的平均值或比例。表中所估計的變項包括第一波調查時綜合分析能力測驗 IRT 分數、第一波是否就讀私立學校、父母是否為結婚狀態、第一波男性比例、第一波學校在鄉村比例、第一波學校在城鎮比例、第一波學校在都市比例、以及第二波就讀普通學程、綜合學程、職業學及五專學程等之比例。表二呈現的是以 2010 年 TEPS-B 面訪追蹤資料中提供的機率變項 sh10wt 來加權估計同樣變項的平均數或比例。⁴

首先，比較表一與表二所得估計值的異同。以第一波綜合分析能力而言，使用第二波調查的 w2stwt2 權數加權估計的 IRT 分數平均值是 1.1872，95% 的信賴區間在 1.1387 到 1.2358 之間，而使用 2010 年 TEPS-B 面訪調查提供的 sh10wt 權數加權估計得到的平均 IRT 分數則是 1.7319，比前者估計的平均分數高，而其 95% 的信賴區間在 1.6798 及 1.7840 之間，與前者沒有重疊。以 w2stwt2 加權估計第一波是否就讀私立學校的比例，得到的數值是 0.4992，95% 信賴區間則是在 0.4827 與 0.5157 之間；而以 sh10wt 加權的估計，則顯示有 39.02% 的受訪者第一

⁴ 本技術報告的分析均使用 Stata 13.1。如以 w2stwt2 加權為例，其執行程式為：
 mean w1all3p w1priv married male rural town city academic comprehen vocational jrcollege [pw=w2stwt2]，其中 w1all3p 為 TEPS 第一波綜合能力分析測驗分數，w1priv 為第一波就讀私立學校，married 為父母為結婚狀態，male 為受訪者為男性，rural、town 及 city 則是學校所在地的三種都市化程度，academic、comprehен、vocational 及 jrcollege 則分別是普通、綜合、職業及五專等四種學程。

波受訪時就讀私立學校，明顯低於 49.92%，其 95% 信賴區間則落在 0.3680 及 0.4123 之間，兩者範圍也沒有重疊。因此，TEPS-B 2010 年面訪追蹤調查完訪樣本在高中職五專三年級時的平均綜合分析能力表現比原來樣本要高，且比較偏向當時就讀公立學校者。

表一、使用 TEPS 第二波機率權數 w2stwt2 所得之估計值 (N=12,179)

	平均數 / 比例	標準誤	95% 信賴區間	
第一波綜合能力測驗	1.1872	0.0248	1.1387	1.2358
第一波私立學校與否	0.4992	0.0084	0.4827	0.5157
父母是否為結婚狀態	0.8683	0.0066	0.8554	0.8812
第一波學生男性比例	0.4775	0.0085	0.4610	0.4941
第一波城鄉地區				
鄉村	0.0691	0.0052	0.0588	0.0793
城鎮	0.4063	0.0081	0.3904	0.4222
都市	0.5247	0.0084	0.5082	0.5411
第二波學程類型				
普通	0.3590	0.0071	0.3452	0.3729
綜合	0.1047	0.0046	0.0956	0.1138
職業	0.4458	0.0089	0.4283	0.4632
專科	0.0905	0.0027	0.0853	0.0957

使用 TEPS 第二波加權變項 w2stwt2 估計父母是否為結婚狀態，得到的比例是 86.83%，95% 信賴區間落在 0.8554 與 0.8812 之間。而使用 TEPS-B 2010 年加

表二、使用 2010 年加權變項 sh10wt 所得之估計值 (N=2,650)

	平均數 / 比例	標準誤	95% 信賴區間	
第一波綜合能力測驗	1.7319	0.0266	1.6798	1.7840
第一波私立學校與否	0.3902	0.0113	0.3680	0.4123
父母是否為結婚狀態	0.9009	0.0069	0.8873	0.9146
第一波學生男性比例	0.4820	0.0117	0.4591	0.5048
第一波城鄉地區				
鄉村	0.0454	0.0060	0.0335	0.0572
城鎮	0.3989	0.0117	0.3759	0.4218
都市	0.5557	0.0118	0.5326	0.5789
第二波學程類型				
普通	0.4741	0.0116	0.4513	0.4969
綜合	0.1032	0.0064	0.0906	0.1158
職業	0.2427	0.0102	0.2227	0.2627
專科	0.1800	0.0088	0.1627	0.1973

權變項 sh10wt 加權估計父母結婚比例，則得到較高比例，有 90.09%，95% 信賴區間與前者沒有重疊，落在 0.8873 與 0.9146 之間。以 w2stwt2 加權估計第一波學生的男性比例時，得到 0.4775，95% 信賴區間落在 0.4610 與 0.4941 之間。而使用 2010 年 sh10wt 變項估計的性別比例則得到 0.4820，男性比例較高一些，其 95% 信賴區間值範圍較廣，涵蓋前者估計範圍，落在 0.4591 與 0.5048 之間。因此，TEPS-B 2010 年面訪完訪樣本男性比例比 TEPS 原來樣本要高些，且高中職五專時期父母是結婚狀態的比例也較高些。

在 2001 年第一波居住城鄉比例部分，以 w2stwt2 估計的比例中，有 6.91% 受訪者學校是在鄉村，95% 信賴區間落在 0.0588 與 0.0793 之間。而使用 sh10wt 加權所估計的比例，則顯示只有 4.54% 的受訪者學校在鄉村地區，其 95% 信賴區間與前項估計範圍沒有重疊，在 0.0335 與 0.0572 之間。以 w2stwt2 加權估計第一波調查時受訪者學校在城鎮的比例，得到的數值是 40.63%，其 95% 信賴區間則是在 0.3904 以及 0.4222 之間。而以 2010 年加權變項估計該波調查受訪者住在城鎮的比例，則是 0.3989，數值與前者估計值相近，其 95% 的信賴區間落在 0.3759 與 0.4218 之間，與前項的估計範圍大致上重疊。第一波調查時受訪者學校在都市的比例最高，以 2003 年加權變項所估計的比例是 52.47%，95% 的信賴區間估計落在 0.5082 以及 0.5411 之間，然而以 2010 加權變項進行估計，得到的比例則較高，有 55.57%，其 95% 信賴區間則落在 0.5326 與 0.5789 之間，範圍與前者估計區間重疊。因此，TEPS-B 2010 年面訪調查完訪樣本當年就讀高中職學校在城鎮及都市的比例與原來 TEPS 樣本相近，但如受訪者當年是就讀於鄉村地區學校者，則 2010 年時被追蹤到的比例偏低。

在受訪者接受第二波調查時就讀學程類型方面，2003 年加權變項估計就讀普通學程的比例是 35.90%，95% 信賴區間落在 0.3452 與 0.3729 之間，而使用 2010 年加權變項估計，得到的比例較高，有 47.41% 受訪者就讀普通學程，其 95% 信賴區間則是在 0.4513 與 0.4969 之間，兩者估計區間不重疊。在綜合學程部分，以 2003 年第二波加權變項所估計的比例是 10.47%，95% 信賴區間在於 0.0956 與

0.1138 之間。以 2010 年加權值加以估計，得到的就讀綜合學程比例則是 10.32%，其 95% 信賴區間則是在 0.0906 與 0.1158 之間，後者估計區間與前者大致相同並涵蓋前者。以 2003 年的加權變項去估計第二波受訪時就讀職業學程的比例，得到 44.58%，其 95% 信賴區間估計落在 0.4283 以及 0.4632 之間。而使用 2010 年加權變項去估計，得到就讀職業學程的比例卻明顯較低，有 24.27%，其 95% 信賴區間則落在 0.2227 與 0.2627 之間，兩者估計區間不重疊。第二波受訪時就讀專科學程的比例，以 2003 年加權變項 $w2stwt2$ 估計，得到的是 9.05%，95% 信賴區間值落在 0.0853 以及 0.0957 之間，以 2010 年加權變項去估計，就讀專科學程比例則較高，有 18.00%，其信賴區間則落在 0.1627 以及 0.1973 之間，與前者的估計範圍沒有重疊。因此，在當年就讀高中職五專學程方面，TEPS-B 2010 年面訪調查完訪樣本與原來 TEPS 樣本相比，追蹤到的受訪者就讀普通及五專學程的比例偏高，就讀職業學程的則偏低，就讀綜合學程的則與原來比例大致相近。

(二) TEPS 第二波 $w2stwt2$ 加權與 TEPS-B 2010 年 $sh10wt*w2stwt2$ 加權的比較

前述以 $w2stwt2$ 加權分析及 $sh10wt$ 加權分析的比較可知，TEPS-B 2010 年面訪調查雖然是依機率抽選原來的 TEPS 樣本，但如果只用 TEPS-B 2010 面訪資料提供之機率權數加權推估一些背景變項的母數的話，應是有偏誤的。造成此偏誤的基本原因之一是此權數是以 TEPS 原來樣本為母群體來設計的，並未考慮用 TEPS 原來樣本推估時就需要加權。因此用 TEPS-B 2010 年面訪調查資料做推估時，其權數應做進一步的調整。最簡單的調整方法之一，就是將 TEPS 第二波的

權數 w_{2stwt2} 與 TEPS-B 2010 年面訪調查提供的權數相乘，並以此相乘得到的權數來加權推估。表三即是以 2010 年加權變項乘上 2003 年第二波調查加權變項， $sh_{10wt} * w_{2stwt2}$ ，推估先前用來評估偏誤之各變項的平均值或比例。

以下說明比較表一與表三中分別使用 TEPS 第二波 w_{2stwt2} 加權及 TEPS-B 2010 年加權變項乘上 2003 年第二波加權變項（即 $sh_{10wt} * w_{2stwt2}$ ）加權得到之估計值的差異。以 2003 年加權變項 w_{2stwt2} 估計第一波綜合能力測驗分數得到的值是 1.1872，而以 $sh_{10wt} * w_{2stwt2}$ 所得估計值為 1.2776，與 2003 年 w_{2stwt2} 加權變項估計所得之值頗為接近。2003 年估計的 95% 信賴區間落在 1.1387 與 1.2358 之間，相乘加權變項得到的估計值則落在 1.1530 與 1.4023 之間，與 2003 年加權變項估計所得之區間重疊。

使用 2003 年 w_{2stwt2} 估計第一波調查時就讀私立學校比例，得到 49.92%，而使用 $sh_{10wt} * w_{2stwt2}$ 估計的比例則是 51.93%，接近以 w_{2stwt2} 之估計。2003 年加權變項估計 95% 信賴區間在 0.4827 與 0.5157 之間， $sh_{10wt} * w_{2stwt2}$ 的估計則在 0.4787 及 0.5598 之間。換言之，兩區間估計有重疊的部份。父母結婚的比例以 w_{2stwt2} 估計得到 86.83%，以 $sh_{10wt} * w_{2stwt2}$ 加權估計，得到的比例則是 90.22%。 w_{2stwt2} 加權估計的信賴區間在 0.8554 與 0.8812 之間，此區間估計被包含在 $sh_{10wt} * w_{2stwt2}$ 加權得到之區間估計 0.8761 與 0.9283 的範圍之內。至於，第一波男性比例部分，使用 2003 年 w_{2stwt2} 變項估計，得到 47.75%，而以 $sh_{10wt} * w_{2stwt2}$ 相乘後估計的比例則較高，是 52.88%。以 w_{2stwt2} 加權估計的

區間在 0.4610 與 0.4941 之間，sh10wt*w2stwt2 加權得到的估計則在 0.4881 以及 0.5695，二者有重疊處。

在高中職五專時期就讀學校所在地部份，以 2003 年 w2stwt2 加權估計居住在鄉村地區比例，得到 6.91%，以 sh10wt*w2stwt2 估計的鄉村比例則是 4.14%，相較下仍偏低。以 w2stwt2 估計的 95% 信賴區間在 0.0588 與 0.0793，此區間估計包含在 sh10wt*w2stwt2 估計 95% 信賴區間 0.0215 與 0.0613 之內。而以 w2stwt2 估計位於城鎮地區的比例是 40.63%，以 sh10wt*w2stwt2 估計的比例則是 41.23%，比例雖高一些，但二者的 95% 信賴區間相互重疊。w2stwt2 估計的信賴區間在 0.3904 與 0.4222 之間，sh10wt*w2stwt2 估計的則是在 0.3708 與 0.4537 之間。再者，居住在都市地區的比例，若以 w2stwt2 估計，得到 52.47%，而以 sh10wt*w2stwt2 估計則是得到 54.64%，較 w2stwt2 估計較高，但二者的 95% 信賴區間估計也重疊。w2stwt2 的估計區間在 0.5082 與 0.5411 之間，sh10wt*w2stwt2 的估計在 0.5047 與 0.5880 之間，涵蓋前者估計區間。

至於 TEPS 年第二波調查時就讀學程類型部分，就讀普通學程的比例，以 w2stwt2 估計得到的是 35.90%，以 sh10wt*w2stwt2 估計，則是得到 37.90%，比只用 sh10wt 加權估計的低了 10%。在信賴區間估計方面，w2stwt2 估計區間在 0.3452 與 0.3729 之間，sh10wt*w2stwt2 的估計在 0.3427 與 0.4153 之間，故包含前者的估計區間。TEPS 第二波調查時就讀綜合學程的比例，以 w2stwt2 加權估計得到的是 10.47%，以 sh10wt*w2stwt2 加權估計則得到 12.55%，稍高一些，但

表三、使用 2010 年加權變項 sh10wt 乘上第二波加權變項 w2stwt2 所得之估計值
(N=2,621)

	平均數 / 比例	標準誤	95% 信賴區間	
第一波綜合能力測驗	1.2776	0.0636	1.1530	1.4023
第一波私立學校與否	0.5193	0.0207	0.4787	0.5598
父母是否為結婚狀態	0.9022	0.0133	0.8761	0.9283
第一波學生男性比例	0.5288	0.0207	0.4881	0.5695
第一波城鄉地區				
鄉村	0.0414	0.0102	0.0215	0.0613
城鎮	0.4123	0.0211	0.3708	0.4537
都市	0.5464	0.0212	0.5047	0.5880
第二波學程類型				
普通	0.3790	0.0185	0.3427	0.4153
綜合	0.1255	0.0145	0.0971	0.1539
職業	0.4239	0.0221	0.3805	0.4673
專科	0.0716	0.0060	0.0599	0.0833

二者的 95% 信賴區間估計則是重疊，w2stwt2 估計落在 0.0956 以及 0.1138 之間，sh10wt*w2stwt2 落在 0.0971 與 0.1539 之間，估計區間較大。在第二波調查時就讀職業學程類型的比例上，以 w2stwt2 加權估計得到的比例是 44.58%，以 sh10wt*w2stwt2 估計得到的比例則有 42.39%，二者頗接近。使用 w2stwt2 加權估計的 95% 信賴區間落在 0.4283 及 0.4632，sh10wt*w2stwt2 的估計落在 0.3805 與 0.4673 之間，涵蓋了以 w2stwt2 加權估計的區間。最後，在估計 TEPS 第二波調查時就讀專科學程的比例上，以 w2stwt2 估計得到的比例是 9.05%，以

sh10wt*w2stwt2 估計的比例則只有 7.16% , 且二者估計的 95% 信賴區間沒有重疊。 w2stwt2 估計的信賴區間在 0.0853 及 0.0957 之間 , sh10wt*w2stwt2 則落在 0.0599 及 0.0833 之間。

四、 結論與討論

前節分析顯示，如以 2003 年 TEPS 第二波提供之加權變項 $w2stwt2$ 估計 TEPS 第一波及第二波一些受訪者基本背景變項的平均值或比例，並以此為基礎來評估使用 TEPS-B 2010 年面訪追蹤調查完訪樣本之適當加權的話，可看出直接用 TEPS-B 2010 年面訪資料提供的加權變項 $sh10wt$ 所估計的結果，與用 $w2stwt2$ 的估計大多有明顯的差異。但如以 $w2stwt2$ 乘上 $sh10wt$ 後得到的新加權變項來估計的話，則能明顯適當修正前述估計上的差異，但仍有無法調整處。因此，研究者也可考慮使用 TEPS-B 技術報告第 1 號建議的權數調整策略，即先計算出 TEPS-B 2010 年被追蹤完訪的機率，然後以此機率的倒數與 $sh10wt$ 及 $w2stwt2$ 相乘，以進一步修正可能的樣本偏誤。

※建議引用本文之參考文獻格式：

關秉寅、詹傑勝 (2015)。**TEPS-B 技術報告第二號：建構 TEPS-B Panel 1 SH**

樣本 2010 年面訪調查完訪成功樣本的機率權數。取自：

http://tepsb.nccu.edu.tw/download/TEPS-B_technical_report_no2.pdf