

教育研究集刊

第六十二輯第一期 2016年3月 頁1-33

國中讀前段班有差嗎？能力分班對 學習成就影響的反事實分析



關秉寅

摘要

本研究以「臺灣教育長期追蹤資料庫」國中樣本兩波資料，分析八、九年級讀前段班對九年級時學習成就的影響。本研究以就讀無能力分班措施學校且自認從無分班經驗者為對照組，分別與學校無能力分班、學校只有部分年級有分班，以及學校各年級均有分班措施等三類型學校內，自認於八及九年級與只有九年級讀前段班者相比較，以評估讀前段班是否影響九年級的學習成就。分析模型包括多層次線性、單層次傾向分數配對分析，以及結合單層次傾向分數配對與多層次交叉分類隨機效應模型等。研究發現，以單層次傾向分數分析或此方法結合多層次交叉分類隨機效應模型的結果觀之，只有就讀於部分年級有分班學校，且於八及九年級讀前段班或只於九年級讀前段班會對正面影響九年級學習成就。

關鍵詞：交叉分類隨機效應、能力分班、傾向分數配對、臺灣教育長期追蹤資料庫

關秉寅，國立政治大學社會學系副教授

電子郵件：soci1005@nccu.edu.tw

投稿日期：2015年08月06日；修改日期：2015年10月23日；採用日期：2016年01月28日

Bulletin of Educational Research
March, 2016, Vol. 62 No. 1 pp. 1-33

Does It Make a Difference in Being Assigned to High-Ability Groups in Junior High? A Counterfactual Analysis of the Effects of Ability Grouping on Students' Academic Achievement in Taiwan

Ping-Yin Kuan

Abstract

The present research assesses the causal effects of being assigned to high-ability groups in junior high according to 9th graders' academic achievement in Taiwan. Using data gathered by the Taiwan Education Panel Survey in 2001 and 2003, the sample is divided into three school types: schools with no ability grouping, schools grouping students only at certain grades, and schools grouping students in all grades. Within each school type, this research focuses on estimating the average treatment effect (ATE) by comparing separately the achievement of 9th graders who have self-identified to high-ability groups in both 8th and 9th grade, or in 9th grade only, to those having no tracking experience and enrolled in schools without ability grouping. The research uses multilevel linear modeling, single-level propensity score matching (PSM), and the

Ping-Yin Kuan, Associate Professor, Department of Sociology & International Program in Asia-Pacific Studies, National Chengchi University

Email: soci1005@nccu.edu.tw

Manuscript received: Aug. 06, 2015; Modified: Oct. 23, 2015; Accepted: Jan. 28, 2016.

combination of single-level PSM and cross-classified random effects model (CCREM) to estimate the ATE of being assigned to high-ability groups. According to the analysis of the model combining PSM and CCREM, the research finds that only students assigned to high-ability groups at schools which group students at certain grades have positive impacts on their 9th grade achievement.

Keywords: cross-classification random effects model, ability grouping, propensity score matching, Taiwan Education Panel Survey

壹、前言

自1968年實施九年國民義務教育（以下稱九年國教）以來，學校依照學生的智能或學習表現予以能力分班的作法大概是最具爭議的教育議題之一。實施九年國教初期，不少國中為提高升學率，依學生能力採階梯式分班，所謂的「好班」、「壞班」、「升學班」、「放牛班」等名稱應運而生，造成一些學生學習及行為上的負面效果，也引起教育資源分配不公及階級歧視等問題，消除能力分班乃成為重要的教育改革議題。

在輿論層面上，能力分班的爭議及關注反映在網路可搜尋到之支持以及反對能力分班的各式各樣報導與網誌中。2015年7月底，如以完全符合「能力分班」四個字為關鍵詞搜尋Google臺灣網頁，可搜尋到約八萬八千多筆的結果。最近一年多來，因為12年國民基本教育（以下稱12年國教）實施的爭議，能力分班又成為熱門話題。如以「能力分班」來搜尋最近一年Google臺灣新聞版，則有一百六十多則相關報導。在法規層面上，教育主管單位曾多次修正編班政策，且逐漸加強要求常態分班（林彥志，2006）。例如：教育部於1979年頒布之《改進國民中學學生編班試行要點》，即明文規定「嚴禁固定階梯式能力分班」。2004年修訂的《國民中學學生編班實施要點》則具體規定國中三年皆須進行「S型排列」常態編班。2011年立法院審查通過了《國民教育法》第12條的修正條文，其中規定「……國民小學及國民中學各年級應實施常態編班，為兼顧學生適性發展之需要，得實施分組學習，其班級及學科能力分組準則，由教育部定之」。此修正條文一方面將常態編班的規定由行政法規提升到法律位階，但另一方面，則又允許能力分組學習，讓學校在對待不同能力學生的學習時留下彈性作法。另外，教育部也制定了《國民小學及國民中學常態編班及分組學習準則》，其中第8條規定：「國中小之分組學習，以班級內實施為原則。但國中二年級、三年級得就下列領域，以二班或三班為一組群，依學生學習特性，實施年級內之分組學習：一、國中二年級得就英語、數學領域，分別實施分組學習。二、國中三年級得就英語、數學、自然與生活科技領域，分別實施分組學習……。」換言之，雖然教育部不允許學校進行各年級以整個班級為基礎的能力編班，但仍准許依領域進行

的能力分組。

以國中同年級的班級為基礎之能力分班的現況言，雖然法規上原則禁止，但人本教育基金會2002年及2003年的調查均顯示，臺灣大多數縣市地區仍有不同比例的國中採取能力分班的作法（人本教育基金會，2002a，2003）。而從「臺灣教育長期追蹤資料庫」（Taiwan Education Panel Survey, TEPS）公布2001年及2003年兩波追蹤之國中學生的班級數來看，接受追蹤學生所屬的班級數目從七年級時的1,244班，增加至九年級時的1,938班，增加了約57%（張苙雲，2011）。至於班級數增加的原因之一，可能就是國中在七年級以後實施能力分班的結果。這些調查均顯示能力分班或分組仍是不少國中學校的作法。

能力分班，事實上，也是不少基礎教育普及之工業化國家常見的措施（Ansalone, 2010; Hanushek & Woßmann, 2006; Huang, 2009; Schofield, 2010）。以美國言，能力分班也是教育界研究最多、爭議熱烈的議題之一（Ansalone, 2010; Loveless, 1999; Oakes, 1985）。這樣的現況，正如Gamoran（2010）所指，能力分班或分流的作法是教育體系內的兩難狀態。這種兩難源自於一方面學校應該提供學生一個共同認知及學習社會技能的架構，以培養學生未來全面參與成人社會所需之公民及經濟活動的基本能力；另一方面，學校又往往必須選擇並分類學生，以便給予符合他們志向及能力的成長軌跡。因此，分班或分流的爭議在於，究竟學校應該是提供所有學生共同的社會化經歷，還是要將學生分化以符合其不同方向的未來發展。如為前者，則學校編班應符合混合不同能力的常態編班；如是後者，則會是能力編組或編班。更重要的是，學校的這類作為可能會連結學生原有在社會階層結構中的位置，例如：學生家庭之階級或族群等，並進而影響整體社會流動或不平等的程度。因此，能力分班的核心爭議是此一作法是否是會擴大原肇因於階級或族群不平等所形成之學習成就上的不平等（Kubitschek & Hallinan, 1996）。

正由於能力分班是社會學及教育學的重要議題，因此，能力分班究竟會帶給學生在學習、身心、行為及未來發展上何種正面或負面的效果，亟需有好的實證研究，以提供教育實務及政策制定的基礎。要探究能力分班對學習成就的效果為何，理想上是要用大規模隨機分派的實驗設計來進行。如具同樣能力及背景條件的學生經隨機分派到前段或後段班後，其後之學習成就上的差異，就可歸因於

6 教育研究集刊 第62輯第1期

能力分班的結果。但這理想上的研究設計，實務上會有諸多爭議而不可行，因而此方面的研究就常須調查或以田野觀察方式來蒐集相關的資料進行分析。但以抽樣調查的資料來觀察實際已經分派到不同能力分班者，就會遭遇到一些會影響分班效果估計的問題。最基本的問題是基準線（baseline）上的差異，亦即進入這些不同能力分班的學生，在未進入能力分班狀態前，其先備學習能力可能就大不相同。此外，進入這些不同能力班級者，可能因為個人先備能力或自我選擇等差異，而使這類班級對不同的個人產生不同的效果。因此，如何利用適當的統計分析來分析調查資料，以逼近實驗設計可觀察到的效果，是十分重要的。此類統計分析最好能夠像隨機分派的實驗研究一樣，這也是本研究企圖達成的目標。

相對於國外學界對此議題長期研究且持續關注，並將研究成果做為學校或政府政策參考的情形言（Gamoran & Berends, 1987; Kulik & Kulik 1982; Kulik, 2004; Oakes, Gamoran, & Page, 1992; Schofield, 2010; Slavin, 1987, 1990a; Van de Werfhorst & Mijns, 2010），此議題在臺灣目前教育學及社會學的基礎研究上略顯遜色。臺灣缺乏這類研究的部分原因是過往並無這方面實驗性的教育研究，也沒有大規模且具代表性的長期追蹤調查。因此，需要如TEPS這類長期追蹤調查資料庫建置完成後，才有可能做比較嚴格的實證檢驗。此外，能力分班對學習是否有效果的議題，實質上可分成兩個相關的研究議題：一為評估能力分班是否有效的議題；另一則是在前一議題的研究發現上，進一步探究能力分班有效或無效的條件及機制為何。由於此二議題從本文以下介紹的反事實因果推論的角度言，如果研究設計不是隨機分派實驗的話，都不容易確認研究結果的有效性，且這兩個議題又以後者較前者更複雜。因此，本研究的目的有二：一為利用TEPS的長期追蹤資料中自認八年級或九年級讀前段班者與從無分班經驗者做比較，先聚焦探究第一個議題，並初步探索影響分班效果的一些學校層次的條件因素；其次則是說明如何利用新的分析方法及其立基之反事實因果推論的方法論，以有效評估國中時期讀前段班對學習成就之影響。

貳、文獻探討

一、能力分班的定義

Kulik (2004) 回顧能力分班效果的相關文獻後，歸類出六種美國教育體系分班或分組的作法，¹並指出一般學者常會交替指稱ability grouping或tracking這些不同方式的分班或分組的作法。不論是哪一種方式的分班或分組，其結果皆是使同一特性或能力的學生在一起學習，形成一種同質性的班級環境。

本研究分析的對象是2001年及2003年參與TEPS調查的公立國中學生，並檢視能力分班對TEPS測量之綜合能力分析分數為學習成就指標的影響。因為國中將學生依照能力分成不同班級上全部課程是不符規定的，且即使學校在合法範圍內進行能力分組的作法不同，也因如此，TEPS只於2003年時由受調查國中自陳何時讀過前段或中後段班。²如學生回答說讀過這類班級的話，則有可能是全部課程與特性相同的同學一起上課，但也有可能是當時法規所允許之能力分組，就某一主要升學考試相關科目，如數學或英文，與同一特性或能力的同學上課的情況。因此，以TEPS的資料言，本研究所稱之能力分班是同時指稱前述的兩種情況。此外，以下相關文獻的探討也同時包括了這些以不同方式分班或分組的情況。

¹ Kulik (2004, p. 159) 提到的依能力分組的六類方式包括：(1) 同一年級的學生依學習表現分成如高、中、低等不同班級，並接受不同課程教導；(2) 大部分時間是在具異質性的班級學習，只有少數課程（如閱讀或數學）是在同質性高的班級學習；(3) 同年級學生依照其在某科目的學習表現分派到不同年級程度的班級上課；(4) 同一班級分成不同能力的群組學習；(5) 資優學生接受較進階的學習使其能跳級學習或提早畢業；(6) 資優學生接受較一般課程更豐富及多元的學習經驗。

² 依照人本教育基金會 (2002a, 2002b) 2002年調查發現，即使七年級時是常態分班，八年級時就英、數兩科分成兩或三組後，並沒有什麼隨學生能力變動而往前移動之類的作法，且八年級暫時性分組，到九年級時就成了永久性分班了。但TEPS於2003年田野調查時，也發現有每次段考後即重新分班的學校。因此，以國中生自己認定來判斷是否就讀過前段或中後段班，雖可能與理想客觀的情況不完全相同，但以學校實際分班或分組的複雜性言，可說是次佳的選擇。

二、能力分班的效果

能力分班效果的爭議中，最常見的對立是贊成能力分班者基本上假定能力分班可因材施教外，也相信此作法會對學習有正面效果；而反對者則認為進入後段班者，對其學習有負面效果，並擴大學習成就的不平等。從國外研究的文獻來看，能力分班是否會造成正面或負面的效果，並未有明確的定論，因此這議題也是能力分班研究者需優先探討的。

對於能力分班效果的研究發現大致可分成三類：

（一）能力分班會增加不平等：這類研究結果的發現，是學習成就會因學生是參與前段班組或是學習進階班組而有正面效果；但在後段班的，則有負面效果。這類比較主要是比較同一學校內不同能力分組者間的學習表現，而能力分班會使學習成就不平等擴大，則通常是歸因於能力分班與學習不同程度的課程、不同學習機會，不同的班級學習氣氛，以及教師的教學經驗、品質及態度等（Carbonaro & Gamoran, 2002; Entwisle, Alexander, & Olson, 1997; Gamoran, 1987, 2004; Kelly, 2004; Lee, Bryk, & Smith, 1993; Loveless, 1999; Lucas, 1999; Oakes, 1985; Van Houtte, 2004）。Gamoran（2010）回顧相關文獻後指出，大多數英、美學界的研究是指向這類結果。例如，Kerckhoff（1986, 1993）根據一項近30年多波長期追蹤一群1958年3月第一個星期出生於英格蘭及威爾斯的人的調查，發現就讀中小學時被分為高能力者，其學習表現比進入混合能力類型學校或班組者為佳；而進入低能力類型學校或班組者，則比較差。更重要的是，進入高、低能力學校或班組者間的差距是有累積性的。依據美國全國性調查資料的幾項研究結果也顯示，學生進入不同能力分組學習後，其學習成就的差距也是逐漸增加（Alexander, Cook, & McDill, 1978; Gamoran, 1987, 1992; Gamoran & Mare, 1989; Heyns, 1974; Hoffer, 1992）。

除了英、美的研究外，其他如日本、韓國、南非、以色列、德國、比利時等國的研究也發現，不論是以能力分派學校或學校內的分班組等，都會擴大學習成就上的差距（Gamoran, 2010）。臺灣學者一項在都會區對能力分班結果的長期追蹤也發現類似的情形（謝小苓、楊挽北，2003）。而黃敏雄（Huang, 2009）以TIMSS 2003的跨國數學表現的資料分析同質性班級的效應時，也發現同質性的班級會擴大能力高、低者的平均數學表現，但與混合能力班級的平均數學表現

比較時則無差別。此外，美國因為能力分班可能擴大學習成就不平等，部分地區開始了消除能力分班（*detracking*）的作為。一些研究結果顯示，消除分班後，在混合能力的班級內，低能力學生的學習表現會增強，但高能力學生的表現則會降低（Argys, Rees, & Brewer, 1996）。這類發現可說是從另一角度支持了能力分班會增加學習成就不平等的證據。

（二）能力分班無明顯的正面或負面效果：雖然大多數實證研究的發現是，能力分班會擴大高、低能力班組的差距，但Kulik（2004）指出這類的研究發現通常是根據調查資料或民族誌的田野觀察。Kulik進一步指出，不同分組學習成就差距，大部分可歸因於學生的自我選擇，而不是能力分班的結果。Slavin（1990b）則認為，如果同校但不同能力組的學生一開始就差異很大，則一般這類資料所做的統計分析，往往無法完全控制這類的差異，進而提高比較蘋果與橘子的可能。此外，他比較有分班及無分班制度之中學間的平均學習成就後，並未發現明顯的差異。

換言之，不論是要比較同學校內能力分班對學習成就的影響，還是跨校比較有或無能力分班制度學校間的學習成就差異時，都需要先設法蒐集到所有影響不同組間或學校間差異的因素，但一般調查並無法達此目標。一般利用這類資料的分析如果採用OLS（ordinary least squares）迴歸分析模型，並無法確實評估能力分班的因果效應。為了克服前述統計分析方法上的問題，有一些研究利用如固定效果模型、工具變項（instrumental variable）或是傾向分數配對（propensity score matching, PSM）等，這些分析方法可相當程度控制不同組間基準線差異及未觀察到變項（unobservable variables）的影響。這類研究的結果顯示，能力分班的效果小於過往以OLS迴歸分析所得到者。例如：Figlio與Page（2002）用了工具變項及二階段最小平方（two-stage least squares, 2SLS）迴歸分析美國國家教育長期追蹤研究（National Education Longitudinal Study）資料後，發現分班並不會對學習成就造成負面影響，且可能因學校的能力分班制度吸引了高社經地位背景學生，反而會對部分後段的學生有正面效果。Betts與Shkolnik（2000）同時以2SLS及PSM的方法分析美國青少年長期追蹤研究（Longitudinal Study of American Youth, LSAY）後，發現能力分班對低能力組沒有影響，對中間能力的有負面影響，對高能力組有正面影響。至於比較不同類型學校時，也可用前述先進統計

分析策略。例如：Galindo-Rueda與Vignoles（2005）用PSM的方法分析了前述Kerckhoff（1986）所使用的英國長期追蹤資料後，發現進入低能力類型的學校對學業成就並沒有影響。此外，Kulik（1992）及Slavin（1990a）評估美國於二十世紀初期在底特律進行的能力分班實驗研究及以後類似實驗的結果後發現，如果能力分班不是配合學生能力給予不同課程，其效果對於高能力者只有很小的正面效果或是沒有效果，而對於中或低能力者言，則是沒有效果（Kulik, 2004）。

（三）能力分班對不同能力者有正面效果：這類研究結果主要由實驗設計或類實驗設計的方法進行能力分班的研究，且配合不同能力給予不同課程而得到。這類的實驗包括了將不同年級但能力相近的小學生組成不同的閱讀班級，或是班級內的能力分組（Kulik, 1992; Slavin, 1987, 1990a）。此外，如果提供特殊資優學生特別的課程，則其效果會比讓這些學生在一般的班級學習要來得好（Kulik, 1992, 2004）。Gamoran（2010）在檢視不同國家能力分班之效果時也發現，如果能力分班是與課程或清楚具體的評估標準相結合，則不論是能力好的班級，或是能力比較差的班級，都能有正面的結果。又如 Brooded（1997）發現，臺灣因有聯考的制度，因此，不論是高能力或中後段能力的班級，都必須努力學習，故導致學習成就的不平等縮小。同樣地，Ayalon與Gamoran（2000）在以色列的研究也發現，如果學校將準備升大學的數學班級區分為不同能力上課，則不同能力組間受家庭背景影響之學習成就差異，會比沒有將學生能力分組的學校來得小。在跨校比較不同類型學校方面，南韓因教育制度改革而使一半高中執行平等化政策並實施隨機分班，另一半則仍實施能力分班。一項研究分析此二類學校學習成就的結果顯示，能力分班學校的平均表現較佳，且以分量迴歸（quantile regression）分析後，發現能力分班對高於中位數能力者有正面效果，對低於中位數者則無負面影響（Kim, Lee, & Lee, 2008）。

三、小結

從前述回顧能力分班對學習效果的文獻來看，能力分班的效果未有定論的原因至少有兩項：

（一）能力分班的效果會受到社會、學校及班級等脈絡的影響：就社會層次言，臺灣或南韓這類社會的全國性標準化入學測驗制度，會影響教學過程與目

標，進而使其能力分班的效果與其他沒有這類制度的社會有差異。就學校或班級層次言，如果不同能力班級的教學過程符合學生的需要，且因材施教，則能力分班不一定會帶來負面的效果，而且學校執行能力分班或分組的同質性程度、進入分班或分組的自願性、換組彈性及實施範圍等，也都會影響能力分班的效果（Gamoran, 1992, 1993）。

（二）能力分班效果會受到研究設計的影響：能力分班效果的觀察往往是用調查或民族誌的方法研究，而不是隨機分派的實驗設計。以前者方法蒐集的資料分析時，如果分析時無法適當控制其他因素的影響，則得到的能力分班效果可能會有偏誤，或產生不一致的觀察。例如：若研究者直接比較實施能力分班學校中的前段班及後段班學生，就有可能無法有效控制兩組學生間一些重要未觀察到的差異。比較好的比較，應該是觀察具有同樣啟始條件的兩組（Betts & Shkolnik, 2000; Mulkey, Catsambis, Steelman, & Hanes-Romos, 2009）。當然，大規模實驗設計往往不可得，而如果要以調查資料分析能力分班的效果，則此調查資料最好是有代表性，能提供個人、班級或學校等多層次資料，且具有類實驗設計特性的長期追蹤資料。此類資料除可有效控制各層次基準線上的差異外，尚可從事嚴謹的因果推論。

為了處理上述兩種造成能力分班效果研究上的問題，本研究即利用TEPS此一具代表性之長期追蹤資料，並以可適當處理未觀察到變項，以及能考量學校及個人等層次干擾因素的幾種分析方法，來分析國中時期能力分班的效果。以下就本研究所採取之分析策略及資料等，做進一步的討論。

參、研究方法

一、以觀察性研究進行能力分班效果之分析策略

如前所述，要能有效推估能力分班對學習成就的效果，理想上是用隨機分派的實驗設計。如果透過調查這類稱為觀察性研究（observational studies）資料推估能力分班之效果時，則會遇到進入及未進入能力分班者在基準線即有的系統性差異，以及分班效果會受到未能觀察到之重要變項的干擾，進而造成推估偏誤的

問題。一般線性迴歸（ordinary least squares regression）分析假定研究者可控制足夠且適當觀察到的變項來處理這些問題。但如違反假定的話，能力分班效果的推估就會有偏誤。在推估能力分班效果時，違反假定的可能性頗高。除了進入及未進入能力分班者可能在各種可觀察到之初始條件，如學習態度、表現、家庭社經地位等有所不同外，在個人及學校層次都可能有無法蒐集到且影響能力分班效果推估之變項。以個人層次言，如智力即此種重要因素。在學校層次，一個學校為何會有能力分班或與學校的大小，以及學生學習表現、家庭社經地位及族群等的異質性程度等有關（Lucas & Beresford, 2010; VanderHart, 2006）。除了這些可能比較容易蒐集到的學校層次變項外，還是可能有其他學校層次且會影響能力分班，以及分班效果的變項無法被蒐集到，例如：學校在地區的競爭力、學校能力分班的篩選標準及實施範圍等（Gamoran, 1993）。換言之，學校是否採能力分班，以及如何能力分班等的作法並不是一種隨機分派的過程，而是一種學校層次自我選擇的過程，但此過程並無法有效地反映在調查中的資料。

為處理前述未觀察到的變項對所欲推估之因果效應的影響，研究者通常可以採取幾個策略。其一為如有同一對象之長期追蹤資料，研究者就可以運用固定效果模型（fixed-effects model，簡稱FE模型）或差異中之差異模型（difference-in-difference model）來從事分析。此種模型之分析策略是如同將個案本身當成控制變項，如此則能有效處理未能蒐集到與個案有關且可能影響解釋變項因果效應的個別因素（Allison, 2009）。但是，以FE模型從事因果推論時也有其限制。例如：若此未觀察到變項與解釋變項無關，則FE模型估計值的標準誤會比隨機效果模型（random effects model，簡稱RE模型）較大。³此外，傳統的FE模型對因果效應的估計是假定沒接受處遇（treatment）者如果接受處遇（如進入前段班）的話，則其結果會與接受處遇者相同（這也是OLS迴歸的假定）。但如果接

³ RE模型與FE模型都可以 $y_{it} = \mu_i + \beta x_{it} + \gamma z_i + \alpha_i + \varepsilon_{it}$ 表示。其中， x_{it} 是解釋變項，其數值不僅隨個案有所不同且隨時間而變動，如能力分班的狀態。 z_i 則是隨個案不同但不隨時間變動的解釋變項，如個案之性別。 α_i 及 ε_{it} 都可看成是無法直接觀察到誤差項，不同的是 α_i 是個案間不同但並不隨時間變動的，而 ε_{it} 則是個案在不同時間點會變動的。RE模型與FE模型不同之處在於前者假定 α_i 與模型中之解釋變項是無關的，而後者則允許有此相關（Allison, 2009; Wooldridge, 2010）。

受與不接受處遇者一開始就有顯著的差異，則此假定不一定成立（Abadie, 2005; Heckman, Ichimura, & Todd, 1997）。處理此類差異的策略之一即為利用傾向分數配對（PSM）的方法來找出接受與未接受處遇者間彼此可比較的個案。

以上說明顯示研究者使用觀察性研究資料進行因果推論時，即使是比較先進的統計分析方法本身，也還是可能無法完全有效處理如未觀察到變項等這類因素之干擾。在此情況下，研究者或可結合不同的分析策略以減少這類的干擾。本研究所用的分析策略之一，即為結合PSM及多層次交叉分類隨機效應模型（cross-classified random effect model, CCREM）。

PSM分析是假定可透過觀察到的配對變項來處理接受與未接受處遇者之間的選擇偏誤（selection bias）。此種配對分析法是在反事實（counterfactual）推論分析的架構下發展出來的，反事實推論分析認為每一個體都存有兩種反應，一種是接受實驗處遇後的反應，另一種則是未接受實驗處遇的反應。以能力分班為例，進入前段班是一種實驗處遇，因此每個學生都有可能進入或不進入前段班，也因此都有這兩種情況下的反應。現實生活中，因每個學生不可能同時是在能力分班及非能力分班的狀態，因此只有其中一種反應可以被觀察到，另一無法觀察到的，則為潛在可能的反應（potential outcome）。換言之，從反事實推論分析架構的角度來看，真正的因果效應是每個人（不論實際上是否接受了處遇）這兩種反應的差異。自然，此個人層次之差異在現實生活中是無法得到的，能做到的就是比較特性上相同或相近，但實際上接受或未接受處遇等兩組人的反應。因此，從反事實推論分析的角度來看，實驗設計是透過隨機分派特性相同或相近者後，建立實驗組和控制組，並以群組的平均處遇結果來推估前述兩種反應的差異。

在觀察研究中，研究者往往無法做隨機分派的設計，所以會造成實驗組（接受處遇組）和控制組（未接受處遇組）有基準線的差異，以及因自我選擇而有因果效應上的差異。這些差異會影響到因果效應的推估。為了克服觀察研究在因果推論上的困難，Rosenbaum與Rubin（1983）提出了傾向分數配對法來選取實驗組和控制組中在眾多配對變項（confounding covariates）上的可比較樣本。傾向分數的估計通常就是用邏輯迴歸（logistic regression）或是probit迴歸來估算每個個案接受處遇的機率。利用傾向分數配對法的優點，是研究者只需要用一個傾向分數進行配對，以減少眾多配對數在配對上的困難（Guo & Fraser, 2015; Morgan

& Winship, 2015)。

以本研究言，以PSM推估能力分班效果的作法之一，是指將進入某類型能力分班者的個人、家庭及學校等特性均視為是個人層次變項，然後與從無分班經驗者進行配對，然後再計算兩者間在學習成就上的差異。此作法可稱之為單層次PSM (Guo & Fraser, 2015)。符合使用PSM分析之假定的前提下，PSM原則上可以分別估算出實際上分派到某類型能力分班者，如果不是在這類班級的平均處遇效果 (average treatment effects on the treated, ATT)，以及實際上無能力分班經驗者，如果是進入了某類型能力分班的平均處遇效果 (average treatment effects on the untreated, ATU)。ATT及ATU依照有無能力分班樣本數比例加權並加總後，即可得到ATE (average treatment effects)，亦即如所有的人都參與能力分班的平均效果。與OLS迴歸分析不同的是，為求接受及未接受處遇者間在配對變項上無顯著差異或達到平衡 (balanced)，此方法在分析時可能會排除一些無法配對的樣本於分析之外，以達成合理比較的目的。此外，OLS迴歸基本上假定ATE與ATT及ATU是相同的，而PSM則無此假定。⁴

二、本研究之分析策略及樣本選擇

本研究對國中時期能力分班效果的分析，使用的資料是TEPS在2001年及2003年調查同一群公立國中學生樣本的限制使用版 (授權碼為TBN00017)。TEPS 2003年實際完訪國中學校數為333所，班級數為1,938班，學生數則約19,000人 (張苙雲, 2011)。本研究只分析就讀公立國中並且未轉學之學生，原因是就讀私立國中者已是經過自我選擇後之結果。因此，初步分析的對象是15,019人。

此外，本研究將進一步針對無能力分班措施之公立國中中從未有分班經驗

⁴ 如果所有人都參與前段班的潛在結果以 Y^1 表示，都不參與的潛在結果以 Y^0 表示，且實際參加是以 $D = 1$ 及實際沒參加以 $D = 0$ 表示，則 $ATE = E(Y^1 - Y^0) = [E(Y^1|D = 1) - E(Y^0|D = 0)] - [E(Y^0|D = 1) - E(Y^0|D = 0)] - (ATT - ATU)q$ ，其中 $E(Y^1|D = 1) - E(Y^0|D = 0)$ 是實際觀察到參與者的結果與實際沒參與者的結果之間的差異， $E(Y^0|D = 1) - E(Y^0|D = 0)$ 是實際參與者如果沒參與的結果與實際沒參與者之結果的差異，而 $ATT - ATU$ 則是實際參與者與實際沒參與者在平均處遇效果上的差異， q 是沒參與者在母群體的比例 (Guo & Fraser, 2015, pp. 53-54)。

者，與八年級及九年級自認都就讀前段班或好班者，⁵以及只有九年級讀前段班者做比較，以瞭解讀前段班是否對於九年級的學習成就之ATE及ATT，也就是如果所有學生都讀前段班，以及實際就讀前段班者（如果不是讀前段班的話）之學習成就差異的估計。此分析策略除符合先前文獻探討如何方能有效評估能力分班效果，以及一般認為讀前段班有效等原因外，也與TEPS於2003年調查能力分班的題目設計有關。TEPS此波調查時詢問追蹤調查之國三生（即九年級生）：「從國一到現在，有沒有念過前段班或好班？」。可複選的答項包括：「學校沒有分前後段班」、「國一有」、「國二有」、「國三有」、「國中三年級都在中段或後段班」、「不清楚」。為求較嚴謹之因果推論，本研究只選擇公立國中七年級（即國一）時均未參與能力分班者，然後分析他們八年級或九年級時是否就讀前段班後的學習成就差異。排除自認七年級即進入分班者的原因是，TEPS資料無法有效知道學生七年級以前的學習表現。此外，本研究也排除了自認只有八年級、七及八年級或七及九年級讀前段班，以及國中時期一向讀中、後段班的學生。此研究設計，一方面可就TEPS資料充分掌握七年級時尚未進入分班者之各種個人、家庭等學校背景情況，另一方面則可掌握在八或九年級時持續就讀前段班的變動狀態。以下分析時，也將進一步區分為只有九年級就讀前段班者，以及八及九年級均就讀前段班者兩個類型。

由於學校是否實施能力分班或分組，以及分班或分組的實際作法不同，因此分析時，進一步依照TEPS於2001年由學校行政人員填答之「學校問卷」中蒐集到的學校實施資優班及能力分班的情況，將學校分成三類：無資優班及能力分班學校、部分年級有資優或能力分班學校，以及各年級都有資優或能力分班學校。進一步比較分析時，將排除後二類學校中學生從無分班經驗者，因為這些學生雖無分班經驗，但客觀上，可能已經被學校篩選過了，但卻不知道。換言之，從實驗設計的角度來看，本研究的控制組是無資優或分班措施之學校中從無分班經驗者。因為這些學生不論是主觀或客觀條件上，都沒有受到分班措施的影響。對本研究言，此種將學校分類及樣本選擇的設計，至少有兩個好處：

（一）可大致掌握學校能力分班之組織特性：雖然TEPS並未蒐集學校層次

⁵ 以下將以「前段班」之名稱涵蓋前段班及好班。

與能力分班或分組相關的詳細資訊，無法如Gamoran（1993）之研究般，就學校能力分班或分組的組織特性做較細緻的分類，並評估其效果。但掌握學校是否分班，以及是否各年級都分班，已可大致掌握學校分班實施影響的範圍，並可在分析時做控制或配對時比較。

（二）可分析自認讀前段班者對學習成就影響的基準：由於學校問卷為學校行政人員自陳是否學校有無資優班或能力分班的訊息，但此訊息與學生自認的學校是否有能力分班的經驗有不一致之處。本研究經過比對學校回答無資優班或分班措施的樣本中七年級及九年級同班同學之比例，並經篩選無變動情況者後發現，即使是這樣本中，仍有154人自認在九年級時讀前段班，46人自認是八年級及九年級都讀前段班（如表1）。此類型學校這兩類自認讀前段班者，正好可以做為這兩類自認讀前段班者對學習成就影響的基準。原則上，如果學校行政人員的回答沒有問題，應可預期控制學生各類背景變項後，這類無分班措施學校中這兩類自認讀前段班者，應與從無分班經驗者無學習成就的差異。如果有差異的話，則或可視為是所有這兩類自認讀前段班者之主觀知覺對九年級學習成就的影響。因此，如果在其他兩類有分班措施學校中，這兩類自認讀前段班學生的學習成就與從無分班經驗者的學習成就有顯著差異的話，就可看做是學生主觀知覺及客觀條件綜合的結果，而可以用以評估分班的客觀影響大致為何。

經上述樣本篩選後，本研究用以分析只有九年級，以及八年級及九年級讀前段班對九年級學習成就之ATE及ATT的樣本數是3,819人（如表1）。

三、本研究之統計分析模型

在統計分析模型方面，本研究是以多層次線性模型（multilevel linear models, MLM）做為其他分析模型的比較基礎。更具體言，本研究將採用兩層次MLM的隨機截距模型（random intercept models），此模型如以maximum likelihood方法估計，則等同於先前提及的RE模型（Allison, 2009）。此處的兩層次分別是學生個人層次及學校層次。MLM可視之為OLS的延伸，基本上也是假定ATE等同於ATT及ATU。因此，本研究將以MLM分析初始樣本及進一步分析做為控制組之從未有能力分班經驗者，分別與三類型學校中兩類有就讀前段班經驗者的兩兩捉對比較的樣本。以MLM分析初始樣本15,019人可視為未考慮學校及個人

表1
國中時期學校能力分班類型、讀前段班類型中個人及學校層次變項的描述統計

變項	學校無能力分班						學校部分年級分班						學校各年級都有分班					
	沒分班		九年級		八及九年級		九年級		八及九年級		九年級		九年級		八及九年級			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD		
九年級學習成就	59.391	11.970	62.135	10.777	64.428	8.897	60.932	11.494	64.722	7.752	58.908	11.691	64.447	8.341				
七年級學習成就	52.238	9.618	53.262	8.242	55.442	7.918	52.657	9.023	55.264	6.537	51.449	9.696	55.881	6.743				
男生	.502	.500	.506	.502	.435	.501	.500	.500	.429	.496	.546	.499	.459	.501				
族群																		
閩南	.734	.442	.792	.407	.783	.417	.716	.451	.758	.429	.762	.427	.847	.362				
客家	.109	.312	.130	.337	.196	.401	.176	.381	.114	.319	.145	.353	.041	.199				
大陸各省市	.157	.364	.078	.269	.022	.147	.108	.311	.128	.335	.093	.290	.112	.317				
父母最高教育程度																		
國中或以下	.230	.421	.325	.470	.283	.455	.248	.432	.215	.411	.198	.400	.184	.389				
高中職	.353	.478	.448	.499	.478	.505	.438	.496	.457	.499	.498	.501	.408	.494				
專科	.189	.391	.130	.337	.130	.341	.173	.379	.176	.382	.176	.382	.204	.405				
大學	.136	.343	.084	.279	.065	.250	.106	.308	.107	.310	.101	.302	.163	.372				
研究所	.078	.267	.006	.081	.043	.206	.019	.138	.024	.154	.009	.094	.041	.199				
其他	.245	.430	.019	.472	.283	.455	.263	.441	.235	.425	.216	.412	.184	.389				
家庭每月收入																		
不到2萬	.063	.243	.130	.337	.087	.285	.101	.302	.118	.323	.141	.349	.071	.259				
2萬至5萬以下	.354	.478	.422	.496	.391	.493	.418	.494	.384	.487	.414	.494	.337	.475				
5萬至10萬以下	.373	.484	.364	.483	.391	.493	.388	.487	.391	.489	.339	.474	.429	.497				
15萬至20萬以下	.132	.338	.065	.247	.087	.285	.065	.247	.093	.292	.079	.271	.143	.352				

(續下頁)

表1 國中時期學校能力分班類型、讀前段班類型中個人及學校層次變項的描述統計

變項	沒分班		學校無能力分班				學校部分年級分班				學校各年級都有分班			
			九年級		八及九年級		九年級		八及九年級		九年級		八及九年級	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
其他	.071	.256	.019	.139	.043	.206	.022	.148	.014	.117	.026	.161	.020	.142
完整家庭	.843	.364	.805	.397	.957	.206	.850	.357	.841	.366	.806	.396	.847	.362
自己教育期望														
國高中畢業	.130	.337	.104	.306	.022	.147	.097	.296	.048	.215	.128	.335	.061	.241
專科	.165	.372	.227	.420	.217	.417	.218	.413	.201	.401	.207	.406	.194	.397
大學	.235	.424	.208	.407	.413	.498	.231	.421	.291	.455	.216	.412	.276	.449
研究所	.272	.445	.266	.443	.239	.431	.295	.456	.315	.465	.291	.455	.286	.454
其他	.197	.398	.195	.397	.109	.315	.159	.366	.145	.353	.159	.366	.184	.389
7年級每週補習時數														
無	.243	.429	.208	.407	.130	.341	.192	.394	.142	.350	.185	.389	.163	.372
4小時以下	.255	.436	.273	.447	.130	.341	.243	.429	.215	.411	.238	.427	.235	.426
4至8小時以下	.268	.443	.253	.436	.391	.493	.286	.452	.339	.474	.260	.440	.337	.475
8至12小時以下	.142	.349	.169	.376	.239	.431	.174	.380	.183	.388	.225	.418	.143	.352
12小時以上	.092	.289	.097	.297	.109	.315	.105	.307	.121	.327	.093	.290	.122	.2904
七年級學習成就學校平均數	51.991	4.177	49.539	2.977	48.522	3.139	48.827	3.274	48.545	3.442	49.199	2.659	49.651	.188
七年級學習成就學校標準差	.862	.188	.899	.154	.833	.191	.907	.196	.915	.174	.868	.173	.904	
樣本數 ¹	2025	154	46	980	289	227	98							

註：各學校類型內各分班類型之樣本數以變項中均刪除遺漏值後樣本數最小者為基準。

層次中可能有未觀察到變項影響學校是否採取能力分班措施，以及學生是否進入前段班等較簡單的分析模型。

本研究也以單層次PSM方法與單層次PSM結合CCREM模型等兩種模型來推估就讀前段班對九年級學習成就的ATE及ATT。當資料為多層次時，估計傾向分數至少有四種方法。這四種方法主要的不同處在於估計傾向分數時，是在屬高層次之同一群組（如學校）內做配對估計，還是跨群組配對估計，以及估計時是否用到MLM（Guo & Fraser, 2015; Thoemmes & West, 2011）。由於本研究的幾種配對比較樣本並非是同一學校中有無能力分班經驗者的比較，因此，本研究為跨群組比較，且此跨組比較也排除了使用MLM來從事多層次邏輯迴歸來估計傾向分數的作法，故本研究採用單層次PSM來估計傾向分數並進行配對，亦即所有配對變項為單一層次資料的估計。但以單層次PSM配對成功後的樣本，即形成配對的群組，因此本研究將用Guo與Fraser（2015）提出的第五種方法，亦即結合單層次PSM及CCREM的模型。此模型一方面使用單層次PSM配對成功的樣本，另一方面，則進一步利用MLM，並視分析樣本的個案為同時交叉隸屬學校及配對組等兩個層次。⁶此作法可彌補單層次PSM無法有效利用資料為多層次的特性，且又利用了PSM的優點。本研究之PSM模型均為一對一配對，但允許配對時，控制組（即無能力分班經驗者）中與讀前段班者有相同傾向分數者可重複被用來配對，且配對時使用四分之一個傾向分數標準差為caliper的設定，以確保配對成功者為傾向分數相近者，並減少估計誤差（Guo & Fraser, 2015）。⁷

四、資料與變項

根據TEPS提供學生個人及學校層次的資料，本研究使用之變項（如表1）可

⁶ 在教育長期追蹤研究中，CCREM可用來瞭解學生從小學到中學的多層次分析。因此，同一學生在分析時是同時交叉隸屬小學及中學兩個群組（Rabe-Hesketh & Skrondal, 2008, pp. 473-501）。

⁷ 本研究所有的統計分析均使用Stata 13。MLM分析使用mixed指令，PSM則使用psmatch2（Leuven & Sianesi, 2012）。除了設定的caliper是傾向分數的四分之一個標準差是將配對的傾向分數限定在一定距離範圍內外，本研究也有common support的設定。本研究樂意提供所有使用Stata之指令。

分成以下四大類：

(一) 三類型學校內與讀前段班相關的變項：本研究初步分析之15,019人中，自認學校沒有分班的有8,834人（58.8%）、只有七年級讀前段班的為522人（3.48%）、只有八年級讀前段班的為541人（3.60%）、只有九年級讀前段班的為1,762人（11.70%）。至於七及八年級、七及九年級，以及八及九年級讀前段班的，則分別為162人（1.08%）、33人（.22%）及529人（3.52%）。七年級至九年級年級都讀前段班的有935人（6.23%），而七年級至九年級都讀中、後段班的，則有1,701人（11.33%）。本研究僅關注七年級時未能力分班，而八年級或九年級時才讀前段班情況的樣本。因此，本研究關注的對象包含了樣本數最大的組別，即無分班經驗者，以及各類讀前段班中最大組別，即只有九年級時讀前段班者。因此，不少國中可能為了高中升學考前的衝刺，採取九年級時才分前段班的作法。⁸

(二) 學習成就之指標：做為能力分班效果的主要結果變項是TEPS測量九年級綜合能力分析之IRT分數，此分數是反映學生學習成就及成長的指標（楊孟麗、譚康榮、黃敏雄，2003）。無論是七年級或九年級測得的IRT為可比較之標準化分數，但不易為一般人理解，故本研究先將TEPS全部七年級時測量到之綜合分析能力之IRT分數轉成平均數為50、標準差為10的分數，然後再以此為基礎轉換全部九年級綜合能力的IRT分數。以本研究分析之初始樣本15,091人，七年級平均學習成就為49,932，標準差為9.754，而九年級的平均學習成就為56,600，標準差為11.971。因此，平均來說，九年級的分數增加了約7分，但整體離散的程度也增加了些。

(三) 個人層次的共變項或配對變項：這些共變項均根據TEPS於學生七年級時之學生及家長資料來建構，包括做為學生先備能力指標之七年級上學期時測

⁸ 另以初始分析樣本（ $n = 15,019$ ）進一步以學校類型及學生分班經驗的交叉分析來看，無分班措施學校中認為學校沒有分班的比例的確最高（88.7%）。部分年級由分班之學校為42.5%，各年級都有分班之學校為55.7%。此顯示學校行政人員的回答有一定信效度。此外，部分年級有分班之學校以九年級讀前段班的比例為最高（17.6%），其次為各年級有分班之學校（9.2%），再次為無分班之學校（6.7%）。各年級均有分班學校中自認國中三年都讀前段班的比例是9.1%，而部分年級有分班學校則是7.0%。

得的綜合能力分析分數，學生人口特性、家庭背景及資源，以及與興趣和努力等有關的變項。分析模型中加入七年級時的綜合分析能力分數，可視為一種設法控制學生個人層次未觀察到變項的策略。代表學生之人口特性、家庭背景及資源的變項包括了性別、族群、家長教育程度（父或母中教育程度最高者為主）、家庭每月收入、家庭結構等。學生自己興趣及努力有關的變項則為自己的教育期望，以及每週上補習班的時數。

（四）做為學校層次的共變項或配對變項：TEPS限制使用版資料提供各國中之代碼。本研究利用此代碼建構了學生七年級時各校學生學習成就的平均數及標準差等兩項學校層次變項。此二變項應可有效反映各校學生之能力水平及分化差異的程度。

肆、研究發現

一、描述性統計之觀察

依據表1所呈現三類型學校內分班的描述統計顯示，做為控制組之從沒有分班經驗者九年級時的平均學習成就是59.391、標準差是11.970，此分配比分析初始樣本的平均數為高，但標準差則相近。九年級學習成就平均數最高的，不論學校是否有分班措施，都是八及九年級都讀前段班者，其平均數都在64.5上下。只有九年級時讀前段班者的平均數，是出乎意料之外的，是以學校無分班措施的最高，62.135。另由表1可看到，不論學校類型，自認八及九年級都讀前段班者九年級時學習成就的標準差，比從沒分班經驗者及只有九年級時讀前段班者都小約2至3分左右。其中，又以部分年級分班學校內八及九年級都讀前段班的標準差7.752為最小。不論學校類型，只有九年級讀前段班者的標準差大致相近，且也與無分班經驗者相近。因此，從九年級學習成就的描述統計初步看來，讀前段班的學習成就確實較高，且以八年級起連續兩個年級都讀前段班，會有更高的學習成就，彼此的同質性也較高。但這是就國中時期最後九年級的學習成就來比較，並未考慮啟始基準點不同的問題。因此，需要進一步觀察比較的是學習成就的成長情況。

表1也清楚顯示，八及九年級都讀前段班者在七年級時的平均學習成就，不論學校類型，相對於其他組別都是最高的，差距約3分左右。其中，以部分年級分班學校中，八及九年級都讀前段班者的平均數55.264為最低，也因此，此組是國中三年學習成就成長最大的。與九年級的平均學習成就比，此組成長了約9.5分，與全部樣本學習成就平均成長7分來比，多了2.5分，也約是四分之三個同樣樣本之九年級學習成就的標準差。但是，這突出的學習成長，與其他兩類型學校同樣自認兩個年級都讀前段班者相比，就不顯得特別了。因為無分班措施學校中，八及九年級都讀前段班的學習成就成長約9分，各年級都有分班學校中同類別的學生也成長了約8.5分。不論學校類型，自認只有九年級讀前段班的學習成就成長，雖大致言比兩個年級都讀前段班者差些，但差距也不多。此類型讀前段班者，成長最多接近9分的是無分班措施學校中只有九年級讀前段班者，其次是屬部分年級分班學校者，約8分，再來就是平均成長約7.5分屬各年級都有分班學校者。後面兩組別七年級的平均學習成就與無分班措施學校中無分班經驗者相近，甚至略低些，而此無分班經驗者的學習成就成長也是最小的，約7分。由表1呈現做初步觀察，似可看出讀前段班是能增加一些學習成就，但不算大，且最有效果的學校是屬於部分年級有能力分班的學校。但這初步觀察還需要進一步地控制其他會影響這表面觀察結果的變項。

二、讀前段班之效果

以七年級學習成就言，就讀前段班者大多高過無分班經驗者。除了此差異外，表1也顯示兩者在一些個人層次及學校層次的部分特性上有可觀察到的差異。因此，如果要分析讀前段班是否會影響學生九年級的學習表現，需要設法控制這些基準線差異的影響，以避免讀前段班對學習成就效果的評估產生偏誤。表2及表3即分別呈現進一步以三種統計分析模型控制各層次變項後，對各類型學校內讀前段班類型對九年級學習成就之ATE及ATT的估計結果。

整體觀之，表2明顯呈現MLM與兩個與PSM有關的分析結果相當不同。MLM的分析結果顯示，不論是只有九年級讀前段班，還是八及九兩年級都讀前段班，對於學習成就大多有達統計顯著的正面影響。此外，MLM分析如果是用全部初始樣本分析得到的ATE（即如所有學生都讀前段班的平均處理效果）都略

表2

以不同分析模型估計國中時期兩種就讀前段班類型學生之國三學習成就的ATE

分析模型	學校類型	讀前段班年級	ATE	標準誤 ¹	納入分析樣本數 ²	讀前段班人數
MLM	全部學校	9年級	2.714***	.213	15,019	1,783
		8及9年級	3.939***	.299		534
MLM	沒分班	9年級	2.236***	.587	2,179	154
		8及9年級	2.207*	1.101		2,071
	部分年級分班	9年級	1.620***	.411	3,005	980
		8及9年級	3.189***	.481		2,314
	各年級都有分班	9年級	.151	.839	2,252	227
		8及9年級	2.023**	.762		2,123
單層次PSM	沒分班	9年級	-.292	1.384	2,013	154
		8及9年級	-2.166	3.803		2,012
	部分年級分班	9年級	-.292	1.642	2,945	975
		8及9年級	.901	1.077		2,269
	各年級都有分班	9年級	-1.878	1.330	2,945	975
		8及9年級	.602	1.288		1,721
PSM及CCREM	沒分班	9年級	1.580	.828	294	148
		8及9年級	2.053	1.323		79
	部分年級分班	9年級	1.549***	.462	1,134	579
		8及9年級	3.265***	.711		445
	各年級都有分班	9年級	-.600	.910	404	209
		8及9年級	.672	.907		181

註：1. MLM模型之標準誤為robust standard error，PSM之標準誤則以bootstrapping方法估計。2. 以MLM模型分析全部學校樣本時，樣本數包括所有分析樣本；分析各學校分班類型時，則為組合完全沒分班經驗及各讀前段班類型之樣本。PSM之樣本數則為納入分析，且在common support範圍內之樣本數。

*** $p \leq .001$. ** $p \leq .01$. * $p \leq .05$.

大於用捉對比較樣本分析的結果。如前所述，用全部樣本分析的估計偏誤可能比較大，因為兩種讀前段班者是與所有其他人做比較，其估計可能受到未觀察到之變項影響，會比較嚴格限定在與完全沒有分班經驗者的比較來得大。這可從用捉對樣本比較的話，則估計得到的ATE較小看出。以MLM分析捉對比較樣本的結

果顯示，八及九年級都讀前段班對學習成就的影響不一定比只有九年級讀前段班來得大，且隨學校類型會有變化。在無分班措施學校類型內，只有九年級讀前段班的ATE（2.236）是三種學校類型內只有九年級讀前段班之效果最大者，且與同學校類型內八及九兩個年級的效果差不多。兩類型讀前段班差距最大的是在全部年級都分班的學校類型內。此類型學校內只有九年級讀前段班對學習成就是無顯著影響的，而八及九年級都讀前段班能增加約2分左右。但八及九年級讀前段班對學習成就影響最大的是在部分年級有分班之學校類型內，約可增加3分左右。此學校類型內，只有九年級讀前段班可增加約1.6分。MLM分析兩兩捉對比較樣本的結果，因可能還是無法有效控制無分班經驗者與讀前段班者間基準線差異的問題，仍可能拿蘋果與橘子比較的問題。因此，其分析結果可能還是有偏誤。這也是進一步用PSM來分析比較的原因。

依據表2比較MLM及PSM相關模型之分析結果時，需要注意納入PSM的樣本數與MLM可能不同。如單層次PSM各類捉對比較樣本數都比相對應之MLM分析的樣本小些，納入PSM結合CCREM分析的樣本數又更小了。以部分年級分班之學校類型為例，單層次PSM分析只有九年級就讀前段班的樣本數為975，比相對應之MLM少了五人，而比相對應之PSM結合CCREM分析的樣本（ $n = 579$ ），則多了將近百人。這是因為PSM視配對方法（如一對一配對方法）會排除不在設定配對方法、設定之配對距離範圍（如caliper之設定）及common support範圍內的樣本。因此，PSM的優點是設法比較可比較的樣本，但其限制則是只能推論到配對成功的樣本。

在配對成功樣本的基礎上，表2顯示兩種PSM相關模型的分析結果也有不同之處。以單層次PSM分析的結果是，不論學校類型，兩種讀前段班類型都與無分班經驗者的九年級學習成就無顯著差異。更進一步將PSM配對成功樣本結合CCREM的分析後，則部分年級有分班的學校類型內的兩種讀前段班類型對學習成就言，都有顯著的正面效果，但其他兩學校類型讀前段班者則都無顯著效果。在部分年級有分班的學校類型內，八及九年級讀前段班可增加約3.265分，大約是只有九年級讀前段班的一倍（1.549分）。這個讀前段班對學習成就的影響會

⁹ 本研究各項PSM經配對後配對變項在配對前如有顯著差異的話，配對後均為無顯著差異。

隨學校類型而異的結果，同樣也在以單層次PSM估計讀前段班之ATT（即就讀前段班者的平均處理效果）的分析中看到。表3顯示，在部分年級有分班學校類型內於八及九年級讀前段班者的ATT約為3分，而只在九年級讀前段班的ATT也很接近，約2.9分左右。¹⁰在其他兩類型學校內自認讀前段班者的ATT則都沒有達到顯著。換言之，此處的發現是，讀前段班的正面效果，不論是對全部學生而言，還是只對有讀前段班經驗者而言，都是要在只有部分年級有分班的學校類型內才會達到顯著。

表3

以不同分析模型估計國中時期兩種就讀前段班類型學生之國三學習成就的ATT

分析模型	學校類型	讀前段班年級	ATT	標準誤 ¹	納入分析樣本數 ²	讀前段班人數
單層次 PSM	沒分班	九年級	1.462	1.313	2,013	154
		八及九年級	3.030	2.720	2,012	43
	部分年級分班	九年級	2.930***	.683	2,945	975
		八及九年級	3.023***	.856	2,269	286
	各年級都有分班	九年級	-.876	1.132	2,945	975

註：1. PSM之標準誤以bootstrapping方法估計。2. PSM之樣本數為納入分析，且在common support範圍內之樣本數。

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

伍、結論與討論

國中學校進行能力分班的議題是多年來臺灣教育改革最具爭議性的議題之一，雖然教育主管當局對整班式的能力分班作法採取嚴禁的立場，但在升學考試

¹⁰ 在部分年級分班學校類型中，表2單層次PSM對只在九年級ATE的估計（1.549）與表3單層次PSM對同群人的ATT估計（2.930）有頗明顯的差異，此顯示實際只有九年級讀前段班者的正面效果（ATT）大於無能力分班經驗者如果同樣只有九年級讀前段班者（ATU）。但在八及九年級都讀前段班的比較上，則較不明顯。也就是說，無能力分班者如果是八及九年級都讀前段班，其正面效果與實際讀前段班的差不多。

的壓力下，仍有國中以各種名目進行能力分班或分組的作法。對此作法，贊成者與反對者也各有其道理，但通常不是建立在嚴謹的實證研究基礎上。可是，要有嚴謹的能力分班效果研究，也非易事。最理想的研究設計是大規模的隨機分派實驗，這在現實上自有極大困難。在無法進行實驗設計的情況下，退而求其次的作法是使用大規模有代表性，且能在一定程度上做因果推論的長期追蹤調查資料。本研究的目的即為利用目前臺灣教育研究可得到最具規模且高品質之TEPS資料，並以各種控制未觀察到變項及選擇偏誤等影響力之統計方法，分析國中生於八年級以後讀前段班對其九年級學習成就之影響為何。

本研究的發現是，不論九年級、還是八及九年級讀前段班，對九年級學習成就都會有一些正面影響，但此影響要看就讀學校的類型，也要看是與誰比較。如果是在沒有分班措施學校就讀的話，雖然有學生不知為何自認在八年級以後讀前段班，但如與同類型學校完全沒意識到分班經驗者中可比較者相比的話，則兩者間的學習成就是沒有差異的。同樣地，在各年級都有分班措施的學校類型內，讀前段班也與沒分班經驗者在九年級學習成就上無顯著差異。此結論需假定PSM結合CCREM的分析因充分考量到資料多層次特性及樣本比較適當性等，且其估計之平均處理效果的偏誤應是最小的情況。此外，本研究也呼應了與Gamoran（1993）之學校分班作法會影響能力分班對學習成就效果的發現。但本研究限於資料，無法如Gamoran般更進一步細緻地指出是學校的哪些措施（如分班選擇性、彈性及範圍等面向）造成此處發現的差異。

為何只有部分年級分班學校八年級以後讀前段班會對學習效果有正面影響，而全部年級都有分班的卻沒有影響？依據TEPS的資料，合理的推測之一是分析樣本中部分年級有分班的學校，確實是在八年級或九年級以後才依照學生過往的學習表現來分班，並選擇學習表現同質性較高者讀前段班，所以可有比較明顯的效果（Hallinan, 1994）。此可能性可由表1觀察到屬部分年級有分班學校中讀前段班者，特別是八及九年級讀前段班者，七年級之學習成就之標準差，相對而言，比其他類型學校讀前段班者略小些。另一可能性是，本研究只針對八年級以後讀前段班的樣本做分析，而全部年級都有分班之學校，可能是從七年級開始就將能力好的學生編成前段班了，因此本研究中此類型學校的樣本是選擇過後的。根據TEPS資料，這兩類型學校所有自認七年級讀前段班者七年級的平均學習成

就，如屬全部年級都有分班學校者會比部分年級有分班學校同類學生在七年級時的平均學習成就來得高些（53.439 vs. 51.289）。但七年級就分班原則上是不符法規的，且就TEPS資料言，並無法符合本研究企圖達成的研究設計。也因此，究竟何者為比較合理的推測，還需要更好的資料與更進一步的研究。此外，需再度強調，本研究PSM分析的結果是只能推論到納入分析且配對成功的樣本。這雖是限制，但也突破了過往可能是比較表面的觀察。要探討能力分班效果是否會擴大學習成就之不平等時，我們需先確定是拿誰跟誰比較。如先前文獻探討所述，過往能力分班效果的研究的最大問題之一是比較的對象並不恰當，而本文最主要的貢獻即是設法在無法大規模實驗能力分班的限制下，以新發展的分析方法比較可比較的對象，並評估國中時讀前段班對學習成就的影響為何。此外，在能力分班仍有不少爭議的情況下，本文也期待成為一個以「證據為基礎」之教育研究及政策討論的示範。

當然，本研究目前的發現還有其他限制及不足之處：

一、除了可能因學校篩選方式不同而影響讀前段班對學習成就的效果外，讀前段班的效果是否需透過其他中介因素，如課程或教學等，也待進一步探討。不過，從反事實分析架構的角度言，因為中介因素本身也是一種potential outcomes，此種研究要比本研究目前所做的更加複雜（Wang & Sobel, 2013）。

二、本研究關注的對象只限於兩種八年級以後讀前段班的公立國中生。研究設計除排除了七年級讀前段班的學生外，也排除了學生進入又退出前段班，同時也未研究讀中段班對學習成就的影響。¹¹特別是讀中、後段班是否會有負面影響，也是能力分班爭議的重點之一。可惜，目前TEPS的資料也無法符合本研究要求的設計。

¹¹ 由於本研究是由TEPS國中學生樣本中選出符合研究目的分析樣本，故可能產生樣本代表性及其推論之母群為何等統計推論的問題。本研究所推論之母群體，比較簡化的說法，即是本研究納入分析之對象所來自之母群體，而納入分析樣本是來自此母群體所有可能樣本之一。而比較複雜的說法，則是比擬隨機分派實驗設計的想像，以納入分析對象參與或未參與前段班的潛在結果所構成之所有可能排列組合的集合，即為此種反事實分析用來做推論依據之母群分配（Guo & Fraser, 2015, pp. 43-48; Morgan & Winship, 2015, pp. 74-76）。

三、本研究能力分班的定義是由學生自己認定的，而非比較客觀的分類。雖然前面提到國中學校能力分班的作法相當多元，但能力分班是用客觀的分類（如美國的研究可以選修的課程分類），還是學生自己的認定也會影響能力分班效果的評估。過往發現學生自陳的資料會測量到能力分班的社會心理面向，並與長期學習成就有相當強的關聯（Gamoran & Berends, 1987）。不過，根據自陳資料的分析結果有可能與根據客觀測量分析的結果不同（Lucas & Gamoran, 2002）。以TEPS的資料言，雖然可以客觀觀察到國中生在七年級班級同學與九年級班級同學是否相同的比例，但如果客觀定義是否讀前段班仍有相當大的限制，需要更進一步的探索。

四、能力分班不只是影響到青少年的學習成就而已，也會對其身心有所影響。這方面的研究自也有待未來有心者的努力。

五、本研究雖然已經使用了相當複雜的統計分析模型，企圖控制不同層次可觀察到及不可觀察到的因素，以求嚴格的因果推論。但這些分析是否達成目的，還是要依靠假定。這是分析觀察性資料的限制，但也需要看未來是否在反事實分析的架構下，能有更先進的分析模型去處理多層次的資料。

DOI: 10.3966/102887082016036201001

參考文獻

人本教育基金會（2002a）。**2002年國中違反狀況調查報告**。取自<http://hef.yam.org.tw/research/08/08-03.doc>

[Humanistic Education Foundation. (2002a). *2002 survey report on the situation of violations of normal class grouping in junior high*. Retrieved from <http://hef.yam.org.tw/research/08/08-03.doc>]

人本教育基金會（2002b）。有教無類難，因材施教易。人本教育札記，**161**，16-19。

[Humanistic Education Foundation. (2002b). It is difficult to educate everyone, irrespective of social background, and easy to teach according to students' aptitudes. *Humanistic Education Journal*, *161*, 16-19.]

人本教育基金會（2003）。**2003年各縣市升學編班狀況調查報告**（台中縣市除外）。取

自<http://hef.yam.org.tw/news/45.doc>

[Humanistic Education Foundation. (2003). *2003 survey report on the situation of ability grouping in counties and cities (except Taichung county and Taichung city)*. Retrieved from <http://hef.yam.org.tw/news/45.doc>]

林彥志（2006）。從學生學習權的角度談我國常態編班政策及展望。《現代教育論壇》，14，294-300。

[Lin, Y.-C. (2006). About the policy and the prospect of ability grouping in Taiwan from the perspective of student's right to learn. *Modern Education Forum*, 14, 294-300.]

張莛雲（2011）。台灣教育長期追蹤資料庫：資料使用手冊（2011.12.01版）、第一波（2001）國中學生問卷。取自http://survey.sinica.edu.tw/srda/teps/W1W2W3W4_JSF_manual_20111201.pdf

[Chang, L.-Y. (2011). *Taiwan Education Panel Survey: Users' guide and the first wave (2001). Student questionnaire for junior high school*. Retrieved from http://survey.sinica.edu.tw/srda/teps/W1W2W3W4_JSF_manual_20111201.pdf]

楊孟麗、譚康榮、黃敏雄（2003）。台灣教育長期追蹤資料庫：心理計量報告：TEPS2001分析能力測驗【第一版】。臺北：中央研究院調查研究專題中心學術調查研究資料庫。

[Yang, M.-L., Tam, T., & Huang, M.-H. (2003). *Psychometric report for the ability tests of TEPS 2001*. Taipei, Taiwan: Center for Survey Research, Academia Sinica.]

謝小苓、楊挽北（2003）。「能力編班」長期追蹤調查—能力分班無法因材施教！人本教育札記，173，34-36。

[Hsieh, H.-C., & Yang, W.-B. (2003). A longitudinal study of ability grouping: Ability grouping could not teach students according to their aptitudes. *Humanistic Education Journal*, 173, 34-36.]

Abadie, A. (2005). Semiparametric difference-in-difference estimators. *Review of Economic Studies*, 72, 1-19.

Alexander, K. L., Cook, M., & McDill, E. L. (1978). Curriculum tracking and educational stratification. *American Sociological Review*, 43, 47-66.

Allison, P. D. (2009). *Fixed effects regression models*. Thousand Oaks, CA: Sage.

Ansalone, G. (2010). Tracking: Educational differentiation or defective strategy. *Educational Research Quarterly*, 34(2), 3-17.

Argys, L. M., Rees, D. I., & Brewer, D. J. (1996). Detracking America's schools: Equity at zero

- cost? *Journal of Policy Analysis and Management*, 15(4), 623-645.
- Ayalon, H., & Gamoran, A. (2000). Stratification in academic secondary programs and educational inequality: Comparison of Israel and the United States. *Comparative Education Review*, 44, 54-80.
- Betts, J. R., & Shkolnik, J. L. (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review*, 19, 1-15.
- Broaded, C. M. (1997). The limits and possibilities of tracking: Some evidence from Taiwan. *Sociology of Education*, 70, 36-53.
- Carbonaro, W. J., & Gamoran, A. (2002). The production of achievement inequality in high school english. *American Educational Research Journal*, 39, 801-827.
- Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1997). *Children, school, and inequality*. Boulder, CO: Westview Press.
- Figlio, D. N., & Page, M. E. (2002). School choice and the distributional effects of ability tracking: Does separation increase inequality? *Journal of Urban Economics*, 51, 497-514.
- Galindo-Rueda, F., & Vignoles, A. (2005). *The heterogeneous effect of selection in secondary schools: Understanding the changing role of ability (CEE Discussion Paper 52)*. London, UK: Centre for the Economics of Education, London School of Economics and Political Science.
- Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of Education*, 60(3), 135-155.
- Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review*, 57, 812-828.
- Gamoran, A. (1993). Alternative uses of ability grouping in secondary schools: Can we bring high-quality instruction to low-ability classes? *American Journal of Education*, 101, 1-22.
- Gamoran, A. (2004). Classroom organization and instructional quality. In H. J. Walberg, A. J. Reynolds, & M. C. Wang (Eds.), *Can unlike students learn together? Grade retention, tracking, and group* (pp. 141-155). Greenwich, CT: Information Age.
- Gamoran, A. (2010). Tracking and inequality: New directions for research and practice. In M. W. Apple, S. J. Ball, & L. A. Gandin (Eds.), *The Routledge international handbook of the sociology of education* (pp. 213-228). London, UK: Routledge.
- Gamoran, A., & Berends, M. (1987). The effects of stratification in secondary schools: Synthesis of survey and ethnographic research. *Review of Educational Research*, 57, 415-435.

- Gamoran, A., & Mare, R. D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology*, *94*, 1146-1183.
- Guo, S., & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and application*. Los Angeles, CA: Sage.
- Hallinan, M. T. (1994). Tracking: From theory to practice. *Sociology of Education*, *67*(2), 79-84.
- Hanushek, E. A., & Woßmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, *116*, 63-76.
- Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, *64*, 605-654.
- Heyns, B. (1974). Selection and stratification within schools. *American Journal of Sociology*, *79*, 1434-1451.
- Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis*, *14*, 205-227.
- Huang, M.-H. (2009). Classroom homogeneity and the distribution of student math performance: A country-level fixed-effects analysis. *Social Science Research*, *38*(4), 781-791.
- Kelly, S. (2004). Are teachers tracked? On what basis and with what consequences. *Social Psychology of Education*, *7*(1), 55-72.
- Kerckhoff, A. C. (1986). Effects of ability grouping in British secondary schools. *American Sociological Review*, *51*, 842-858.
- Kerckhoff, A. C. (1993). *Diverging pathways: Social structure and career deflections*. Cambridge, UK: Cambridge University Press.
- Kim, T., Lee, J. H., & Lee, Y. (2008). Mixing versus sorting in schooling: Evidence from the equalization policy in South Korea. *Economics of Education Review*, *27*(6), 697-711.
- Kubitschek, W., & Hallinan, M. T. (1996). Race, gender, and inequality in track assignment. In A. M. Pallas (Ed.), *Research in sociology of education and socialization* (Vol. 11, pp. 121-146). Greenwich, CT: JAI Press.
- Kulik, C.-L. C., & Kulik, J. A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, *19*, 415-428.

- Kulik, J. A. (1992). *An analysis of the research on ability grouping: Historical and contemporary perspectives*. Storrs, CT: National Research Center on the Gifted and Talented.
- Kulik, J. A. (2004). Grouping, tracking, and de-tracking: Conclusions from experimental, correlational, and ethnographical research. In H. J. Walberg, A. J. Renolds, & M. C. Wang (Eds.), *Can unlike students learn together? Grade retention, tracking, and grouping* (pp. 157-182). Greenwich, CT: Information Age.
- Lee, V. E., Bryk, A. S., & Smith, J. B. (1993). The organization of effective secondary schools. In L. Darling-Hammond (Ed.), *Review of research in education* (Vol. 19, pp. 171-267). Washington, DC: American Education Research Association.
- Leuven, E., & Sianesi, B. (2012). *PSMATCH2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing*. Retrieved from <http://ideas.repec.org/c/boc/bocode/s432001.html>
- Loveless, T. (1999). *The tracking wars: State reform meets school policy*. Washington, DC: Brookings Institution Press.
- Lucas, S. R. (1999). *Tracking inequality: Stratification and mobility in American schools*. New York, NY: Teachers College Press.
- Lucas, S. R., & Beresford, L. (2010). Naming and classifying: Theory, evidence, and equity in education. *Review of Research in Education*, 34(1), 25-84.
- Lucas, S. R., & Gamoran, A. (2002). Tracking and the achievement gap. In J. E. Chubb & T. Loveless (Eds.), *Bridging the achievement gap* (pp. 171-198). Washington, DC: Brookings Institute Press.
- Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal analysis: Methods and principles for social research*. Cambridge, MA: Harvard University Press.
- Mulkey, L. M., Catsambis, S., Steelman, L. C., & Hanes-Ramos, M. (2009). Keeping track or getting offtrack: Issues in the tracking of students. In L. J. Saha & A. G. Dworkin (Eds.), *International handbook of research on teachers and teaching* (pp. 1081-1099). New York, NY: Springer Science + Business Media.
- Oakes, J. (1985). *Keeping track*. New Haven, CT: Yale University Press.
- Oakes, J., Gamoran, A., & Page, R. N. (1992). Curriculum differentiation: Opportunities, outcomes, and meanings. In P. W. Jackson (Ed.), *Handbook of research on curriculum: A project of the American Educational Research Association* (pp. 570-608). New York, NY: Macmillan.

- Rabe-Hesketh, S., & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. College Station, TX: Stata Press.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- Schofield, J. W. (2010). International evidence on ability grouping with curriculum differentiation and the achievement gap in secondary schools. *Teachers College Record*, 112(5), 1492-1528.
- Slavin, R. E. (1987). Ability grouping and achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, 57, 293-336.
- Slavin, R. E. (1990a). Achievement effects of ability grouping in secondary schools: A best evidence synthesis. *Review of Educational Research*, 60, 471-499.
- Slavin, R. E. (1990b). Ability grouping in secondary schools: A response to Hallinan. *Review of Educational Research*, 60, 505-507.
- Thoemmes, F. J., & West, S. G. (2011). The use of propensity score for nonrandomized designs with clustered data. *Multivariate Behavioral Research*, 46(3), 514-543.
- Van de Werfhorst, H. G., & Mijs, J. J. B. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, 36, 407-428.
- Van Houtte, M. (2004). Tracking effects on school achievement: A quantitative explanation in terms of the academic culture of school staff. *American Journal of Education*, 110(4), 354-358.
- VanderHart, P. G. (2006). Why do some schools group by ability? Some evidence from the NAEP. *American Journal of Economics and Sociology*, 65(2), 435-462.
- Wang, X., & Sobel, M. E. (2013). New perspectives on causal mediation analysis. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 215-242). New York, NY: Springer.
- Wooldridge, J. M. (2010). *Econometrics analysis of cross section and panel data*. Cambridge, MA: The MIT Press.

期刊徵稿：<http://www.edubook.com.tw/CallforPaper/BER/?f=oa>

高等教育出版：<http://www.edubook.com.tw/?f=oa>

高等教育知識庫：<http://www.ericdata.com/?f=oa>