

【附件三】教育部教學實踐研究計畫成果報告格式(系統端上傳 PDF 檔)

教育部教學實踐研究計畫成果報告(封面)

Project Report for MOE Teaching Practice Research Program (Cover Page)

計畫編號/Project Number：PBM1090784

學門專案分類/Division：商業及管理

執行期間/Funding Period：2020/08/01 – 2021/07/31

(計畫名稱/Title of the Project)

**以學科概念之高維度語意空間結構演變
量測經濟學的學習成效**

Measuring Learning Outcomes in Economics through Changes in the High Dimensional
Spatial Structure of Terminologies within an Academic Domain

(配合課程名稱/Course Name)

經濟學
Economics

計畫主持人(Principal Investigator)：鄭保志

共同主持人(Co-Principal Investigator)：張智宏

執行機構及系所(Institution/Department/Program)：國立中央大學經濟學系

成果報告公開日期：

☒ 立即公開 ☐ 延後公開(統一於 2023 年 9 月 30 日公開)

繳交報告日期(Report Submission Date)：2021/09/17

以學科概念之高維度語意空間結構演變 量測經濟學的學習成效

Measuring Learning Outcomes in Economics through Changes in the High Dimensional Spatial Structure of Terminologies within an Academic Domain

一. 報告內文(Content)

1. 研究動機與目的(Research Motive and Purpose)

我跟許多經濟學界的朋友都有一些共同的經驗。

喝喜酒時同桌不認識的賓客在聽說自己是經濟學家以後，通常會問「你覺得哪些股票可以持有」、「台灣明年的景氣會如何」等類型的問題。

在審查申請入學生的書面資料時，常常會看到學生寫道：「因為媽媽擔任公司會計，所以我從小就對經濟學很有興趣…」、「我如果能夠錄取貴系，一定會利用開學前的時間好好閱讀財訊（商業周刊）…」。

這些不全然是錯誤認知，但有些刻板印象與我們所認識的《經濟學》落差很大！

研究動機

一般素人眼中所見的經濟學（心理學、哲學…）與受過專業訓練的經濟學家（心理學、哲學家…）不同，這當然是一件再正常不過的事情，而素人與專家對於學科本體的認知差距，必然會反映在雙方對於詞語（形容詞或名詞）或陳述句的使用上。舉例來說，面對以下這個問題

(可複選) 下列詞語可以良好地描述我對《經濟學》的印象

- ☐ 有趣 ☐ 艱澀 ☐ 抽象 ☐ 很能解釋社會現象 ☐ 很多數學
☐ 圖形複雜 ☐ 雜亂無章 ☐ 貨幣金融 ☐ 日常生活處處可見
☐ 無聊 ☐ 投資理財 ☐ 很有系統性 ☐ 與實際社會脫節 ☐ 其他

時，素人與經濟學家所選擇的詞語群組可能大相逕庭。

又比如說，在知名《經濟學》教科書中，Gregory Mankiw 提出廣為經濟學者所接受的經濟學十大原理，包括：取捨、機會成本、理性人進行邊際思考、誘因、交易互惠、市場通常是組織經濟活動的好方法、政府有時可以改善市場結果、生產力決定生活水準、過多的貨幣發行會造成通貨膨脹、通膨與失業間的短期取捨等。而台大經濟系王道一教授與幾位學界友人則整理出「鄉民經濟學十大原理」，很有笑果地指出素人與專家之間的差距：

1. 價格只能反映成本。	6. 每個人都應該有豐厚的月退休俸（從公務員開始）。
2. 努力決定薪水。	7. 許多產業都很神聖，絕對不能商品化。
3. 政府的背後金主來自火星。	8. 教育只是一個認證，不會增加智識。
4. 如果市場失靈，一定是政府沒管好。	9. 貨幣貶值可以促進經濟成長。
5. 如果政府失靈，表示經濟學家沒做好。	10. 隱匿資訊才能避免不必要的恐慌。

問題是我們有沒有方法來量測這詞語使用上的不同程度有多大？在經過專業學科的學習之後能縮小這個差異嗎？如果以上的答案都是肯定的，那麼這套創新的量測模式是否可以作為傳統工具（例如學生自我效能評估、滿意度問卷、試題測驗成績等）外另一種檢視學生學習成效的方式呢？它跟傳統量測工具之間的關聯又是如何呢？

傳統的考試題目答對得愈多就代表愈有學習成效嗎？我們如何知道學生答對試題是猜的、背的還是真正懂了？我們又如何知道學生得到的是片段的還是融會貫通的知識呢？有些試題說不定連大學教授在沒有準備的情況下也無法立即得出正確解答，但我們難道會認為這些教授不具有高層次的知識能力嗎？透過詞語量測的評估結果是否比傳統工具更加穩定呢？

研究目的

傳統教學現場採用的選擇、填充和申論等考試方法，雖然能概要地評估學生對教師所設定的特定知識主題的記憶與理解程度，卻難以窺見學生在該領域的整體知識表徵，也缺乏有系統比較其知識結構表徵在學習初期與晚期如何變遷，遑論將大量的學生和專家學者放在同一套評估方式下量化比較其對於某些課程理解方式有何不同。

本研究計畫企圖以融合心理學實驗與大數據分析的創新方法，來回答「素人與專家對於經濟學的關鍵概念詞彙認知表徵有何差異」與「素人在經過經濟學專業學科學習之後，其對於關鍵概念詞彙之認知表徵有何改變」這兩個問題。將 Q-sort、自然語言處理 (Natural Language Processing, NLP)、推薦系統 (recommender system) 等行為或大數據分析技術，應用於本計畫所收集之學科概念詞彙判斷資料，並建構學科概念於學習過程中轉變以及專家與初學者差異之模型，以作為一套檢視學生學習成效的新工具。

在發展出這套新工具之後，我們將之應用在正式課程中來蒐集相關數據，依照各班級所採用的教學模式（遊戲或單向講授）、遊戲內容，分為多個實驗組與對照組，對象以大學生為主，搭配傳統量測工具所得到的結果，更進一步探討這些工具之間的關聯性，並對新工具的實用價值有更為完整的評估。

如果能夠透過本計畫在《經濟學》這門學科上得到明確的成果，未來可將這套「高維度語意空間結構演變」的創新量測模式推廣到其他學科，對於同一學科內的各教學實踐研究計畫而言，便有了一套共通且具可比性的成效評量工具。

2. 文獻探討(Literature Review)

對專業學科知識的重要概念詞彙結構表徵，理論上應當隨著個體的經驗與學習而動態改變，才能對應到個體在各種領域知識理解與應用的表現進步上。本計畫的重要研究方法之一，即為運用行為實驗與《計算語言學》的方法來量化個體在《經濟學》領域的概念詞彙結構，進而能探討其個別差異與學習造成的變化。我們結合《發展與人格社會心理學》中常用的 Q-sort 技術，以及《計算語言學》中的潛在語意分析 (Latent Semantic Analysis, LSA)，嘗試建構兼具考慮「個體知識個別差異」以及「概念詞彙意涵一般性」

的語意向量空間，藉以量化地比較素人與專家在《經濟學》重要概念詞彙或陳述的語意空間結構有何差異，以及學生在專業學科學習的初期與晚期（例如修習《經濟學》課程前後），其學科的語意空間結構變化。以下簡要介紹本計畫所使用的兩個研究方法，「Q-sort」和「高維度語意空間」，特別著重於其應用在探討概念表徵與其改變的面向上。

概念表徵 (concept representation)

人類的概念結構是個錯綜複雜的網路：沒有任何知識概念孤立存在於人類的認知系統中，所有的概念都與其相關概念共同被表徵和處理，而形成概念的語意網路。當個體閱讀、聽聞或自發地想到某個學習過的概念詞彙時，會增益與其有連結之語意網路在長期記憶中的活化程度，致使位於網路節點之其他詞彙被辨識和分類的效能有所改變。

在典型的語意促發典範 (semantic priming) 實驗典範裡，受試者對電腦螢幕上每次呈現的詞彙作出是否屬於某類別的判斷（動物/植物；人造物/自然物），而時間順序上相鄰出現的詞彙被設計為兩兩一組，先出現者為「促發詞」(prime word)，後出現者為「目標詞」(target word)；當「促發詞」與「目標詞」彼此高度相關時（醫生→針筒），受試者對「目標詞」判斷的反應速度會比兩者不相關時（警察→果凍）來得快。上述語意促發效果是支持人類認知記憶系統中詞彙語意依其相關程度被表徵的證據 (Quillian, 1967; Collins and Quillian, 1969; Rosch, 1973, 1975; Smith et al., 1974)。

Q-sort 技術及其在概念表徵變化上的應用

Q-sort 技術是一種將多個概念依某種條件維度進行排序的行為實驗方法 (Brown, 1980)；每一位參與實驗的受試者（通常 40-60 人，稱之為 P-set）依據指導語定義的向度（喜好程度、與自己相似程度、重要性...等），分別將一群印制在卡片上的概念或陳述句（通常為 30-60 個，稱之為 Q-set），依序分成 5 或 7 堆放入一個金字塔狀的 Q-table 中（參見圖 1），其中每一堆的比例經特別安排近似於常態分佈；所有詞語的順序分數 (rank score) 形成一個「個體 x 概念」的矩陣，而能計算概念或個體之間兩兩相關性之相關矩陣，再進一步藉由因素分析 (factor analysis)、類聚分析 (cluster analysis) 或其他多變項分析方法萃取出個體或概念的子群組。這一整套運用 Q-sort 的實驗程序、刺激材料設計、和資料分析方法，又合稱為 Q 方法學 (Q-methodology)。

不同意				同意				
-4	-3	-2	-1	0	1	2	3	4
非常			還好		還好			非常
(2)	(3)	(4)	(5)	(6)	(5)	(4)	(3)	(2)
1	3	6	10	15	21	26	30	33
2	4	7	11	16	22	27	31	34
	5	8	12	17	23	28	32	
		9	13	18	24	29		
			14	19	25			
				20				

圖 1：一個假想登錄有 Q-sort 完成結果的 Q-table

改編自 He, Hu, & Fan, 2017；Encyclopedia of Personality and Individual Differences

Nagasima (1961) 讓接受心理治療的受試者將一堆關於自己的描述依「像我」到「不像我」的程度以 Q-sort 方法分堆，然後計算不同向度判斷之間的相關性（父親眼中的我、母親眼中的我、朋友眼中的我、理想的我）；受試者在心理治療前後各作一輪，然後看這些不同向度判斷間相關程度的改變。結果發現心理治療組的向度間相關在治療後有提升，而控制組則否。高程度的向度間相關代表各種向度的自我比較一致，代表一個人內在的整合程度較高、較少內在衝突；相關由低變高反應了心理治療的效果。Rogers (1951; 1958) 等一系列研究亦有相同的發現。

高維度語意空間

高維度語意空間 (High dimensional semantic space; Lund & Burgess, 1996) 是《計算語言學》與《認知心理學》中相當廣為研究者使用的語言表徵模型。透過分析大規模文本語料庫（例如 Google 的搜尋結果、中研院平衡語料庫、報章雜誌等），計算各個詞彙在特定共現窗口字數 (word width window) 條件下，在所有語料內容中與其他詞彙的共現關係值(=窗口寬度-間隔字數+1)，構成一個共現向量 (co-occurrence vector)；所有詞彙的共現向量構成一個共現矩陣，也可稱為這些詞彙的高維度語意空間。表 1 是以五個詞為寬度的共現窗口套用在「時下最夯的拍照打卡景點」這句話時，所得到的共現矩陣；矩陣中每一個值代表其對應欄與例的標題詞彙間之「共現強度」(strength of co-occurrence)，越大的值關係越強。每個欄的值代表該欄標題的詞彙在五個字的分析窗口中與文本中較其晚出現的詞彙（各列標題）之順序先後關係；例如「拍照」出現於「時下」之後第四個位置，故在「時下」這一欄中，「拍照」的共現值 = 窗寬(5) - 順序(4) + 1 = 2。相對的，每個列的值代表該列標題的詞彙在五個字的分析窗口中與文本中較其早出現的詞彙（各欄標題）之順序先後關係；例如「拍照」出現在「景點」之前第二個位置，故「景點」這一系列中，「拍照」的共現強度值 = 5 - 2 + 1 = 4 (Lund & Burgess, 1996)。以前述對單一句子的共現矩陣定義方式建構高維度語意空間的流程概要如下：

表 1：一個簡單的共現矩陣實例。

	時下	最	夯	的	拍照	打卡	景點
時下							
最	5						
夯	4	5					
的	3	4	5				
拍照	2	3	4	5			
打卡	1	2	3	4	5		
景點		1	2	3	4	5	

註：空格的共現強度值為 0。

1. 將設定寬度的文本分析窗口掃過巨量的文本，累加所有詞彙各自的共現強度值
2. 去除與所有字詞都有高共現強度的功能字詞，以及去除與其他字詞過少的交集致使共現矩陣過度稀疏 (sparse) 的罕見字詞。

3. 將共現強度值分數轉換為標準常態分數。
4. 運用巨量文本的詞彙共現矩陣中所有成員間之向量空間關係，分析結構。

當兩個詞彙的共現型態越相似，也就是它們與其他詞彙的向量組成型態越接近時，在超高維度語意空間 (Hyper-space) 中的距離就越短。以圖 2 為例，「消費者」是一個 12 個維度的向量，當它與「供給」、「勞工」和「所得」等其他高維度詞彙向量投射在一個二維平面上，這些向量間的距離或彼此間之餘弦投影 (cosine projection) 可以作為其相似程度指標。

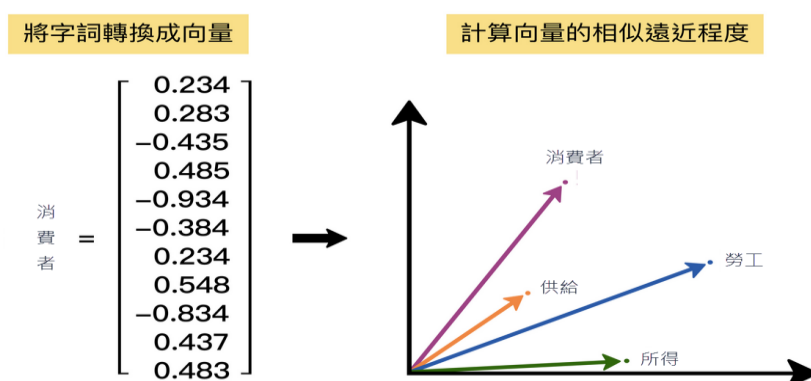


圖 2: 高維度語意空間中不同詞彙向量示意圖

改編自 <http://research.sinica.edu.tw/wp-content/uploads/2018/07/ma-natural-language-processing-04.jpg>

對標準常態化的詞彙語意空間矩陣可以進行各種多變項分析來探討其結構；圖 3 中以奇異值分解 (singular value decomposition, SVD) 演算法，將一個 100 萬×100 萬的高維度語意空間矩陣 A，拆解為 X、B、和 Y 三個組成矩陣的乘積，而這三個組成矩陣的維度可以彈性選擇，其三個矩陣的表徵內容分別為：

第一個矩陣 X 中的每一欄表示意思相關的一類詞，其中的每個非零元素表示這類詞中每個詞的重要性 (或者說相關性)，數值越大越相關。第二個矩陣 B 則表示詞彙類別間的關係，其維度代表詞彙有多少類別。第三個矩陣 Y 中的每一列表示一個類別，其中每個元素表示該類別中每個詞的重要性。

因此，只要對關聯矩陣 A 進行一次奇異值分解，就同時完成詞彙分類，以及得到詞彙類別間的相關性，也就是研究者想知道的詞彙概念結構。

$$A_{1,000,000 \times 1,000,000} = X_{1,000,000 \times 100} \times B_{100 \times 100} \times Y_{100 \times 1,000,000}$$

$$A = \begin{pmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{ij} & \dots & a_{in} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & \dots & a_{mj} & \dots & a_{mn} \end{pmatrix} \quad \begin{matrix} \boxed{A} \\ \\ \\ \end{matrix} = \begin{matrix} \boxed{X} \\ \\ \\ \end{matrix} \quad \begin{matrix} \boxed{B} & \boxed{Y} \end{matrix}$$

圖 3: 奇異值分解示意

改編自：<https://www.twblogs.net/a/5b87db942b71775d1cd94b79>；吳軍，2014

經由 SVD 所得之詞彙類別關係矩陣 B，通常以 100-300 個維度，能達到最佳的概念結構化約與類別相似性區辨間的平衡 (Landauer & Dumais, 1997)。對 B 亦可進一步利用多變量分析技術降低維度方便理解其主要成份。例如圖 4 係以主成份分析 (Principle Component Analysis, PCA)，將圖 X2 中 SVD 中所得的詞彙類別從多維度降維成為 2D 平面。

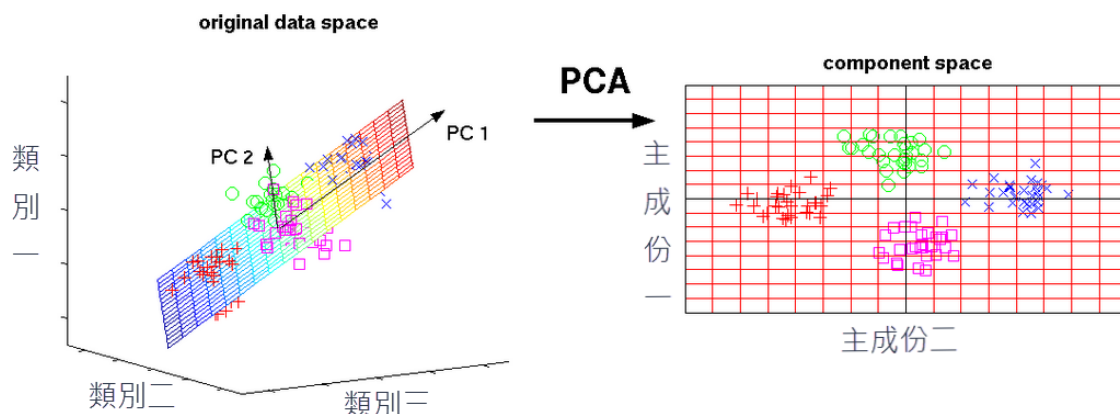


圖 4: 對類別相似矩陣進行主成份分析結果

改編自：https://miro.medium.com/max/1619/1*-I2LNBs9hpZT56hzHu2P9A.png

3. 研究問題(Research Question)

「如何量測學習成效」無疑是近年教學實踐研究計畫與高教深耕的核心問題所在。嚴格地說，在研究架構中採用實驗組與對照組的設計並執行前、後測，對於有效驗證某一創新教學方法能否有效提升學生學習成效是不可或缺的。

然而，前測與後測要測的能力包括了哪些？又該如何設計相對應的測試內容？

美國教育心理學家 Benjamin Bloom 於 1956 年將知識的認知範疇區分出記憶、理解、應用、分析、評鑑、創造等不同層次的能力，對於教學現場的老師而言，當然希望學生不僅僅停留在記憶和理解的基礎程度，還要能夠繼續向上攀升。對於如何做才能提升學生的能力層次，國內外學界目前已有各式各樣充滿創意的作法來嘗試挑戰。

但要怎麼樣才能知道學生有沒有達到某個水準呢？讓學生自行評估學習效能？透過某個專題在執行後由老師依據成果加以判斷？還是讓學生來回答傳統的考試題目？又，答對得愈多就代表愈有學習成效嗎？我們如何知道學生答對試題是猜的、背的還是真正懂了？我們又如何知道學生得到的是片段的還是融會貫通的知識呢？有些試題說不定連大學教授在沒有準備的情況下也無法立即得出正確解答，但我們難道會認為這些教授不具有高層次的知識能力嗎？

本研究計畫企圖【提出一個頗有創意的作法來量測學生的程度】，建構一個與學科概念相關之語意空間結構，並藉其演變來量測學習成效。

4. 研究設計與方法(Research Methodology)

本計畫利用 Q-sort 行為實驗取得素人 (學生) 與專家之《經濟學》概念詞彙結構，以及利用高維度語意空間表徵演算取得《經濟學》教科書和專業期刊論文等文本資料中對應於 Q-sort 結果之概念詞彙結構概念。來自行為實驗以及來自文本資料的概念詞彙表徵，可以互相比較對照受試者群體間差異，以及受試者與經典文本間之差異；亦可以再整合兩種來源概念詞彙之資訊，形成有更多特徵維度的複合矩陣表徵，建構出概念結構評分指標，應用於評估未來新增受試者之概念結構專業程度。

基於上述研究目的與理論基礎，在行為實驗方面，本計畫以 Q-sort 方法測量受試者的概念結構。受試者區分為修習《經濟學》相關主題課程之學生以及《經濟學》相關專業之大學教授。其中學生於第一學期之期初、期中與期末接受測量，以評估其概念結構變化。

在文本語料分析方面，本計畫以計算語言學中各種自然語言處理 (Natural Language Processing) 的工作流程，將來自《經濟學》領域文本語料建置為高維度語意空間。

在研究架構方面，本計畫採用【實驗分析】的架構 (如圖 6) 來進行研究，共有 21 位教授《經濟學》課程的大專經濟學老師參與 (原 24 位，3 位因故退出)，受試者計有 1592 位修課學生。我們依照各班級所採用的教學模式 (遊戲或單向講授)、教案內容以及講授順序，將所有參與的老師以隨機方式分為四組，彼此互為實驗組與對照組。各班學生一起進行一次前測與數次後測，在控制班級與學生個人特性的前提下，藉由前後測差距的比較，也就是差異中差異法 (Difference in Difference)，來驗證各種教學模式的學習成效。



圖 6: 本計畫之實驗分析架構

在知識結構量測工具方面，本計畫將採用線上版本的 Q-sort (例如利用 Q-Assessor 或 Q-sortware)¹ 以進行快速與大量的施測；由專家學者建議以及從《經濟學》教科書章節和論文關鍵字與摘要中，決定要包含於 Q-set 中詞彙或命題陳述。

在文本分析部份，我們利用網路爬蟲 (web crawler) 自教科書中搜集《經濟學》主題

¹ Q-Assessor: <https://q-assessor.com/> ; Q-sortware: <http://www.qsortware.net/>

領域之中文語料，再使用 python 的 jieba 套件將所得語料斷詞。斷詞後以 Google 的開源工具 word2vec 將詞彙轉成上述向量表徵，再藉由 singular vector decomposition (SVD) 演算法將高維度的詞彙向量空間降階為特定維數的語意向量空間（參見陳明蕾，王學誠 & 柯華葳, 2009）。工作流程階段以及各階段採用之計算程式工具如圖 7 所示：



圖 7：經濟學文本高維度語意空間分析流程

5. 教學暨研究成果(Teaching and Research Outcomes)

利用前述的這套工具，本計畫分析了三本具有代表性的經濟學教科書（兩本中文一本英文），我們在所有找出來的關鍵詞對 (keyword pairs) 當中隨機地挑選了 120 組制做問卷，問卷中的每一組詞對我們都要求填答者填寫 1 到 7，來代表在他們心目中這兩個詞彙彼此之間的關聯強度由最弱到最強。我們也針對受試者的問卷填答進行了有效與否的檢測，並剔除這些違反檢測標準的樣本。

此外我們也邀請了三位國內知名的經濟學家來填答這份問卷，每一組詞對的關係強度取其中位數來做為「專家共識值」，以作為量測知識結構的絕對標準。三位專家的填答結果各自跟「專家共識值」的相關係數由 0.7558 到 0.8421。之所以我們不採用三本教科書的共識值來做為絕對標準呢？這是因為教科書在編寫的時候有其目的性，是為了向沒有學過的素人介紹這一整套知識，因此在編寫的順序上面有其引導作用。但專家的知識事實上是超越了單本教科書，在學位養成過程中，專家所閱讀的書籍與論文在數量與內容上都不是一本入門用的經濟學教科書可以比擬的，其知識結構的來源要比教科書而言龐大許多。

為了要與我們新創的這一套「語意空間結構」測量結果進行比對，本計畫也讓參與的一千多位同學做了傳統的學科測驗，接著介紹學科測驗在前測跟後測中的分布狀況。如圖 8 的測驗成績分布所示，在轉換為百分制的情況下，前測的平均成績為 35.13 分，經過一個學期，後測的成績進步到 50.52 分，代表在一個學期的上課之後，學生在學科測驗方面的表現確實是進步不少。

再將每一位受試者所填答的詞彙關聯問卷結果拿來跟專家共識值進行比對，利用兩者之相關係數來測量學生們的知識結構完整度。如圖 9 的相關係數分布所示，前測時學生的知識結構完整度平均為 0.1429，經過一個學期，後測則上升到 0.1574，代表在一個學期的上課之後，學生對於經濟學的知識結構在完整度方面確實有一些進步，但是相較於專家平均值的 0.78 則還是差了一大截，不過這並不令人感到意外。

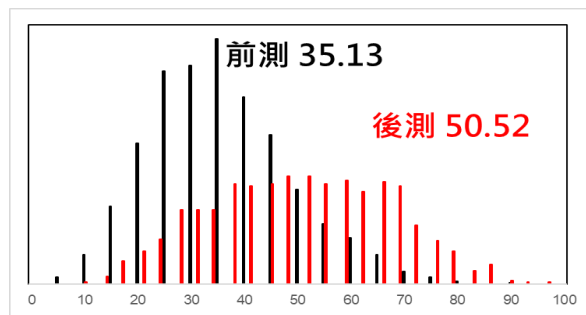


圖 8: 學科測驗成績分布

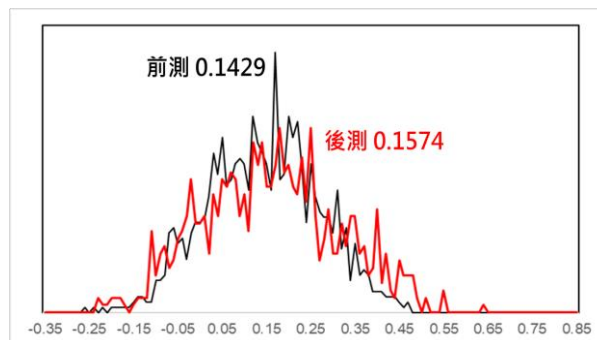


圖 9: 知識結構完整度之分布

若是比較學科測驗成績與知識結構完整度，我們發現不論是前測或後測，兩者之間的相關係數大約落在 0.36 到 0.42 之間，代表知識結構完整度越高的學生，其學科測驗也會表現得比較好。

6. 建議與省思(Recommendations and Reflections)

本計畫建立了一套量測經濟學知識結構完整度的工具，從建構流程即可發現，這一套方法的應用範圍其實不僅僅是經濟學，未來也可以應用在其他學科上。未來我們準備更進一步深入探究，利用幾位專家填答結果的一致程度高低，將詞對分為幾種類型來分析。

除了量測學生知識結構的完整度之外，本計畫的研究成果還可以應用在教科書的挑選或者是教學素材的選用上。此外，我們也必須思考如何讓學生有誘因，願意認真填答這份詞彙關聯問卷，畢竟以目前發展的初期階段而言，要拿這套工具的量測結果來作為學期成績的計算標準還不是時候。

二. 參考文獻(References)

- Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H.; Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. *Handbook I: Cognitive domain*. New York: David McKay Company.
- Brown, S. R. (1980). *Political Subjectivity: Applications of Q Methodology in Political Science*.
- Burgess, C., & Lund, K. (1997). *Modelling Parsing Constraints with High-dimensional Context Space. Language and Cognitive Processes*, 12(2–3), 177–210.
<https://doi.org/10.1080/016909697386844>
- Colander, David and Kim Marie McGoldrick (2009), The Economics Major as Part of a Liberal Education, *American Economic Review*, Vol. 99, No. 2, pp. 611-623.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*. [https://doi.org/10.1016/S0022-5371\(69\)80069-1](https://doi.org/10.1016/S0022-5371(69)80069-1)
- Durham, Y., McKinnon, T., & Schulman, C. (2007). Classroom experiments Not just fun and games. *Economic Inquiry*, 45(1), 162-178.
- Emerson, T. L., & English, L. K. (2016a). Classroom experiments: Teaching specific topics or promoting the economic way of thinking. *The Journal of Economic Education*, 47(4), 288-299.
- Emerson, T. L., & English, L. K. (2016b). Classroom experiments: Is more more? *American Economic Review*, 106(5), 363-67.
- Eisenkopf, G., & Sulser, P. A. (2016). Randomized controlled trial of teaching methods: Do classroom experiments improve economic education in high schools?. *The Journal of Economic Education*, 47(3), 211-225.
- He, J., Hu, B. Y., & Fan, X. (2019). Encyclopedia of Personality and Individual Differences. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–4). <https://doi.org/10.1007/978-3-319-28099-8>
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, 28(2), 203–208.
<https://doi.org/10.3758/BF03204766>
- Nagashima, S. (1961). OBJECTIVE APPRAISAL OF EFFECTS OF GROUP THERAPY : A Q-TECHNIQUE STUDY OF CHANGES IN SELF CONCEPTS. *Japanese Psychological Research*, 3(4), 179–192.

Rosch, E. H. (1973). *Natural categories*. *Cognitive Psychology*. [https://doi.org/10.1016/0010-0285\(73\)90017-0](https://doi.org/10.1016/0010-0285(73)90017-0)

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/0096-3445.104.3.192>

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*.
<https://doi.org/10.1037/h0036351>

Yandell, D. (1999, March). Effects of integration and classroom experiments on student learning and satisfaction. In *Proceedings: Economics and the Classroom Conference*, pp. 4-11.

周郁凱，《遊戲化實戰手冊》，商業周刊出版社，2017。

亞倫·迪格南，《加入遊戲因子，解決各種問題》，先覺出版社，2012。

黃教益、陸定邦與孔憲法，教學相長：學生社會網絡中介電腦繪圖學習動機與成效，
《設計學報》，17卷3期，2012。

陳明蕾, 王學誠, & 柯華葳. (2009). 中文語意空間建置及心理效度驗證：以潛在語意分析技術為基礎. *中華心理學刊*, 435(1), 415 – 436.

三. 附件(Appendix) (請勿超過 10 頁)

與本研究計畫相關之研究成果資料，可補充於附件，如學生評量工具、訪談問題等等。