

大數據

遊走書字行 考出來

證中國大學收生比美國更公平

香港文匯報訊（記者 歐陽文倩）教育制度的優劣是一個廣闊無垠的議題，要掌握當中現象作深入探討，收集大量資料並加以分析的大數據研究方法或許是其中一個出路。科技大學人文社會科學學院院長李中清收集北京大學和蘇州大學前後逾半世紀、約15萬份歷屆本科畢業學生的資料，以剖析中國的教育制度如何增強社會階層流動性，實現一場「無聲的革命」。有關研究更打破過去人們以為美國大學教育更平等的觀念，以數據說明中國高校招生更「英雄不問出處」。

就針對北大蘇大兩校作數據挖掘對象，李中清解釋說，前者是中央直屬重點大學，其生源反映全國範疇精英教育的狀況，後者則是省屬重點大學，反映省級區域精英教育的生源；再加上中國不同檔次的高校招生標準相似，一所國家級加一所省級大學已能做到見微知著，有助了解整體情況。

蒐半世紀 15萬畢業生資料

雖然只是兩所大學，但當中數據仍龐大。李中清指，在收集到兩校1933年至2003年的數據庫中，共涉及約15萬位本科畢業生個人及家庭背景資料，其中北大佔約6.5萬個，蘇大佔8.6萬個。他的研究則主要有系統地選取1952年至2002年的情況。

是次研究的數據內容複雜，涉及數百個院系專業、8,000多種家長職業名稱、5萬多個學校名、住址也有近13萬個，而首要是要將實體的學籍卡電子化，李中清直言有關過程歷經逾30年，「我是由1983年就開始做的，頭半年只有我一個人，因為要對資料有一定理解，所以親自輸入。後至（上世紀）90年代，再聘請退休知識分子，約五六人做，一直做了30多年。」

研究揭美教育制度較不公平

但有關努力無白費，透過對資料的挖掘與分析，李中清的研究團隊取得令人意外的結論。他說：「其實美國教育制度較不公平。我曾以為中國也是一樣，但十多年前，中國比美國公平得多。」美國最精英的私立大學至1969年才接受女學生、對少數民族學生入學仍有諸多限制，但是次研究反映，在1980年至1999年間北大農村子弟學生在10%以上，蘇大則有40%以上，而隨中國社會發展，整體農村人口比例後來不斷減少，但至21世紀，兩校農村生比例卻無大幅下降。此外，對外國大學以幹部、專業技術人員和商業服務業人員子女學生佔絕大多數，北大於1952年至2002年間有逾30%學生屬工農子弟，蘇大有約40%。

高考助勤苦工農子弟升大

李中清認為，有關現象主要與教育政策有關，而藉高考成绩收生是其中主因，「高考制度的內容與形式，於文化資本相對缺乏的社會大眾有幫助，工農子弟透過苦練亦能考取佳績。若高考分數非唯一標準，富裕家庭子女便可透過其他課外才藝增加入讀精英大學機會。」

除高考外，中國各級學校制度也對教育公平性有貢獻，李中清說，國家級、省級重點大學的生源多樣性在很大程度上依賴重點中學的生源多樣性。雖然數據呈現中國早年教育改革能一定程度趨向公平，但他強調，工農子弟的實際入學比率與其相對工農人口比率仍有

很大差距，「而近十年的生源亦有改變，值得再研究下去。」



李中清認為，大數據應用於社會科學研究而言是更客觀、更科學。歐陽文倩攝



港大電子學習發展實驗室將校內電子學習項目擴展至70所港澳及內地中小學。記者劉國權攝

冀結合電子教學 即析成績提效

三地共研

學生數據對剖析教育制度是重要原始資源，但因貢獻數據者多已完成學業，他們難以受惠；但如將大數據理念結合新興電子學習（E-learning），便為學生提供實時分析回饋。香港大學電子學習發展實驗室積極將其電子學習項目擴展至70所港澳及內地中小學，又與北京的國家開放大學合作推試點計劃，跟進逾6,000名法律課程學生數據，甚至微觀追蹤學生於個別題目表現，讓科技與數據成為提升教育成效的關鍵。

每次公開試、入學篩選等均會產生大量教育數據，但港大電子學習發展實驗室總監霍偉棟指，那都屬於「終結性評核」，用以量度學生完成整個階段後的成果，並有助檢視課程或教育體制，惟涉事學生因已離校，數據不會對其學習有直接幫助。

全國法律課成績比拚 京生跑出 霍偉棟表示，結合電子學習的「進展性評估」數據可針對性為學生作實時回饋，最典型做法是大批學生透過電子平台答題，老師能藉當中共同錯誤了解其弱項，並在教學中即時調節強化。他又指，在教育普及及下相關數據顯得更重要，「當學生人數少，老師可憑觀察與經驗判斷不同人的學習進度，但當學生增加、學習差異擴大，判斷便變得困難；電子學習的其中一個目標便是透過數據而非『感覺』，讓教與學分析來得更快、更有依據。」

他指，實驗室幾年前為大學四年制下約千名工程本科生開發電子學習平台，隨後項目獲各地教育界重視並逐步推展，現時香港、澳門、杭州及蘇州有約70所中小學參與，涉及的教育數據也大增；而實驗室今年更與註冊學員達300萬的國家開放大學合作推試點計劃，結合電子學習剖析逾6,000名全國各地法律課程學生背景資料及成績表現，初步顯示北京學生答題成績較優秀，團隊希望透過電子學習數據與開放大學合作，促進教育成效。

香港文匯報記者 任智鵬

助社科研究更客觀科學

開闢新道

社會科學研究往往都以抽樣和個案研究去進行，做起來有一定限制。李中清認為，大數據的出現，並將它應用在社科以至歷史研究中，可令研究變得更科學，所有人都可憑這些客觀數據，去再現（Re-produce）一樣的結果，令社科研究亦能像物理或化學的實驗一樣，「大數據有着真正的『現象』。」李中清解釋說，以往社會科學一些較大規模的研究會以抽樣為主，但當中包含很多複雜問題，「抽樣首先要社會人口比較一致，而且母、分子的比例性樣本要有代表性。」他進一步說，抽樣不易之處，還要考慮現象、分布和可行性的問題，「例如貴州的情況就和江蘇的很不一樣，中國那麼多個省、那麼多個縣，跑一躺也不容易，除非是國家級單位幫忙，否則很難。」

李中清：中國為最佳實驗地

用大數據去做研究則開闢另一種新渠道。李中清表示，大數據是根據資料檔案去做，而中國更是進行此類研究的最佳地方，「中國很早以前就有很多檔案，像科學資料亦保留很多年，由漢朝

的選拔制度起已有記錄，而且保存得很好，有些適合用大數據方式去做。」他又指，用大數據做研究不需統計專業背景、不需電腦培訓，「但對史料的保存要下很大工夫，後人才可根據他們的基礎去做研究。」

出身自理科家庭、父親為諾貝爾物理學獎得主李政道的李中清形容，這是很科學、很前沿的研究方法，「我的家庭就常常說，大自然比我們聰明得多。要做研究，就要尊敬現象。先收集材料，承認自己的理解和看法不一定對，再找現象，讓現象決定結論。」在李中清看來，大數據所呈現的就是真正現象，「現在有社交媒體、網絡，發展的速度也很快，創造了很多訊息可以分析，而且這很科學，任何人都可以再現，學界也認可。」

欠個體代表性 須有相同想法

不過，大數據用作社科研究亦有其缺點，就是欠缺個體代表性，而研究速度亦比較慢，怎樣利用、怎樣計算，都很看重研究人員有沒有相同想法，「但它在時間上有厚度，對理解轉變和延續十分有用。」

歐陽文倩

「三大」學生多出身基層

身窮智富

除內地大學外，李中清亦有研究香港院校生源狀況。他發現，至2008/09學年，本港精英大學的學生有相當比例來自並不富裕的普通家庭，香港大學、中文大學和科技大學這「三大」中，只有少於25%學生的家長受過專上教育，更有多於25%只得小學或更低教育程度，「很多孩子都是他們家裡第一代的大學生。」

在美國任教多年的他指，當地大學較向富裕家庭子女傾斜，期望將來進一步研究不同地方的升學制度，「香港、美國、英國都有考試，為何香港的更公平？」

盼「多數人創造黃金時代」

李中清特別提到，近年本地大學經聯招入讀的學生似有減少、非聯招的則上升，對此不無擔心，「前者是沒甚麼機會的家庭，後者則是精英家庭。很多有才能孩子因為在社會下半層，沒機會發揮，浪費潛在人力。所以多收聯招的更重要，不要浪費人力資本，但如何幫社會、國家認出這些『資本』，好好培訓，這是關鍵所在。」

他認為，隨着教育普及，若能好好運用人力資本，更能投射出美好遠景，「以前的文明由少數人創造，未來可以由多數人創造，這是我們的黃金時代。」

歐陽文倩