

美國學生學習歷程縱貫性大數據系統 P-20W 的啟示與借鏡

洪錦珠 蕭佳賓 李育齊*

摘 要

建置教育大數據主要目的在於針對社會各界所關注的教育議題尋找影響學生學習成果的關鍵因素，以做為改善教育制度與教學現場問題、預測未來教育趨勢之參據，落實以「證據為本」的教育決策模式。

本研究旨在探討如何建構和推動學生學習歷程的縱貫性教育大數據，爰美國建構 P-20W 縱貫性大數據系統多年並獲得豐碩成果，故以其為研究對象，蒐集相關法令及實際案例，採用文獻分析法進行探討和說明。

研究發現 P-20W 是以改善學生學習成效為策略目標，建置學生自學齡前教育至就業的個人化歷程資料和教育研究為手段，擷取與分析足以有效達成策略目標的資訊，提供教育決策及改善教育現場問題參考。為落實該項策略並確保資料品質，投入資源並訂定相關法案，建構有利的環境和摒除可能的障礙，並輔以關鍵配套措施。

本研究結合教育與資訊技術的角度，參考資訊系統規劃實務，

洪錦珠，華夏科技大學資產與物業管理學系副教授

電子郵件：cch@cc.hwh.edu.tw

蕭佳賓，華夏科技大學資訊管理學系兼任助理教授

電子郵件：xiao@mail.csf.org.tw

* 李育齊（通訊作者），國立臺灣師範大學全人教育中心博士研發管理師

電子郵件：jomire@ntnu.edu.tw

投稿日期：2018年8月26日；修正日期：2018年12月4日；接受日期：2019年2月19日

從資料策略，確認教育需求與問題，建立資料治理結構與流程，到學生個資隱私、安全與保密，依序呈現並提出建議，以作為我國建置此類資料庫的參考。

關鍵詞：教育大數據、縱貫性資料、資料策略、資料探勘、品質保證

ENLIGHTENMENT AND EXPERIENCE OF THE P-20W LONGITUDINAL BIG DATA SYSTEM OF AMERICAN STUDENTS' LEARNING PROCESS

Ching-Chu Hung Jia-Bin Hsiao Yu-Chi Li*

ABSTRACT

The main purpose of constructing an educational database is to find the key factors that affect students' learning outcomes in view of the educational issues that are of great concern by all sectors of society, so as to improve the educational system and teaching site problems, to predict future educational trends, and to implement an educational decision-making mode that is "evidence-based".

The purpose of this study is to explore how to construct and promote the longitudinal educational database of students' learning process. Since the United States has already obtained fruitful results from its construction of the P-20W longitudinal educational database, this study takes it as the research object, collects relevant laws and practical cases, and uses the document analysis approach to explore and explain the findings.

This study notes that P-20W has taken a series of actions to achieve its strategic objective at improving students' learning effectiveness, which include building personal data on learning for each student from pre-school through workforce entry, capturing information sufficient to effectively achieve its strategic objective, analyzing and providing this information to help stakeholders make educational decisions, and improving problems in the classroom. To implement the strategy and

Ching-Chu Hung, Associate Professor, Department of Assets & Property Management , Hwa Hsia University of Technology, New Taipei City, Taiwan.

E-mail: cch@cc.hwh.edu.tw

Jia-Bin Hsiao, Adjunct Assistant Professor, Department of Information Management , Hwa Hsia University of Technology, New Taipei City, Taiwan.

E-mail: xiao@mail.csf.org.tw

*Yu-Chi Li (corresponding author), Research & Development Manager, Center of Holistic Education, National Taiwan Normal University, Taipei, Taiwan.

E-mail: jomire@ntnu.edu.tw

Manuscript received: August 26, 2018; Modified: December 4, 2018; Accepted: February 19, 2019

ensure data quality, P-20W has invested a lot of resources, formulated relevant bills, set up a helpful environment, removed related barriers, and supported it with various activities.

Based upon the perspective of education and information technology and referring to the planning practice of an information system, this study presents the research results through the sequence of working out data strategy, identifying the educational needs and problems, establishing data governance structure and processing, setting up privacy, security, and confidentiality policies and facilities for students' data, and finally proposing some suggestions as a reference for the establishment of such databases in Taiwan.

Keywords: educational big data, longitudinal data, data strategy, data mining, quality assurance

壹、前言

當今世界各國教育單位正為支持學生成功（Student Success）而殫精竭慮，也因課程與教學典範的轉移，強調以學生為中心的學習模式，重視學生學習成果的認證評量，成為各國教育變革所追求的目標之一（彭森明，2010）。為此，教育單位蒐集各種教育現場的教學資料，探究其影響學生學習的因素，做為規劃措施以提昇學生學習成果的依據。過去學生學習的資料取得與分析較為不便，且以總結性評量為主，故造成資料分析時效果有限。但在教育大數據時代，長期大量的學習行為資料數據之蒐集、分析、應用將大大提升學習成效，積極支持以「大數據」（Big Data）來改善教學模式及落實適性教學（張佳琳，2014），教育單位可擷取過去不易取得的資料，讓學生「學習過程」也可資料化，並利用數據分析，將之回饋至教育決策機關及教育現場，以改善教育體系（Mayer-Schönberger & Cukier, 2014）。透過持續性蒐集學生資料，進行跨層次的資料整合與分析，分析及預測學生的學習表現，提供上級決策者改善教育政策和教育現場工作者改進教學方式的參考，讓所有學生得以提升學習成就。換言之，教育大數據資料庫不僅是累積學校辦學的知識庫，也是支援學校運作及決策的資產，運用相關教學數據以提升學校效能、教師教學與學生學習的時代已經來臨（王金龍，2015；曾元顯，2016；彭森明，2013；Mandinach & Gummer, 2013）。

舉例而言，新加坡教育部成立教育資訊情報中心，針對教育現場中相關的訊息進行彙整，從中進行整理分析以了解教育趨勢與變化；美國也於2012年揭櫫「大數據研究與發展方案」（Big Data Research and Development Initiative），在84項計畫中，教育領域也蘊含其中；此外像是歐盟、日本、韓國也有類似策略的規劃（蔡明學、黃建翔，2015）。上述各國試圖利用大數據，針對社會各界所關注的教育議題，尋找影響學生學習成果的關鍵因素，做為改善教育、預測未來趨勢之參據，落實「以證據為基礎」（Evidence-based）的教育決策模式。

教育大數據分析在於探索「好的議題」、「有用的資料來源」與「合適的研究方法」，以創造學校所需的分析結果（Rios-Aguilar, 2015），然而教育大數據的建置，實務面仍有許多問題尚待解決，例如建置哪些型態

的大數據、涵蓋範圍、如何建置，以及建置過程中可能面臨的挑戰，均需全盤瞭解背景脈絡並未雨綢繆。根據美國國家教育統計論壇（National Forum on Education Statistics [NFES]）研究顯示，美國自 2001 年起逐步立法，並在 2009 年通過「美國永續富強投資法案」（The American Recovery and Reinvestment Act），提撥美金 5,360 億，由教育部統籌並建立縱貫資料體系，以政策和經費支持教育大數據的建置和應用（NFES, 2010）。多年推動下已獲初步成果，以及彌足珍貴的經驗。

在我國教育先進期盼政府評估建置教育資料庫，落實教育大數據分析模式之際，本研究以文獻分析為研究方法，探討美國 P-20W 縱貫性大數據系統（longitudinal data system [LDS]）的建置歷程，以回應下列四個研究問題：

- 一、美國建置國家級縱貫性教育大數據系統之目的？
- 二、建置國家級縱貫性教育大數據系統應具備哪些戰略思維及其成功關鍵要素？
- 三、如何建置一個品質保證機制來確保縱貫性教育大數據系統品質？
- 四、臺灣若要建置類似的縱貫性教育大數據系統，美國經驗有可為借鏡之處？

貳、國內縱貫性教育資料庫建置概況

就縱貫性教育資料而言，長年以來我國各級學校幾乎都儲存有學生的基本資料及在校的學習歷程資料，只是這些資料都是由不同廠商協助建置，其資料項目及格式是否一致則尚待調查研究。這些資料皆獨立運作，且多以支援校務行政為主。瀏覽各大專校院，幾乎每一所學校都建有學習歷程網站（e-profile），或許相對完整的資料是須要密碼才能閱讀到更深層資訊，否則單從網站頁面觀察，資料項目非常短少。其次，近幾年教育部鼓勵大專校院進行校務研究，但觀察其發表的研究結果，尚少有橫向跨校或跨領域者，但至少此皆是以學生為本的資料。它們雖然屬縱貫性資料，可惜的是尚未全然具備大數據應有的 5Vs 條件—規模、多樣性、高效性、準確性和價值性。

根據楊孟山與林宜玄（2016）研究，國內大型教育資料庫共有：（一）臺灣教育長期追蹤資料庫（Taiwan Education Panel Survey [TEPS]），是由中央研究院、教育部、國家教育研究院籌備處（2004~2007）和國家科學委員會（2000~2008）共同資助，並由中央研究院之社會學研究所和歐美研究所共同負責規劃與執行的一項全國性長期的資料庫計畫。這個資料庫是以問卷調查方式，蒐集國中、高中、高職、及五專學生資料。再以這些學生為核心，擴及學生家長、老師及學校。與學生相關的資料，可利用學生代碼做連結。在公共版部分，不提供學校與班級代碼，學生資料亦不能與教師資料連結，僅在 TEPS 限制版可連結學校與班級代碼，學生及教師資料。這項計畫分兩期，第二期至 2017 年結束。（二）臺灣學生學習成就評量資料庫（Taiwan Assessment of Student Achievement [TASA]），是由教育部函請國家教育研究院辦理的常態性計畫，蒐集國小、國中、高中、高職此三個教育階段國語文、英語文、數學、社會、自然此五科之學習成就表現，做為教育部研訂課程與教學政策依據。它是採抽樣方式取得學生在前述五個領域的學習成就表現資料。根據國家教育研究院網站公佈資料（2018）顯示，其目的為追蹤、分析學生在學習上變遷之趨勢，進而檢視目前課程與教學實施成效；並提供完整、標準化的學習成就資料，做為分析學生學習成就上差異表現變項資料，評估學生未來在學術方面能力之發展與社會期許；且瞭解國內學校教學及學生學習成效之現況，做為課程與教學政策改進之參考，並為縣市政府教育局及學校推動補救教學之重要參據；提供各縣市學生學習表現資料，建立與縣市合作機制，以擴大資料庫應用效益；以資料庫的量化資料，提供國內外相關研究人員，深入探討學生學習成就方面的相關政策議題；最後希冀建立本國學生學習成就評量資料庫，同時考慮與國際接軌，利於加入國際比較行列，藉以瞭解臺灣教育之獨特面與優缺點。（三）臺灣高等教育整合資料庫（Taiwan Integrated Higher Education Database [TIHED]），是由彭森明教授帶領下將「大一大三調查」、「大專教師調查」及「大專校院畢業生流向調查」整合而成的橫斷性及縱貫性資料。這項計畫主要是以網路問卷調查方式蒐集資料。這項計畫自 2004 年開始執行，可惜 2016 年起資料庫已正式結束，無法永續執行。（四）後期中等教育長期追蹤資料庫（Taiwan Upper Secondary Database [TUSD]），是由教育部中等教育司委託臺師大教育研究與評鑑中

心建置，以因應十二年國民基本教育的推動，滿足不同層級、不同對象使用者對其中等教育資訊的需求。該資料庫以普查暨網路問卷方式，蒐集全國公私立高中、高職及五專（僅五專前三年）、校長、教師、學生及家長資料，其調查內容包含不少心理或課堂外學習的面向，如國三學習狀況、升高中期望、生涯規劃與身心健康等，調查週期也隨對象分每年或每四年一次。

臺灣除上述楊孟山與林宜玄（2016）蒐集的四個具縱貫性特質的大型教育數據資料庫外，還有教育部目前正在興建的「高級中等教育階段學生學習歷程檔案」。這是教育部為落實十二年國民基本教育適性揚才的理念，達成引導多元適性發展之目標，於 106 年 7 月 26 日公布「建置高級中等教育階段學生學習歷程檔案作業要點」做為發展這套數據庫的依據。該項作業要點將於 108 年 8 月 1 日生效，冀望透過系統化記錄個人學習歷程資料，引導學生發展多元智能、性向及興趣，進而選擇適合自己的升學或就業進路。該歷程檔案已具以學生為本的縱貫性教育資料的特質，其目的為「掌握學生學力、追蹤課綱實施成效，做為相關教育政策研擬、資源配置運用之依據。」、「大學選才參採」、「促進教育機會均等」。它蒐集的資料包括學生的基本資料、修課紀錄、課程學習成果、多元表現（包含幹部經歷、競賽成果、檢定證照、志工服務等）、自傳（包括學習計畫）。由於負有「大學選才參採」的使命，該系統加入防弊用途的檢核機制並將備審資料的格式統一為 PDF。依照教育部（2018a）公佈的資訊顯示，該資料庫包括學業表現歷程資料庫、非學業表現歷程資料庫、學習歷程紀錄模組、實驗教育學生學習歷程資料庫，但這些資料在學生申請入大專校院起 5 年後即被封存。

此外，在國民中小學部分，2000 年進行「全國各縣市國中、小教育行政電子化推動實施計畫」因應九年一貫課程開始實施，為滿足學生成績評量及學生學籍管理電子化之需求，教育部為因應十二年國民基本教育課綱的實施，於 2018 年修正建立「國民中小學學生學籍交換規格標準 4.1 版」正式版，以做為中小學學生資料交換、校務行政系統欄位規劃，以及教育政策分析之用（教育部，2018b），但尚未串接至高中職、大專校院等之相關資料庫，甚是可惜。

參、P-20W 縱貫性大數據系統推動的策略與內涵

一、確立國家發展願景及具戰略目標、形成政策並投入資源

P-20W 縱貫性大數據資料蒐羅範圍由幼教至學生就業階段，其中涵蓋國高中及大學求學經驗，其目的是建構學生學齡前至就業的縱貫性教育大數據系統，應用教育歷程資料進行研究分析，進而制訂相關策略以改善學生學習成果（NFES, 2010），其關心的關鍵政策問題（Guidera, 2011; Kowalski, 2013），如圖 1。

NFES（2010）認為縱貫性教育大數據系統為一個具有前述特質的資料系統，蒐集及維護詳細、高品質的學生層次及教職員層次資料，並將這些資料做跨個體和跨時間的連結，為每位學生建置完整的學業及成就歷程資料。進而透過製作報表和分析，使資料得以被清楚擷取。此外藉由持續性蒐集個別學生資料，將片段資料進行連結，依據關鍵學習資料做跨層次的整合，分析關鍵學習資料對學習成果影響，及變項間的關係。再者亦分析及預測學生的學習表現，其結果將成為中央和地方的政策制訂者，及教育現場工作者改良教育政策和教學方式的依據，讓所有學生得以提升學習成就。

美國政府之所以建置 P-20W 是符合國家發展願景並具有其戰略性目的（Shah, 2012），因多數的教育研究者，都使用總體但橫斷性的資料做統計分析，但橫斷的分析無法提供個別學生進步及增值（value-added）狀況的資訊。縱貫性資料則可補強橫斷性資料的不足，提供研究者進行更豐富分析的機遇，有助於將資料轉化成可據以行動的資訊。正如 Data Quality Campaign [DQC] 創辦人 Guidera 於 2013 年在美國教育研究學會年會裡的論文報告中所言：有效用的教育資料應為：（一）縱貫性的資料，能長期追蹤學生經驗與成果，且是跨系統和跨部門；（二）可據以行動、即時回饋，及對使用者友善且客製化的資料；（三）具情境脈絡、嚴謹且可分析比較的資料。要確實探究學生長期增值動態，單靠橫斷性資料是有限的。綜言之，美國為培養具國際競爭力的下一代，將各層級學生的成功視為國家級教育目標，決心建置縱貫性教育大數據來落實，並以立法及經費來支持該項決策。

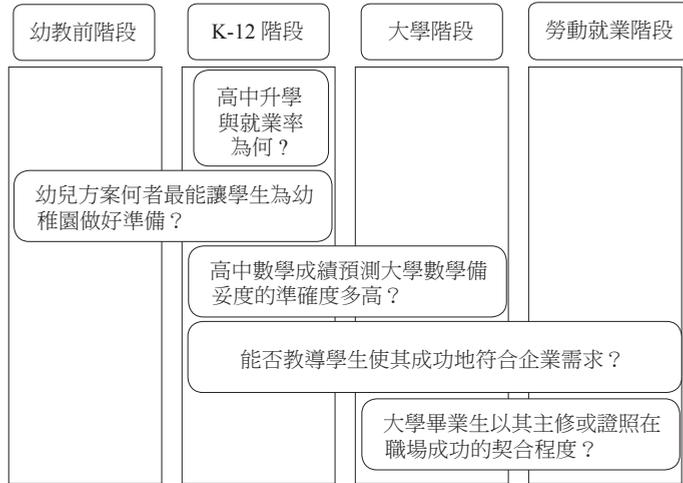


圖 1 P-20W 關鍵政策問題

資料來源：Kowalski (2013). *P-20W Data Government: Critical Roles for Success*, (p.4).

為實現其目標，美國國會自 2001 年後通過一連串的法案，形成政策並投入龐大的資源。例如 2001 年通過「不讓孩子落後法案」(The No Child Left Behind Act)，允許各州教育機構將評量資料納入州政府開發的縱貫資料系統，連結學生的測驗成績、就讀年數、畢業成績。2002 年通過「教育科技支援法案」(Educational Technical Assistance Act)，授權州政府得以設計、開發和實施縱貫資料系統，以有效且正確地管理、分析、細分並使用學生資料，並要求教育部長應檢視流程以確保資料的品質、促進跨州連結、保護學生隱私，及資料的正確性與即時應用，希冀以研究改善學生的學習成就，並縮短教育落差 (U.S. Department of Education, 2002)。該法案同時要求生效 3 年後，教育部長在徵詢國家統計學術委員會 (National Academies Committee on National Statistics) 的意見下，製作一份調查分析報告，就聯邦、州、地方政府為設立該法案的目的，所做規劃與努力公諸於世。其內容包括：確認並分析全州縱貫資料系統的發展與使用狀況；評估該系統處理學生個人資料，及促進跨州連結和保護學生隱私的能力；確定持續改善最佳的工作實務項目與領域等。

2007 年再通過「美國競爭力法案」(The America COMPETES

Act），法案宗旨在經由研究與開發、投資與創新，提升美國的競爭力。並要求各州建置全州性的 P-16 教育縱貫性資料系統。當學生入學時，即給予一組獨一無二的辨識碼。所有學生在 P-16 就學期間的資料將被保存，同時要求此縱貫性資料，必須具備與高等教育資料系統介接，並審查及評估資料的品質及信、效度。2009 年又通過「美國永續富強投資法案」（The American Recovery and Reinvestment Act），核定州財政穩定基金 5,360 億美元，由教育部統籌管理運用，建立縱貫資料系統，以政策和經費支持教育大數據的建置和應用，並於當年度開始推動 P-20W 計畫。教育當局認為學生學習具連貫性的，想提升學生學習成果，必須參考各階段的教育歷程資料。再者擁有縱貫資料分析結果，相關決策者便可據此做更明智的教育決策。例如改善學生的學習，提高學生的學習成就並縮短落差，期許在尚未對學生的學習成功產生負面影響前，確定並排除潛在且重複發生不利學生學習的因素，進而落實因材施教的理想（NFES, 2010）。

爾後，美國教育部揭櫫 2014—2018 年六項國家教育策略目標，持續改善美國教育制度，運用大數據的資料分析模式，以提升各的州教育品質，並將教育大數據分析的結果，做為政府補助之依據，落實證據本位的決策模式（U.S. Department of Education, 2015）。

以上述得知，美國政府將應用教育數據協助學生成功學習當作戰略目標，並透過制訂相關法案，利用由上而下的行政執行模式，促使各州政府及轄下各級學校，理解並適切運用教育大數據的分析結果，再以立法及資源投入予以實現。

二、制訂 P-20W 的資料策略

設定國家級戰略性目的並獲得立法支持後，P-20W 就開始為資料策略做定調。P-20W 將資料途徑分為四個層次，做為各級教育單位規劃縱貫性大數據系統架構的依據。其中，第一層是資料來源層，教育工作者詳細評估，回答或解決某些教育現場問題。第二層是資料儲存層，透過類似 Apache Hadoop 的分布式檔案系統或 Google 檔案系統，取得不同來源的資料並儲存，惟這些原始資料皆供系統判讀之用。因此，還須應用管理者可理解的方式，組織和分類資料的系統來儲存和管理。第三層是資料處理 / 分析層，供教育工作者使用資料去處理和分析，探究有用的教育資訊。例

如利用 MapReduce 這類工具來選擇欲分析的教育現場資料，並且將它轉換成可從中獲得可洞見教育現象的格式；或使用諸如 Apache Pig 或 Hive 這類資料倉儲軟體工具去查詢資料，或使用自動化的樣態辨識工具去決定未來教育發展趨勢，及從教育資料分析中導出結論。第四層是資料產出層，交給教育決策或工作者，應用各類工具，產出清晰、扼要的報告和建議（Bernard, 2014a）。

舉例來說，美國北卡羅萊納州政府，即秉持此資料分析模式取得、連結、提供和執行做為該州的教育大數據資料策略，以網頁服務即時取得 179 個地區的資料，集中存放於資料庫，再經萃取—轉換—上載（Extract-Transform-Load [ETL]）至州教育資料倉儲；這個資料倉儲透過統一辨識碼，連結幼教、高等教育、職業訓練、就業等資料庫並持續更新，再按照共享式學習基礎結構（Shared Learning Infrastructure），提供教育決策或工作者進行分析研究，並讓州教育廳以學習管理系統產出管理報表，而教育研究者則產出研究報告，並據以化為具體行動方案（North Carolina Department of Commerce, 2014）。為完成此一戰略性目標，該州結合了公共事務、財務、教育、勞工、經濟、資料分析、幼教、健康等 8 個單位協同作業，此協同合作架構與其整體的戰略性方向和行動方案是相呼應的。

三、提出 P-20W 的行動方案

資料策略訂定後，各州政府即接續研擬縱貫性教育大數據的方向和行動方案。Guidera 歸納各州政府執行 P-20W 行動方案十項要點：

- （一）全州 K-12 資料系統與學齡前、大學教育、職場、社會服務和其他關鍵的州立機構資料系統連結。
- （二）為嚴謹的州縱貫資料系統，建立穩定、持續性的支持。
- （三）發展系統治理結構（governance structure），引導資料蒐集、分享與使用。
- （四）建立州立的資料倉儲，整合學生、教職員、財務與設施資料。
- （五）保護學生隱私的同時的前提下，執行系統並提供相關人士資訊的即時擷取。
- （六）提供教育者、父母親和學生用以改善學習表現的資料，製作學習歷程相關的報表。

- (七) 製作以學校系統和學生群組的縱貫性統計報表，做為學校、區域和州政府制定相關政策方案之參據。
- (八) 發展具目的性的研究時程表，並與大學、研究者和中立團體協同合作，探索益於產出對教學現場有效的資料。
- (九) 執行政策並推動各種系統實務訓練，包括專業發展與認證，確保教育者知道如何適當擷取、分析和使用資料。
- (十) 提升相關人士對資料的覺察，包括州政策決策者，知道如何擷取、分析與應用資訊。(Guidera, 2013, p.4)

綜合上述，行動方案執行符合 P-20W 縱貫性教育大數據的方向，包含擴充州立各教育階段縱貫資料系統連結，以跨 P-20W 教育管道並跨州立機構為目標；並確保資料庫資料可被觸及、分析和應用，並可傳遞給相關教育利害關係人，落實教育現場持續性的改善機制；最後更鼓勵所有相關人士使用縱貫資料，促進有效決策。

四、確立 P-20W 資料的基本要素

依據 NFES (2010) 的報告，P-20W 將資料區分成基本及擴充兩大功能類，基本功能在於符應 P-20W 的共同核心目標，擴充功能則是為更大的效率和能力而設，如隨堂和形成性評量資料、財務資料、教育設施資料、與社會服務機構資料分享、地理資訊系統等。

據 Guidera 研究指出在 P-20W 推動初期，各州資料皆被要求應具備 10 個基本要素，再據此蒐集數據，分別是：

- (一) 建立全州學生獨一無二的辨識碼。
- (二) 學生層次註冊、社會背景和教學方案的參與等資訊。
- (三) 比對個別學生逐年的考試成績，以測量其學業表現與成長。
- (四) 瞭解未參加考試學生的資訊。
- (五) 可將教師與其班上學生配對的教師辨識碼系統。
- (六) 學生的學習紀錄，包括完成哪些課程和獲得的成績。
- (七) 學生大學入學的測驗成績。
- (八) 學生的畢業和退學資料。
- (九) 比對學生在 P-12 與大學系統間紀錄的功能。

(十) 重視評核資料品質，建立具信、效度的州立資料審查系統。

(Guidera, 2013, p.3)

綜上所述，P-20W 為讓教育工作者能有更完整的資料進行教育決策與分析，在確立資料需求時，決策當局首先創造學生獨一無二的辨識碼，並將之與教師配對，更將學生考試以外的學習歷程資料納入，例如學生個人背景資料、註冊、出缺勤、課堂學習資料、學習障礙、形成性和總結性評量結果、大專學習成就、學習成果、休退學記錄等，以及教師教學與家庭經濟等，並且決定這些資料必須跨區、跨教育部門、跨幼教至大專和就業階段，最後更重視蒐集資料之品質，建立信、效度的州立資料審查系統以評核之，如此確保資料的大數據 5Vs 原則。

五、培育大數據創新應用人才

從 NFES (2010) 的報告顯示在規劃建置 P-20W 的軟硬體系統時，其也同步培訓教育大數據創新應用能力的人才。另由於翻轉學習的帶動，教育現場既是大數據的使用者，同時也將是大數據的提供者。大數據將改變世界，物聯網、雲計算勢必逐漸應用於教育現場及各個角落，感知器、機器學習和視覺化的技術人才將極為欠缺，因此美國政府積極培養相關人才，規劃以華盛頓大學為基地，開設大數據研究與教育課程，傳授資料管理、資料視覺化、感知、開放資料萃取、系統與網路等專業知識與技能，並積極建置機器學習研究中心，培育相關人才。

根據 Bernard (2014b) 建議教育人員應用教育大數據須具備六種技能：

(一) 經營技能 瞭解經營目標與資訊需求。(二) 分析技能 決定蒐集哪些資料和如何分析的能力。(三) 創造力 以新方法蒐集、分析和詮釋資料的能力。(四) 電腦科學技能 熟練數據演算法，熟悉 Hadoop、Python、Pig 等工具。(五) 數學與統計技能 探究及分析數據的能力。(六) 溝通技能 呈現資料 (口頭及書面) 以確保分析報告內容被了解並付諸行動。實務上，要具備這些技能，資料科學家、演算專家是不可或缺的角色。Davenport 與 Patil (2012) 認為執行教育大數據之人員應具備技術專業能力，並具備發現問題、探究問題的好奇心；以及運用分析資料剪報與溝通能力，以及多元而富創意的方式分析問題的能力。

在人力資源管理方面，Shah (2012) 以政策制定、治理的觀點提及人

員的適才適所是縱貫性資料庫永續經營的關鍵之一。此外 Zeke (2017) 檢視 50 個州立長期追蹤資料庫系統後，指出跨部門間的合作、組織人員間建立一套共同的語彙、持續為資料庫建置與運作無私奉獻，及提升組織人事穩定度，將有助於系統的成功和永續推展。

六、編輯指導手冊、完整解說

為了讓社會各界充分瞭解何謂縱貫性大數據系統 (LDS)、如何規劃及開發 LDS、如何有效地經營 LDS 資料，並做高階的應用，NFES 分別在 2010 年 7 月、2010 年 10 月、2011 年 1 月、2011 年 7 月分別發佈了四本「Traveling Through Time」指導手冊 (NFES, 2010; 2011)。第一冊說明什麼是 LDS、什麼不是 LDS、為何要建置 LDS。第二冊說明規劃 LDS 的先備條件、盤點手中已有以找出欠缺、如何開發 (自建或委外)、如何撰寫招標文件。第三冊說明如何治理資料、如何提升資料品質、如何取得資料並保護個資。第四冊說明如何以微觀或巨觀方式應用資料、要有效應用資料的先備條件、如何專業地發展 LDS 的資料和舉辦教育訓練，教導各方人士有效應用資料。

肆、P-20W 縱貫性大數據系統的資料品質保證機制

卓越的資料品質是資料庫建立的目標之一，P-20W 的建置便將確保資料品質列為首要重點。綜合學者、研究機構的觀點 (Baker, 2014; Christer, 2015; Dumbill, 2012; Erhard & Hong, 2015; Guidera, 2013; Kowalski, 2013; NFES, 2011; Shah, 2012; Swan, 2012; Wagner & Davis, 2013)，可理出以下脈絡。

一、確認教育需求與問題

合用、適用是資料品質最基本的要求，因此確認問題與需求是確保資料品質的第一步。為避免連結、比對和分享不必要的資訊，教育決策者界定資料範圍和項目，釐清跨部門、跨區域、跨領域和跨資料庫的需求，選用合適及所需的資料，並考量學生個資的隱私、安全與保密問題，以設計適切的資訊分析系統。

NFES (2010) 認為學生的成功學習往往受到家庭、教職員、學生和體制等四類因素的影響；因此，縱貫性教育大數據必須涵蓋這四方面的資料，方能提供教育研究者豐富完整的研究資訊。但教育大數據亦非萬靈丹，任何研究者很容易從中找線索，但未必可形成具體行動方案。Christer (2015) 建議採用大數據前，先決定想處理或解決什麼問題、形塑研究問題意識。Dumbill (2012) 也發現大數據的實務工作者，大約花 80% 的時間處理原始資料，因此建置前必須慎重考量資料的適切性，是否符合教育決策與工作者的需求與問題。

Rios-Aguilar (2015) 認為身為教育研究者，得先從解決教育實務的觀點思考，並非從技術角度來切入，避免陷入徒求工具理性之窘境。是故教育大數據的分析其關注重點在於能否發現好的研究議題、資料來源的有效性，以及是否具備適當的研究方法，以創造出符合學校需求的研究結果。

二、建立資料治理架構與流程

NFES (2011) 將治理定義為「一項組織流程，也是一個架構；在提升資料品質的前提下，其建立資料責任，組織與計畫有關領域的工作成員協同合作地和持續地透過有系統地設立和強化政策、角色、責任和流程。」

Guidera (2013) 研究發現進行資料品保時，可能會遇見四項困難，分別為：勢力範圍 (Turf) — 教育界的文化和結構尚具傳統思維；信任 (Trust) — 對資料品質和應用心存疑慮，早期的資料品質保證係用來當作績效之評判，而非指引持續改善政策與制度的依據；技術議題 (Technical issues) — 一部分技術問題仍待克服，在各類解決方案如雨後春筍般湧現之際，要落實尚需主管機關的領導和政策支持；時間 (Time) — 需明訂建構縱貫性資料庫需求的優先順序。因此，政府需以政策建立一套資料治理的架構。

根據 Kowalski (2013) 的報告顯示 DQC 對資料治理建議：建立對的結構、選擇對的人、賦予該結構下決策及執行的權力、確保結構是永續的、確保美國各州教育願景得以持續 (Shah, 2012)。NFES 的 P-20W 資料治理模式 (Kowalski, 2013; Shah, 2012)，如圖 2。

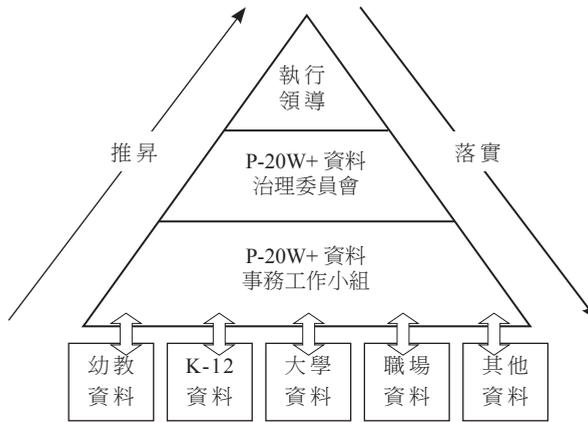


圖 2 NFES 的 P-20W 資料治理模式

資料來源：Kowalski (2013). *P-20W Data Government: Critical Roles for Success*. p.6.

政策制訂者的領導治理是這項結構的關鍵，資料治理要成功必須有策略性方向與權力。其次，DQC 指出資料系統彼此之間的聯繫雖至關重要，但徒有資料仍不足以滿足政策需要。使相關機構或部門主管的參與達最佳化的正式治理結構和流程，才是引導 P-20W 資料連結，以及建立資料管理責任績效的關鍵。當時仍有許多州的資料治理結構缺乏執行階層的政策制訂者有系統地督導並克服一些障礙。資料治理不僅須要明確的結構和流程，政策決定者的領導和堅持才是關鍵（Shah, 2012）。

在資料治理過程中，中央與地方政府必須協作。就美國為例，各州政府的角色是統籌領導、政策制訂和提供資源，州政府要做的是聚焦地方教育現場的需求、需要分析的教育資料，及適度的下放權力。此外跨州立機構成立的治理辦公室（Governor's Office），及獲得地方的支持亦是重要的一環。

鑑於有些執行狀況較佳的地區已建立適切的資料庫、標竿評量，並積極與當地教育者協同合作，分析學生的學習進步狀況，評核學校和學生的學習表現。所以 Guidera（2013）建議州與地方政府彼此應形成夥伴關係，將資料投資極大化，降低成本和負擔；並確保跨區和跨州的相容性，未來資料系統可相互介接，以建構無縫式的資料傳輸與分享系統。同時資

料要符合所有教育工作者的需求，並適度賦權給地方政府，讓資料流（data flow）流向州政府，而可行動的資訊流（information flow）流回地方政府。

三、資訊系統連結、資料配對及障礙排除

（一）資訊系統連結與資料配對

縱貫性數據的資料可在前述資料治理的規範下，讓教育研究者能按制訂的規則和規格，匯出資料並按時將資料傳送至特定資料庫，讓資訊系統更具備相互操作性的原則。

P-20W 的作法是由州政府規劃 P-20W 縱貫性資料系統，並發送至各區機關和學校，各機關和學校依據規定時程和格式上傳資料，當州政府收到各單位上傳的資料時，先做資料核對並驗證資料，若確認資料無誤後則編碼 P-20W 的 ID 至資料來源檔，將資料去辨識化後，再行將資料上傳至 LDS 資料庫，同時也刪除來源資料檔；若資料經核對並不吻合，或無法通過驗證，則進行其他補救措施，以確保資料的正確性和完整性。

美國教育研究者部分數據會取用來自動態即時像網站行為和社交互動等的數位資料，及各類傳統資訊系統的數位資料，包括各種跨結構化、非結構化和多結構化資料，來補 P-20W 資料的不足。隨著數位瓦解（digital disruption）轉變溝通和互動的管道，以及教育單位不斷強化學生跨設備、跨 web 屬性、面對面的互動和社交平臺的技術，多結構化資料等趨勢將會持續發展，但不宜與正規的資料相互混雜。

（二）資料障礙排除

系統所蒐集資料格式及內容往往不同，資料經常不一致、遺漏、重複或錯誤，因此，往往需要做資料清理，清理作業的複雜程度也隨著來源是單一或多方而有所差異。參考 Erhard 和 Hong（2015）的見解，資料清理的方法通常包括：資料分析、轉換工作流程和鏡射規則的定義、規則驗證、資料轉換、已清資料返流（backflow of cleaned data）。在資料分析前須先取得後設資料，在分析過程中將有問題的資料挑出。並依據資料異質和混亂的程度決定資料轉換的工作流程和鏡射規則。資料轉換工作流程的正確性和有效性，以及轉換的定義均應進行測試和評核，藉由 ETL 工作流程，

將資料上傳和更新資料倉儲。當錯誤被移除後，資料來源會更為清晰、純淨，確保教育工作者獲取正確資料。資料清理的工作既專業又繁雜，因此從文獻上可以觀察到美國多數學校將此工作委由廠商協助。

四、制訂學生個資隱私、安全與保密的策略

若教育大數據中學生資料的蒐集過程或應用無法完全保證學生個資的隱私與安全，縱使末端資料的品質極佳，也將無法成功蒐集與應用。因此學生資料的隱私、安全與保密是高品質資料最基本的要求，美國制訂個人資料及隱私保護的相關法規，防制全面蒐集長期儲存個人資訊，特別是管控資料的蒐集、被蒐集者的知情同意，以及賦予直接擷取、分析的公司或機構更重的法律責任（Mayer-Schönberger & Cukier, 2013）。

根據Guidera（2013）的報告指出P-20W的推動有三種配套措施：（一）符合道德與法律義務，尊重學生個人可辨識資訊的隱私與保密性；（二）降低資料誤用的各項風險；（三）確保使用者角色與義務的明確性，包括分享資料的權限、資料分享的格式、資料詳細度、保護的對象等。而且教育當局應朝善盡以下三項責任而努力，以保護學生個資的隱私、秘密和安全：

1. 建立資料管理者的角色：

定義並溝通資料的決策、管理和安全的權限、績效責任。P-20W的資料來自各級地方政府，及相關單位所提供的教育資料及職場資料，這些資料集中儲存於州層級的資料系統，且經過加密保護。教育當局決策者、合格授權機關，或研究者欲想擷取與政策、研究相關的資料，所得資料必須是去身份辨識化的（de-identified），教育現場工作者或社會大眾，僅能擷取經處理過後呈現在報表上的綜合資料，獲取教育大數據的分析數據，但絕無個人身份辨識資料。

2. 確保政策文件化、透明與落實：將與資料管理相關的法律、政策和決策，製作成文件，並讓相關利害關係人，包括機構員工、學生、父母親和社會大眾可擷取的方式溝通與頒佈相關政策和程序。

3. 強化專責單位的人力與結構：確保各州有推動和維持這些政策和程序的能力，包括人力培育和技術系統基本結構。唯有當監督、管理和推動這些政策的專責單位力量足夠，保護學生個資才更有效能。

根據北卡羅蘭納商業司 2015 年的報告，該州 Common Follow-up System [CFS] 的資料是存放在資訊科技系統的主機和商務處的本地區域網路（LAN），依照資源存取控制設施安全協定，區域網路端則建立個別使用者資料擷取記錄檔，以及個別使用者的 ID 和密碼。使用者在登入系統時和發佈任何一筆資料時，系統會告知資料的保密等級和使用上的法律規定，讓其了解個資的重要性。

五、將數據分析資料轉化為行動方案與應用

資料品質的良窳需要經過不斷地萃煉，在前述的品質保證行動後，實際成效如何則在開始取得分析資料後便逐見真章，能否產出可據以行動的資訊，將是資料品質最終的驗證。回饋在品質管理活動中是很重要的一環，藉由回饋做必要的調整、修正和補強，是資料品質提升的關鍵契機；因此，將資料轉換成有用的資訊需要創新的分析技能和資訊使用者持續的回饋建議，以確認需求和分析方向是否適切。

NFES（2011）建議各州教育當局考慮擬定一份年度研究時程表與報告，確定所做的分析研究確實是符合教育決策與實務工作者資訊需求或教育現場問題，並朝以確保各區可取得適當的分析資源；並將比對資料紀錄的效率和最佳化，以產出分析用的資料集；以及產出可跨州比較的資訊指標和分析報告。報告的形式可分：（一）診斷式報告：提供學生能改善學習成果的學習歷程資訊；（二）預測式報告：呈現學生不同時期學習經驗與成果之間的關係，如高中成績和測驗分數與大學就讀科系、課程選擇樣態和成績的關係；（三）預警式報告：依據研究所確認的預測指標，探究學習低落甚至可能流失的學生；（四）回饋式報告：在學生畢業後，提供個人學習成果的相關資訊，協助學生生涯規劃與發展。

然而要有豐碩的研究和應用，除教育部門的投入並產出教育政策決策所需報告外，尚賴更多的教育研究者以更個人化的教育資料探勘（Educational Data Mining [EDM]），產出可為教學發展依據的證據，結合相關報告相互參照比對並整合剖析，提供更深層的洞察，讓政策的決定與教學實務有更為具體的資訊做為依據，達到相輔相成的效果。目前美國常見的探勘模式有預測、結構探究、關係探勘、與模式發覺（Discovery with Models）等，其下又包含多種資料探勘技術，如決策樹、隨機森

林、貝氏知識追蹤演算法 (Bayesian Knowledge Tracing) 和績效因素分析 (Performance Factors Analysis)、社會網絡分析、領域結構發覺 (domain structure discovery)、關連規則探勘 (association rule mining)、相關探勘 (correlation mining)、時序樣態探勘 (sequential pattern mining)、肇因資料探勘 (casual data mining) 等 (Baker & Yacef, 2009; Pavlik Jr., Cen & Koedinger, 2009)。Swan (2012) 指出 EDM 技術在美國已相當普及且成果豐碩，例如 WCET 於 2012 年執行一項由蓋茨基金會資助的大數據與預測分析報告 (Predictive Analytics Reporting [PAR]) 框架的研究，此框架是一項縱貫性資料探勘專案，目的是應用大數據技術，整合跨校多元的高等教育機構的學生紀錄，並就此大量蒐集的學生學習歷程紀錄進行預測分析，以瞭解影響學生線上課程學位的持續就學和進步狀況，並據此檢視大數據技術對教育研究的意涵與效能 (Wagner and Davis, 2013)。Baker 與 Inventado (2016) 指出 EDM 社群已開發適合教育資料獨有特質的自動化或半自動化方法，處理來自教育軟體、遠距學習課程、課程管理系統、州和國家教育資料庫等多元的教育資料。有些 EDM 在教育研究領域已被證實是一項有助於教育決策及改善教學實務問題的分析技術，包括研究是哪些因素導致學生流失退學、了解某學習領域的知識結構、探究學生學習投入狀況、預測哪些學生準備繼續升學及了解線上學習學生分組做功課時的樣態等，此分析技術開始被廣泛運用。換言之，協助教育決策及改善教學實務為 EDM 在教育研究領域之最終目的，以下針對教育決策以及改善教學實務分舉一例陳述。

在教育決策的實例，Gane、Okoroh 與 DiBello (2015) 使用個案研究法，從大數據中探索和萃取有關教師使用 Diagnoser 教學工具軟體和學生成績的資訊，Diagnoser 是一套診斷式教學工具軟體，用以評量國高中學生物理學習成就。他們運用個案研究法和視覺化方法評估 Diagnoser 評量的成效。該研究蒐集了 2004 至 2011 年的 194,000 筆學生的物理作業紀錄，從分析學生的思考面向去驗證 Diagnoser 的成效。在應用教育大數據進行研究後，指出此教學工具軟體確實有助教學診斷工作，研究者將此結果提供給教育主管機關參考，做為未來持續推廣 Diagnoser、資源挹注的依據。

另在改善教學實務方面，也有許多教育工作者正應用 EDM 多元模式，從事各種有利於學生發展的研究。以 Shelley、Dasgupta、Moher 與 Lyon

(2014) 的研究為例，其針對五年級學生進行支援學習者從大影像數據建構對動物行為理解的研究，教導學生如何應用錄影和錄音設備蒐集動物行為的資料，整個過程共分為四階段，首先是觀摩其他學校作品，其次是觀看教師拍攝的校園景象，判定哪些環境可能效果最佳，第三階段則是要學生以自家後院為場域錄影錄音，回答「在所住社區，有那些多元性的物種？」的問題，最後，則要學生設計他們自己的研究問題，並應用攝影機蒐集資料回答之。過程中，研究團隊給予學生必要的指導，並協助他們將照片上傳至 PhotoMAT，供學生從這些影像資料萃取出資訊。

最後在教學實踐研究方面，為了鼓勵社會各界活用 LDS，由州政府出題，讓教育研究者應用教育大數據來解題。並制訂資料申請程序，以為有興趣者遵守。例如 Virginia 州政府提出六大議題，共 38 個研究題目。也制訂了資料申請辦法，透過 5 個步驟，教育研究者便可取得所需大數據。這項措施已經產出 9 篇研究，正在進行中的有 3 篇（Virginia Longitudinal Data System, 2018）。

藉由政府部門資源及工具的提供，配合教育研究者的共同努力來豐富研究成果，使教育大數據的建置效益得以彰顯。

六、確立資料產出時效性及以實務角色為本的應用模式

Guidera (2013) 勉勵各州政府，教育資料不應僅是做教育當局討論之用而已，還須將資訊客製化和提供適切的分析報告給教育決策、教育現場進行參考應用，落實以實務角色為本（role-based）的理念。並且建議應做到資料透明化，製作容易擷取學校和學生表現的資料，提供關於教育體系績效、成就和挑戰的背景脈絡資訊，協助家長更了解並支持子女的教育，並提升社會大眾認同度。

根據 DQC (2013) 的報告顯示，美國 Oregon 和 Kentucky 州是應用大數據資料提升學生成就的兩個成功典範，尤其是 Oregon 州允許教師擷取各項教學數據資料，並給予應用資料的教育訓練，讓教師判讀及運用教育大數據時，更能駕輕就熟，教師受過這項資料分析應用訓練的學生，其州考試的成績比其未受訓練的進步幅度較大；Kentucky 州則產出清晰的學生學習報表，將學生在大學的表现歷程資訊，提供給高中，提升高中學生生涯輔導效能。也因兩州對教育大數據分析模式的重視與運用，讓其學生

學習成果獲得明顯提升。

Oregon州推動推動的是「直取成就」(Direct Access to Achievement [DATA])的專案，提供教師直接擷取高品質教學資料的途徑，並使用資料充實教學決策。從2007年起至2013年止，這個專案總共訓練了將近5,000名教學者，讓教師樂於在課堂中應用資料，提升學生的學習成就。Oregon州的策略是投資在教師與學校領導者身上，授予擷取學生學習資料的權限，提供他們周全且與職務相關的教育訓練，並教導在教室如何應用資料進行決策，目的則是回應其教育現場的需求與問題，亦改變他們對資料對影響學生成就的價值的認知。其具體作法為評估教學現場的需求，依據其需求與問題開發訓練模式，建立指導團隊做為種子教師，並進行資料分析的教育訓練。

七、培養教育決策與工作者使用資訊的能力

Guidera (2013) 研究發現，雖然各州都投入經費連結 P-20W 資料系統，但就分享合適且有限的關鍵資料所需的權限，尚缺乏統籌管理結構；各州皆使用縱貫性大數據資料產出報表，並用控制儀表板降低閱讀門檻，但就確保使用者具備擷取資料的能力上，則尚有改善空間。這顯示當走過建置期後，進入應用期前仍需要持續的努力。

大數據動輒以 tera-bytes 計，有效的資料視覺化有助於教育決策與工作者更容易分析和瞭解教育研究產出的各項資訊和事實，將複雜的資料變得更容易理解和應用。依據研究目的和使用者的不同，研究者將選用或設計各式各樣的圖表或圖示呈現研究分析結果，教育決策與工作者則必須具備認識、理解和應用大數據研究分析的能力，方能在教育決策或教育現場增添教學利器。

八、其他配套措施

一個具支持性和穩定的環境是優質資料的保障，大數據的蒐集和應用需要政策的引導、支持，及排除可能的疑慮和障礙。其定位將隨著國家對精進學生學習成果的重視程度而異，除政策面的宣示、組織的架構和經費的支持以外，在執行面仍需要解決各項實務問題，例如，美國教育部於2010年專案委託各州學校主管委員會(Council of Chief State School

Officers [CCSSO]) (CCSSO, 2015) 協調和建構全國教育資料模式 (National Education Data Model [NEDM])，成為 P-20W 的資料來源和標準化核心教育資料集的執行參考依據，因此了解政策訂定者、教育者、研究者、資料管理者和委外廠商們在教學、學習和行政管理系統的資訊需求，建置一份常用資料集的說明書，並根據現有標準，建置一個開放性的架構，做為建立教育資料系統的依據，如標準的資料定義、程式編碼策略、資料元素標準、資料集關係描述、最佳實務範例。藉由訂定一致性的資料定義和體系結構，確定哪些資料有效，建構明確邏輯和實體資料模型的內容與結構，以提升資料品質和相互操作性，是在前述七項品質確保策略外，配套措施亦值得重視。

綜合上述，美國政府為了解縱貫性教育大數據系統 P-20W 的執行狀況與成效，在 2016 年進行了一次全國性調查，檢視 LDS 的發展與應用並發表了一篇政策報告，目的是探索 50 個州及華盛頓 D.C. 的 LDS 的發展與應用狀況，包括連結教育資料的益處與障礙，此外並以三州做個案研究。從這篇報告可以看出這些地區都已具備資料能在系統彼此之間關聯互通的能力，其中有 47 個州和 DC 有獲得聯邦政府的補助、37 個州和華盛頓 D.C. 已經能連結到至少 2 個州級政府單位的資料、有 16 個州和華盛頓 D.C. 都已完整建置好 P-20W 系統 (Zeke, 2016)，此外 Zeke (2017) 研究指出美國可持續推動縱貫性資料庫的關鍵在於是否有穩定多元的經費挹注、人事面的異動程度，包含領導者、部門角色，可能導致無法有效維護和監督資料系統；最重要的則是資料連結的安全性、隱私是否受到完全保障。由此顯見美國政府單位及教育現場對縱貫性教育資料庫的重視與落實。

伍、臺、美兩國縱貫性教育大數據策略之比較與借鏡

綜觀前述我國六種教育資料庫，涵蓋全國國小至大專畢業學生，且無不以瞭解及改善學生學習成效為共同目的，由於所具教育目標不同，其蒐集、儲存及應用資料的範圍、項目及方式也各具特色。臺、美兩國縱貫性教育資料的建置，雖然著眼處都是為滿足不同利害關係人需求、最終目的

都是為協助學生成功學習和就業，但在策略及具體行動上，兩國彼此之間尚存在極大差異，分述如下。

一、對象及規模

P-20W 以州為單位，為所有學生建置從學前至就業的個人化縱貫性的學習歷程資料，形成全國性的教育大數據庫。臺灣的 TEPS、TASA、TIHED 與 TUSD 主要也是以所有學生為對象，但因目標不同，其資料的蒐集除 TUSD 採普查外，其餘則多以抽樣方式，對象隨抽樣而變化，人數規模相對較小，較難以在同一系統下以個人化資料進行縱貫性的研究追蹤。縱以高級中等教育階段學生學習歷程檔案而言，其涵蓋年級亦與之相去甚多。

二、資料內容及應用

P-20W 是依據國家政策及共同標準於每一州建置一個資料倉儲，透過統一辨識碼，連結幼教、高等教育、職業訓練、就業等資料庫並持續更新，再按照共享式學習基礎結構，提供教育決策或工作者進行分析研究，產出研究結果，讓州教育廳以學習管理系統產出管理報表，而教育研究者則產出各州或跨州的研究報告，並據以提出具體行動方案。它是以各州區域性教育或學校機構的校務或學生學習紀錄之既有資料為基礎，做蒐集、儲存及應用。相較於 P-20W，我國各類學生資料隨當初建置的目的而分佈在各大型教育資料庫，資料內容與應用也必以原創目的為依歸，若要進行跨資料庫的應用，自須更多的前置性處理。

三、計畫永續性

美國是以戰略性的位階，以立法及行政資源支持這個計畫，創造有利環境、培育教育大數據人才、鼓勵應用數據進行研究，無論在人事、經費，以及資料安全性、隱私性皆相當重視且投入。而前述我國的 TEPS、TASA 或 TIHED 等大型計畫往往隨著政策變遷或其他因素而中斷，誠為可惜。

陸、結論與建議

一、結論

（一）美國建置國家級縱貫性教育大數據符合國家發展願景及具戰略性目的

解決橫斷性分析無法提供個別學生進步及增值狀況的資訊的不足，藉由品質更好、內容更豐富的資料，加上更深入的問題的探究、找出更具有內涵的解答。翻轉過往除了監控學生成績和對傳統匯總數據做出反應外，政策制定者和教育工作者可以更加主動地使用 LDS 數據去識別趨勢、預測結果，就政策，管理和教學策略做出更明智的決策；並更有效地利用資源，採用最有效的教學策略來滿足學生的個人需求。儘早發現潛在問題，並瞭解能做些什麼努力來促進學生成功。

（二）建置國家級縱貫性教育大數據時，應具備戰略思維及行動方案規劃

首先是政府要有戰略性的目標，以立法及行政資源來推動。在推動前要制訂資料戰略及具體的行動方案，帶領全國教育界及相關單位朝該戰略性目標邁進。其次，資料的基本要素必須統一，哪些資料是基本必須，哪些資料容許擴充，以為各級單位遵守及應用。在這同時，必須編輯指導手冊，完整解說 LDS 的來龍去脈及建置的要領，並大量培養大數據創新應用能力及人才，讓資料庫及系統得以順利建置，俾在建置後，有更多的人懂得如何上傳資料、管理資料和應用資料，這些都是建置 LDS 的成功關鍵要素。

（三）以全面品保的概念建置一個優質機制來確保資料品質

參考美國經驗，瞭解資料品質確保必須從確認教育需求與問題開始，建置完善清楚的資料治理結構及流程，地方及中央密切分工合作，將資訊系統無縫連結、檢核並排除在匯集各方資料的過程可能發生的障礙，制訂學生個資隱私、安全與保密的策略，做好個資隱私保護、保密及資訊安全。宣導鼓勵社會各界，在既定規範內，應用得來不易的資料進行研究，發覺

問題或改善機會，產出有價值的報告。當應用這些資料時，記得秉持以實務角色為本；為使研究結果更有價值，資料源頭的資料產出和送出必須及時。在建置及應用LDS時，各基層機構的表現會參差不齊，因此必須培養教育決策者及工作者應用資訊的能力；此外，在技術面及行政面，中央也必須給予地方必要的支援，最重要的是，能以全面品保機制強化資料庫的永續發展。

二、啟示與建議

誠如彭森明（2010）認為在支持學生成功的目標上，唯有運用長期追蹤的評量分析，方能有效了解個人或群體學生的學習成長。教育是百年樹人的大事，美國在應用教育大數據的戰略目的及努力是值得注意和參考的。我國在縱貫性教育數據的建置、管理與應用上已有多年基礎和經驗，要建置一個國家級的教育大數據應無需從零開始。盤點過我國縱貫性教育資料庫的現況後，綜合以上探討及參考上述美國經驗所提出的策略及品質保證方案，本研究提出以下建議。

- （一）以支持學生成功為策略性目標，擬定我國教育資料策略，擘劃我國整體教育大數據建設藍圖並給予充裕資源和立法支持，建設實現策略目標的環境。規劃歷程可參考美國經驗，分短、中、長期計畫，逐步規劃、設計、建置、維運、應用和調整。
- （二）在推動時，盤點並整合各教學或研究機構長期建構之橫向、縱向資料庫，橫向如勞動部、衛福部等；縱向如國教院及教育部所轄各級教學或研究機構，包含幼兒教育階段學前資料庫、中小學新生入學資料庫、各縣市校務行政系統資料、高中階段高級中等學校學習歷程資料庫到高教司大學校院校務資料庫等，串接不同階段之學生資料，以增加資料多元性，提高資料分析效能。由於高中階段高級中等學校學習歷程檔案已具備類似美國教育大數據的基本條件，故可考慮以其為基礎，向前和向後延伸，發展出具有更高潛在效益的教育數據。
- （三）將此縱貫性教育大數據建構成一個持續累積型的教育資料庫，

並具備資料品質確保的機制和工具，以持續累積優質教育資料，使資料具有永續性、可追蹤性等。

- (四) 設立大數據相關科系、學程或教學實務之教育訓練，以培育大數據技術人才，提升大數據資料處理、管理及分析技術，如資料管理、資料視覺化、感知、開放資料萃取、系統與網路及機器學習等專業領域，提升國內巨量分析人員能量與素質。
- (五) 善用各領域教育專家學者，形成跨領域的教育大數據資料探勘團隊，應用現成或新創模式探勘資料，並依據教育當局、教育現場或相關利害關係人的需求，開採出具行動價值的資訊。
- (六) 明訂資料庫使用相關法源以保障資料蒐集的必要性、正當性與永續性確實進行教育大數據資料庫的建置與推展；此外必須擬定跨部會合作計畫，以利資料整合以及充實資料的內涵與完整性；另上級政府單位必須依據證據為基礎的決策模式，了解大數據資料庫對於政策擬定與決策有所助益，強力主導與資助大數據資料庫的規劃與執行，俾利推展。
- (七) 在實務研究者的應用層面，執行大數據資料的研究者必須深度思考界定研究架構、研究問題以及可能應用的層面，以指引資料蒐集與研究設計。
- (八) 教育相關工作者，尤其是教學現場的教師，應用教育大數據於決策上，並將應用成果回饋給大數據資料庫，以確保和提升它的品質，使其得止於至善。
- (九) 建議教育主管機關及學校舉辦國際學術研討會，邀請美國執行縱貫性大數據系統專家學者與會，與國內專家學者彼此交流、分享、觀摩，期許對我國建置教育大數據資料庫有所啟示及借鏡。

綜言之，豐富、完整、正確、即時、優質的資料是優質教育研究的基本條件，其需要治理機構和流程，亦在確保學生資料隱私、安全和秘密的前提下，綿延不斷地供應研究之需。蒐集、萃取、清理、轉換、儲存、上傳、安保、分析、呈現和分享等涉及的資訊技術已然成熟，只要教育現場的需求及問題明確，我國教育當局可依據縱貫性教育大數據的建置策略，與各級地方政府及學校著手共同規劃進行這項縱貫性教育大數據的建置長程計畫，以落實以證據為本位的教育決策模式。

參考文獻

- 王金龍（2015）。銘傳 Moodle 大數據分析與學生學習成效。評鑑雙月刊，56，22-27。
- [Wang, C. L. (2015). Moodle big data mining and student learning performance in Ming Chuan University. *Evaluation Bimonthly*, 56, 22-27.]
- 張佳琳（2014年5月1日）。「大數據」趨勢對中小學教育的影響引發討論。教育部電子報，612。取自 http://epaper.edu.tw/windows.aspx?windows_sn=15157
- [Cheng, C. L. (2014, May 1). Impact of big data on primary and secondary education. *Newsletter of Ministry of Education*, 612. Retrieved from http://epaper.edu.tw/windows.aspx?windows_sn=15157]
- 國家教育研究院（2018）。「臺灣學生學習成就評量資料庫」建置計畫。取自 <https://www.naer.edu.tw/files/11-1000-1408.php?Lang=zh-tw>
- [The national academy for educational research (2018). *Taiwan Assessment of Student Achievement*. Retrieved from <https://www.naer.edu.tw/files/11-1000-1408.php?Lang=zh-tw>]
- 教育部（2018a）。高級中等學校學習歷程資料庫系統建置成果與後續推展宣導規劃。取自 <https://ws.moe.edu.tw/001/Upload/23/relfile/8059/55480/b4edd469-e9a6-4bd3-b49f-b642da1982f1.pdf>
- [Ministry of Education (2018a). *Construction results and promotion digital learning portfolio for senior high school*. Retrieved from <https://ws.moe.edu.tw/001/Upload/23/relfile/8059/55480/b4edd469-e9a6-4bd3-b49f-b642da1982f1.pdf>]
- 教育部（2018b）。國民中小學學生學籍交換規格標準 4.1 版。取自 <http://public.hlc.edu.tw/apfiles/A201811051612421695-1.pdf>
- [Ministry of Education (2018b). *The version 4.1 revision of Taiwan's common education data exchange standards*. Retrieved from <http://public.hlc.edu.tw/apfiles/A201811051612421695-1.pdf>]
- 曾元顯（2016）。校務研究資料庫的建構與分析應用。當代教育研究季刊，24（1），107-134。
- [Tseng, Y. H. (2016). Development and application of databases for institutional research and analysis. *Contemporary Educational Research Quarterly*, 24(1), 107-134]
- 彭森明（2010）。大學生學習成果評量：理論、實務與應用。臺北市：高等出版。
- [Peng, S. S. (2010). *Assessing college student learning outcomes: Theory, practices, and applications*. Taipei, Taiwan: Higher Education.]
- 彭森明（2013）。高等教育校務研究的理念與應用。臺北市：高等教育。
- [Peng, S. S. (2013). *Institutional research in higher education: Concepts and applications*. Taipei, Taiwan: Higher Education.]

- 楊孟山、林宜玄(2016)。臺灣教育資料庫的現在與未來。*臺灣教育評論月刊*, 5(7), 10-18。
- [Yang, M. S., & Lin, Y. S. (2016). The present and future of longitudinal databases for education in Taiwan. *Taiwan Educational Review Monthly*, 5(7), 10-18]
- 蔡明學、黃建翔(2015)。大數據分析在我國教育發展應用上之探討。*教育脈動*, 4, 154-164。取自 <https://pulse.naer.edu.tw/Home/PrintPdf/65dac715-1ffd-4ab9-854e-365c1feefbbc>
- [Tsai, M. H., & Huang, C. H. (2015). An investigation on the application of big data analysis in the development of education in Taiwan. *Pulse of Education*, 4, 154-164. Retrieved from <https://pulse.naer.edu.tw/Home/PrintPdf/65dac715-1ffd-4ab9-854e-365c1feefbbc>]
- Baker, R. S. (2014). Educational data mining: An advance for intelligent systems in education. *IEEE Intelligent Systems*, 29(3), 78-82.
- Baker R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3-16.
- Baker, S., & Inventado, P. S. (2016). Educational data mining and learning analytics: Potentials and possibilities for online education. In G. Veletsianos (Ed.), *Emergence and Innovation in Digital Learning* (pp. 83–98). doi:10.15215/apress/9781771991490.01
- Bernard, M. (2014a). *6 key skills every business needs*. Retrieved from <https://www.dummies.com/programming/big-data/6-key-big-data-skills-every-business-needs/>
- Bernard, M. (2014b). *Big data: The 4 layers everyone must know*. Retrieved from <http://www.slideshare.net/BernardMarr/big-data-4-layers>
- Christer, J. (2015). *Advancing analytics to predict specific needs*. Retrieved from http://www.huffingtonpost.com/christer-johnson/ibm-advancing-analytics_b_1760680.html
- Council of Chief State School Officers (2015). *About the standards*. Retrieved from <http://www.corestandards.org/about-the-standards/>
- Data Quality Campaign (2013). *Investing in educator data literacy improves student achievement*. Retrieved from <http://files.eric.ed.gov/fulltext/ED548266.pdf>
- Davenport, T. H., & Patil, D.J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 2012, 15-23.
- Dumbill, E. (2012). *What is big data?* 'O'Reilly Media. Retrieved from <http://orm-atlas2-prod.s3.amazonaws.com/pdf/e11376d1c19a651736042656f2aae705.pdf>
- Erhard, R., & Hong, H. D. (2015). *Data cleaning: Problems and current approaches*. Retrieved from https://www.betterevaluation.org/sites/default/files/data_cleaning.pdf
- Gane, B. D., Okoroh, C., Dibello, L. V., & Minstrell, J. (2015). *Making sense of big data from classroom assessments: Teacher Case studies and facets-based physics assessments*. Retrieved from https://www.academia.edu/12105549/Making_Sense_of_Big_Data_from_Classroom_Assessments_Teacher_Case_Studies_and_Facets-based_Physics_Assessments

- Guidera, A. R. (2011). *Using data to improve student achievement: A conversation with members of the montana education and local government interim committee*. Retrieved from <https://slideplayer.com/slide/3832574/>
- Guidera, A. R. (2013). *The four Ts of state data systems (turf, trust, technology, and time): Policy perspective on empowering education stakeholders with data*. Retrieved from https://drive.google.com/open?id=1WXfGbGnIHCLbMwHY1hxCvjOFq_4yAj2n
- Kowalski, P. (2013). *P-20W data governance: Critical roles for success, education, commission of the states*. Retrieved from <https://drive.google.com/file/d/128U9qswrplu4gfUmwa-qm-MnUyviXLaA/view?usp=sharing>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston, MA: Houghton Mifflin Harcourt.
- Mayer-Schönberger, V., & Cukier, K. (2014). *Learning with big data: The future of education*. Boston, MA: Houghton Mifflin Harcourt.
- Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42(1), 30-37.
- National Forum on Education Statistics (2010). *Traveling through time: The forum guide to longitudinal data systems. Book one of four: What is an LDS?* (NFES 2010-8015). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- National Forum on Education Statistics (2011). *Traveling through time: The forum guide to longitudinal data systems. Book three of four: Effectively managing LDS data* (NFES 2011-805). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- North Carolina Department of Commerce (2014). *A Report on the operations of the north carolina common follow-up system*. Retrieved from <https://docplayer.net/16538676-A-report-on-the-operations-of-the-north-carolina-common-follow-up-system-may-2014.html>
- Pavlik Jr. P. I., Cen H., & Koedinger, K. R. (2009). *Performance factors analysis – A new alternative to knowledge tracing*. 14th International Conference on Artificial Intelligence in Education. Brighton, England. Retrieved from <http://files.eric.ed.gov/fulltext/ED506305.pdf>
- Rios-Aguilar, C. (2015). Using big (and critical) data to unmask inequities in community colleges. *New Directions for Institutional Research*, 163, 43-57.
- Shah, R. (2012). *Pivotal role of policymakers as leaders of P-20/Workforce data governance*. Retrieved from <https://files.eric.ed.gov/fulltext/ED538830.pdf>
- Shelley, T. R., Dasgupta, C., Moher, T., & Lyon, L. (2014). *Supporting learners' construction of understandings of animal behaviors from large image sets*. Retrieved from https://drive.google.com/open?id=12MzcakTmgvIGJW_5QGfsqkj2SCbTNYF
- Swan, K. P. (2012). *Big data and the predictive analytics reporting (PAR) framework*. Retrieved from https://drive.google.com/open?id=1rS4Eggr_e1H3Cy0VLTW6BAee-NpTuy99

- U.S. Department of Education (2002). *Education sciences reform act of 2002*. Retrieved from <https://legcounsel.house.gov/Comps/Education%20Sciences%20Reform%20Act%20Of%202002.pdf>
- U.S. Department of Education (2015). *Strategic plan for fiscal years 2014-2018*. Retrieved from <http://www2.ed.gov/about/reports/strat/plan2014-18/strategic-plan.pdf>
- Virginia Longitudinal Data System. (2018). *Insights of VLDS*. Retrieved from <https://vlds.virginia.gov/insights>
- Wagner E., & Davis B. (2013). *The predictive analytics reporting (PAR) framework, WCET*. Retrieved from <http://er.educause.edu/articles/2013/12/the-predictive-analytics-reporting-par-framework-wcet>
- Zeke, P. J. (2016). *50-State comparison: Statewide longitudinal data systems*. Retrieved from <https://www.ecs.org/state-longitudinal-data-systems/>
- Zeke, P. J. (2017). *Examining SLDS development and utility*. Retrieved from <https://www.ecs.org/examining-slds-development-and-utility/>