

教育科學研究期刊 第五十六卷第一期

2011 年，56 (1)，129-166

內容探勘技術在教育評鑑研究發展 趨勢分析之應用

曾元顯

國立臺灣師範大學資訊中心
研究員

林瑜一

臺北市立教育大學教育學系
研究生

摘要

由於大型資料庫時代的來臨，各種教育評鑑研究的數位文件，累積的速度相當快，相關的統計、探勘分析方法，需求漸為殷切。本研究以專家問卷、資料庫搜尋等方式，來界定教育評鑑研究領域的核心期刊。另外，透過引文分析、文字探勘技術等方法，自動從核心期刊中偵測出校長領導、班級大小、評鑑、改革、教師認證、學習機會、留級、教育券等概念，為顯著的研究議題。從評鑑方法的主題中，本研究探勘出 Chen、Weiss 與 Rossi 等學者討論的方法，比 Cook、Compbell 等人的還新，此點從領域學者 Alkin (2004) 在主題樹中對上述專家的擺設位置，也觀察到相同的現象。最後，本研究透過主題趨勢的分析，發現教育評鑑裡的主要議題，其論文數量有愈來愈少的趨勢。然而，這些結果的意義，仍有待領域專家的深入解讀。本研究所提方法可應用在不同學科領域，不僅能增進內容分析的效率，還可定期追蹤主題的演變，甚至標示出重要學者的位置，提供後續相關研究的線索與素材。

關鍵字：主題偵測、科學計量學、核心期刊、評鑑理論樹、趨勢分析

通訊作者：曾元顯，E-mail: samtseng@ntnu.edu.tw

收稿日期：2010/01/17；修正日期：2010/06/21、2010/09/03；接受日期：2010/09/14。

壹、前言

統計分析方法，在教育領域的各種研究中，扮演相當重要的角色。針對各項結構化數據（structured data），如學生年齡、智商、家庭收入等資料，透過各種統計運算，可以客觀地驗證研究假說、推論觀察結果。因此，適當地運用統計技術，對提升研究的水準與效率有莫大的貢獻。隨著近年來大量教育數據的累積，更引發進階統計技術的需求，如資料探勘（data mining）等方法的運用，以便從一堆數據中，探勘其中隱而未顯、潛在有用的資訊或知識（Baker & Yacef, 2009）。

相對於結構化數據，另一種重要的資料是由自由文字（free text）組成的非結構化資料（unstructured data），如科學研究論文、專利技術文件、質性訪談資料，以及開放性問卷中的回答內容等具有分析或研究價值之文獻。分析這些非結構化資料的探勘方法，仰賴對自由文字的進階處理與統計運算，也就是文字探勘（text mining）的技術。

近 20 年來學術文獻資料庫蓬勃發展，不僅文獻數量快速成長，文獻內容也同時包含結構化與非結構化等豐富資料。針對大量學術文獻的分析探索，如主題辨識與趨勢分析，需結合上述不同技術的內容探勘（content mining）方法，以便快速掌握學術領域中各研究主題歷年的脈絡演變或發展軌跡，從而對後續研究的發展方向，提供有價值的參考資訊。

本文提出一套文獻內容探勘方法，以教育領域中的教育評鑑研究為對象，進行相關文獻的分析。教育績效的評鑑，近年來在臺灣受到極大的重視，由大專院校的系所評鑑，到中、小學的校務評鑑、課程評鑑，甚至特殊教育評鑑等不勝枚舉。此現象一方面肇因於近年來教育政策的變動、大量資源的投入，需要瞭解其成效；一方面希望透過評鑑的回饋來發現問題，提升整體的教育品質。在探討教育評鑑方法或策略時，若能參酌文獻上記載過的發展軌跡，對後續教育評鑑的研究與發展，當具有啟發反思之功效。

對於學術研究的主題趨向，常常依賴領域專家的知識與經驗，或以人工方式分析整理大量文獻，以提供資訊作為參考判斷之依據。此種方法雖然可以得到精確有效的結果，但也往往容易流於主觀的篩選或判斷，或受人力、物力與時間上的限制，而難以持續。若有自動化的分析模式與方法，可以全面性、持續性地長期觀察教育評鑑研究領域的發展趨勢、主題演變、研究活動，以及相關重要研究人員的社會網絡分布狀況，當可彌補或輔助人工分析的不足。

基於上述需求，本文提出一套文獻探勘的模式與方法，運用在教育評鑑研究的主題發展與趨勢分析上。其主要的目的與內容歸納如下：

- 一、發展可應用於學術文獻分析與探勘的技術和工具。
- 二、界定教育評鑑領域的核心期刊，以供主題、趨勢分析之用。

三、運用上述方法與資料，探索、解析出教育評鑑領域近數 10 年來的研究主題、趨勢脈絡、並確認此研究領域的重要學者，以供後續研究的參考。

本文結構如下：第貳部分將介紹相關的研究，第參部分則詳細介紹本文提出的分析方法，第肆部分則以教育評鑑領域為例，根據專家意見與資料庫收錄狀況，決定教育領域的核心期刊，再運用本文所提之方法，分析、比較、解讀教育評鑑領域近數 10 年來的研究主題、趨勢與重要研究學者。本文特別著重分析結果的可靠程度，因此不僅以不同的分析方法來互相比較，另外也根據專家的知識，來檢驗結果的有效性。最後，則摘要本文的重要結果，並說明其意涵。

貳、相關文獻探討

針對具備學術價值的文件所進行的內容分析或探索，在數個學術領域中都有各自的研究與發展。例如，在圖書資訊領域有科學計量學 (scientometrics) (蔡明月, 2003; Andres, 2009; Leydesdorff, 2001)，在人文社會學領域有內容分析法 (content analysis) (Krippendorff, 2003)，在電腦科學中則有文字探勘技術 (Feldman & Sanger, 2006)。其具體內容與技術細節，在各領域中或有不同，但目標或方向卻大致一樣。以文獻計量學而言，國內學者何光國 (1994) 提到：舉凡利用數學、統計和邏輯理論及方法，對各類文獻的特質、結構，做數量、品質及運用上的研究和分析，通稱為文獻計量學。國外學者 Holsti (1969) 則定義內容分析法為：任何客觀、系統化的訊息特徵辨識，以供推論的技術。Lasswell (1964) 闡述內容分析的核心問題，在瞭解英文中的 W 開頭疑問字的意義，亦即：“Who says what, to whom, why, to what extent and with what effect?”。Hearst (2003) 說明文字探勘為：由電腦從各種文件自動擷取資訊，以發現新知的過程，其重點在連結各種被擷取的資訊，以形成新的線索或假說，俾利傳統方法的進一步探索與驗證。

為達上述目的，各領域皆發展出相關的模式與方法。由於相關的研究範圍很大，以下僅回顧幾項啟發本文的研究與其分析技術，並做簡要的評論，以凸顯本文所提分析方法的特色。

一、日本學者村田忠禧的歷史分析法

日本學者村田忠禧曾提出了一種探勘歷史新聞，來分析政治變遷的方法 (村田忠禧, 2002)。他以 1949 年到 2002 年《人民日報》每年的元旦社論為材料，先自訂待觀測的關鍵詞彙，然後統計這些詞彙出現的篇數與年代。依此資料分析每年焦點詞彙的變化，並觀看某年焦點詞彙在其他年份出現的狀況，以瞭解該年份與其他年份的連動關係。村田忠禧認為這些詞彙的變化可以宏觀地觀察歷史變動，且此種分析結果不是在推翻過去對歷史的解釋，而是做進一步的補充 (亦即讓資料自己說話)，他認為這正是其詞彙頻率統計分析有效性之所在。其分析方法使用到詞彙索引、年代分布圖、互動式的年代重點詞彙提示與查詢，以滿足歷史

學者的分析需求。然其最重要的起始步驟，需自訂「待觀測的關鍵詞彙」，是此方法最耗費專業知識、人力與時間的地方。

二、IBM 的術語趨勢分析法

與村田忠禧（2002）的方法比較，IBM 的 Lent、Agrawal 與 Srikant（1997）則提出較自動化的作法。他們以資料探勘中的關聯探勘（association）技術，自動擷取專利文件中常常連續出現的字詞，來獲得專利術語，並根據其出現的文件篇數與年代分布，計算出每個術語的趨勢。將這些術語按照其趨勢分類成：漸熱門、漸冷門、熱門後冷卻等類別，提供用戶瀏覽或查詢，以方便瞭解某些術語的趨勢。此方法提供相當有效的研究趨勢演變分析。可惜其所擷取出來的重要詞彙相當龐雜，而且大部分屬於「詞彙」的層次，而非概念、主題或領域等較上位層次。因此，若能將相同主題的詞彙，歸類成概念或主題，再進行趨勢分析，當可改進其缺點。

三、荷蘭萊登大學科學與技術研究中心的領域分析法

荷蘭萊登大學的 Noyons 與 van Raan（1998）從研究文獻中，分析與觀測科學技術的演進，經過長期的經驗累積，發展出一套通用的方法：透過學術文獻蒐集、詞彙擷取與過濾、索引檔建構、共現字分析、文件分類、主題地圖視覺化呈現、專家解讀等步驟，來分析科學與技術文獻。然而，即便文獻經過各式文字探勘的歸類處理，並繪製成主題圖，其解讀仍不容易。他們的網站曾提到：如果知道如何解讀的話，這樣的地圖可提供研究領域的簡要概觀，以供進一步發掘資訊。因此，若能加入更多額外的資訊（特別是額外的結構化資料，如作者、國家、引用文獻等），應當能提升其結果的可解讀性。

綜上所述，這些方法雖具成效，但仍屬專家層級才能使用、解讀；而且在應用時，視待分析文件的範圍、數量、特性與應用而定，各有特殊環境，因而在某種程度上需要客製化分析流程或方法。這些緣由，促成本文擬自行開發一套分析流程，期能整合上述方法的優點，避開其缺點，讓有興趣的研究者，如研究生、初入門之學者、跨領域研究者等，分析探勘學術文件，進而達到探索瞭解領域概觀之目的。

參、研究方法

參考上述文獻的作法，本文提出的分析技術比較接近荷蘭萊登大學 Noyons 與 van Raan（1998）提出的方法。然而在一些重要的步驟，如多階段主題歸類、類別描述詞擷取、主題趨勢分析、結果印證等部分，本文有顯著的不同。詳細的內容探勘分析方法，描述如後。

一、相關文獻界定

根據文獻對一個領域進行分析之前，先界定待分析文件的範圍是一件非常重要的工作。界定方法相當多元，視成本與應用情形而定。可以簡單到以某個特定詞彙的查詢結果為待分析文件的集合，也可以複雜到每一篇文件都經由專家反覆討論，來確認是否包含在待分析文件中。有關教育評鑑方面的文獻，無法用簡單的概念詞檢索來涵蓋。在考量時間與成本之後，本文透過數篇種子論文與資料庫收錄狀況，配合專家的問卷調查，整理出該領域的重要期刊，簡稱為「核心期刊」，將其包含的論文視為相關文獻，依此界定出待分析文件的範圍。

二、文獻下載剖析

本文提出的分析方法可適用於任何具備自由文字的文獻，如學術論文、專利文獻、新聞、訪談稿等具備主題論述的文件。惟本文分析的對象以期刊論文為主。考量到文獻收錄範圍、資料整理、名稱權威控制、引文加值處理等面向，Thomson Reuters 的 Web of Science (簡稱 WoS) 論文資料庫，乃現今做得較好的期刊資料庫之一，方便本文的運用或自行再處理。因此，依據上一步驟界定的文獻範圍，本步驟從 WoS 資料庫下載歷年來的資料，並進行更細緻的文件記錄切割、分欄擷取、資料正規化 (例如，處理大小寫不一致的情形)、偵測重複或不完整資料等處理，以驗證資料的正確性、一致性，確保後續分析結果的可靠性。同時，將其論文編號、篇名、摘要、關鍵詞、描述詞、分類詞、出刊年代、作者、機構、國別、期刊名稱、引用文獻等資訊，解剖出來儲存至關聯式資料庫中，以便於管理與運用，甚至直接進行結構化資料的量化分析與圖表製作，從而掌握其統計特性，對其有初步之認識。

從 WoS 資料庫下載的論文紀錄，均為純文字檔案，每一篇皆包含約四十個項目，每個項目占據一到數行，以項目簡稱開頭，並以空格和換行字元排列項目內容版面。由於其版面具有規則性，經歸納整理後，以程式擷取其中十三個重要項目供後續應用，其項目簡稱、意義與內容範例列舉如下：

AU：作者，例：Anagnostopoulos, D; Bali, VA; Roberts, R。

TI：論文標題，例：Toward a political explanation of grade retention。

SO：期刊全名，例：EDUCATIONAL EVALUATION AND POLICY ANALYSIS。

DE：論文關鍵詞，例：district policy; grade retention; race。

ID：論文描述詞，例：CLASS SIZE; REFORM; URBAN SCHOOL DISTRICTS;。

AB：論文摘要，約 100-300 個英文字。

C1：作者所屬之國家 (原為作者地址，經程式剖析後，得其國家名稱)。

CR：參考文獻，例：CIBULKA JG, 2001, EDUC POLICY, V15, P12。

NR：參考文獻篇數，例：103。

TC：被引用的次數，例：4。

PY：論文出版年，例：2005。

SC：論文所屬領域別，例：Education & Educational Research。

UT：WoS 之論文主鍵，例：ISI:000230839500002。

三、文件關聯計算

為顯現待分析文件隱含的知識結構，此步驟在建立兩兩文件之間的關係，供後續主題自動偵測運用。有三種常見的文件關聯計算方式：共被引（co-citation）、書目對（bibliographic coupling）以及共現字（co-word）（Moed, 2005）。如圖 1 範例所示，共被引是指 2 篇文件（如 Paper A 與 B）被同一篇文件（如 Paper 1）所引用，則此 2 篇文件產生共被引關係，且這 2 篇文件被愈多論文共同引用到，則其關係（通常為主題上的相似度）被認為愈強，如圖 1 中論文 A 與 B 的強度為 1，但 C 與 D 則為 2。書目對則指 2 篇文件（如 Paper C 與 D）若引用到共同的書目資料（如 Ref. 2），則此 2 篇文件產生了耦合關係（coupling）。同樣地，共同的書目資料愈多，表示此 2 篇文件在主題上愈相似、關係愈強。至於共現字，則是指 2 篇文件的內文（如標題或摘要）裡，使用到相同的關鍵詞彙（亦即某關鍵詞彙共同出現在此 2 篇文件的情形，如 Word 2 同時出現於 Paper A 與 B 中），當共現字愈多，則此 2 篇文件的主題相似度也愈高。

不管是哪一種關聯強度（整數值），最後都會以適當方式將其正規化為相似度（0 到 1 的分數值）。本文採用 Dice 係數（Salton, 1989）方式進行正規化，其計算方式如公式(1)所示：

$$\text{Sim}(X, Y) = 2x|S(X) \cap S(Y)| / (|S(X)| + |S(Y)|) \quad \text{公式(1)}$$

以書目對為例，S(X)代表文件 X 的參考書目所組成的集合， \cap 代表兩集合的交集，|S(X)|代表集合 S(X)的元素個數。依公式(1)，圖 1 論文 C 與 D 的書目對相似度為 $2 \times 1 / (3 + 2) = 2/5 = 0.4$ 。

在計算共現字關係時，需要用到資訊檢索或文字探勘的技巧，以刪除停用詞、正規化詞彙（大小寫一致化，再取英文詞幹）、擷取關鍵詞（如 multi-word phrases）、進而建立詞彙到文件的反向索引資訊檔案，以方便任意兩文件之間的共現字快速運算。這部分我們已投入多年的研究與實作，有成熟的程式可以直接應用。在計算書目對關係時，則需要將參考文獻正規化（或進行權威控制），使不同作者引用到的相同論文，可以容易被辨識、比對出來。從 WoS 下載的資料，其參考書目已被正規化在其 CR 項目裡，因此可以直接進行書目對計算。至於計算共被引關係所需要的引用文獻，則要針對每一篇待分析文件，到 WoS 另行下載，不僅相當費時費事，而且不同時間點下載到的資料可能會有所不同。這是因為 1 篇文件可能隨著時間愈久，而獲得愈多的引用。亦即不同時期進行的共被引分析，其結果將會不同！對共現字與書目對而言，這是不會發生的問題。基於上述說明，在時間與成本的考量下，後續將僅依書目對與共現字來計算文件之間的關聯相似度。

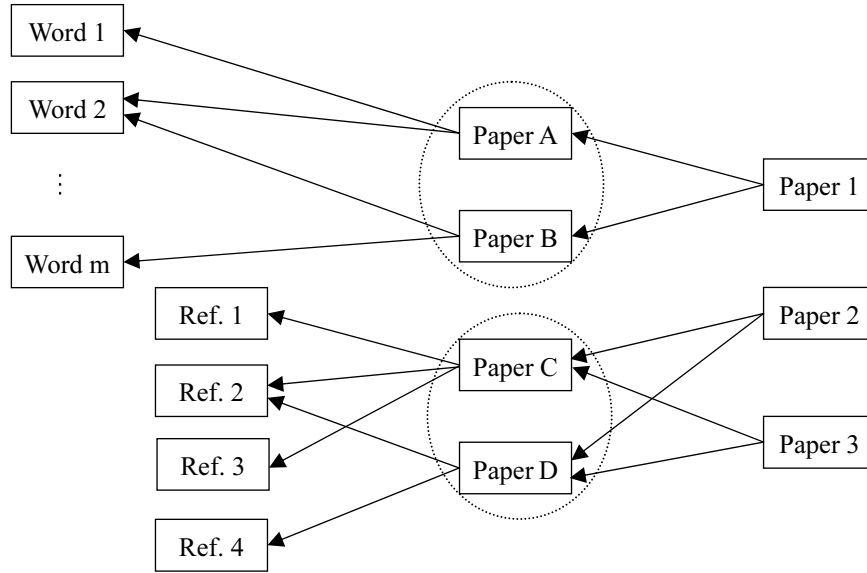


圖1 共被引、書目對、共現字關係示意

四、主題自動偵測

在上述步驟中，不管是根據哪一種方法，針對 n 篇文件的兩兩文件（如 d_i 與 d_j ）計算，最後均可算出如下的相似度矩陣：

$$\begin{matrix}
 & d_1 & d_2 & \cdots & d_n \\
 d_1 & \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \end{bmatrix} \\
 d_2 & \begin{bmatrix} s_{21} & s_{22} & \cdots & s_{2n} \end{bmatrix} \\
 \vdots & \begin{bmatrix} \vdots & \vdots & \ddots & \vdots \end{bmatrix} \\
 d_n & \begin{bmatrix} s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix}
 \end{matrix}$$

依此矩陣，有三種方法可以自動將多篇文件群聚起來，達到主題偵測的效果。第一種是以「奇異值分解」(Singular Value Decomposition) 為基礎的「特徵向量分析」(Eigen Vector Analysis) (Golub & Van Loan, 1996) 或「因素分析」(Factor Analysis) (Gorsuch, 1983)。其可把這 n 篇文件投影到少數幾個特徵向量上，每個特徵向量可代表一個主題，且每篇文件在每個主題上都有其權重。根據這些權重以及適當的門檻值，可以得知哪些主題包含哪些文件，或哪些文件屬於哪個主題。第二種方法是以「多維縮放」(Multidimensional Scaling, MDS) 技術 (Kruskal, 1977)，將這 n 篇文件依照其相似度，投影到 2 維或 3 維人眼可見的地圖上，使得相似度高的文件群聚在互相靠近的位置，而相似度低者則相距在較遠的位置。達到類似這種效果的方法還有「自我組織圖」(Self-Organization Map) (Kohonen, 2000) 等技術。第三種

方法則是「文件歸類」(Document Clustering) (Jajuga, Sokolowski, & Bock, 2002; Salton, 1989), 其中的「層次凝聚歸類法」(Hierarchical Agglomerative Clustering), 不需使用者指定類別數, 可反覆地將最相似的文件或類別聚集在更大的類別中, 如此逐漸由下而上的將所有文件組織起來。特別是其中的完全連結法 (complete link), 可將彼此相似度都高的文件歸在同一群組。根據過去的研究經驗, 實際應用多維縮放與層次凝聚歸類於主題偵測上, 可獲得良好的結果 (Tseng, Lin, C.-J., & Lin, Y.-I., 2007)。因此, 本文也就以這兩種歸類方法為主。

五、多階段主題歸類

理論上, 層次凝聚歸類法可將文件組織起來。實際上, 受限於文件關聯的實際情形 (例如, 待分析文件的共同參考書目不多, 僅一群群少數的文件可以形成書目對), 直接進行文件歸類的結果, 可能造成為數眾多的小類別。另外, 若以較高的相似度門檻, 作為文件聚類的條件, 以期得到議題上較為一致的類別, 也同樣可能會造成為數眾多的小類別。如此的資訊組織結果, 對主題分析的助益並不大。但若把這些探討類似議題的少數文件所組成的小類別, 視為一個個不同的「概念」(concepts), 那麼將這些小類別 (實做時將類別中的文件合併成一份虛擬文件) 再做一次上述的歸類動作, 如圖 2 所示, 則可將「概念」歸類成「主題」(topics), 如有需要, 可將「主題」再歸類成「次領域」(domains), 「次領域」再歸類為「大領域」(fields)。如此, 可不受文件實際關聯情形的限制, 總是可以得到類別數量與類別大小都便於解讀的歸類結果。這樣的歸類策略, 在此稱為多階段主題歸類 (multi-stage clustering)。這裡提到的概念、主題、領域等知識範圍, 均是主觀的認定, 因此嚴格的對其加以區分並不重要。只要人工分析歸類結果時, 可以辨識出待分析文件中的知識結構, 則此種多階段歸類過程, 可在任何階段停止。

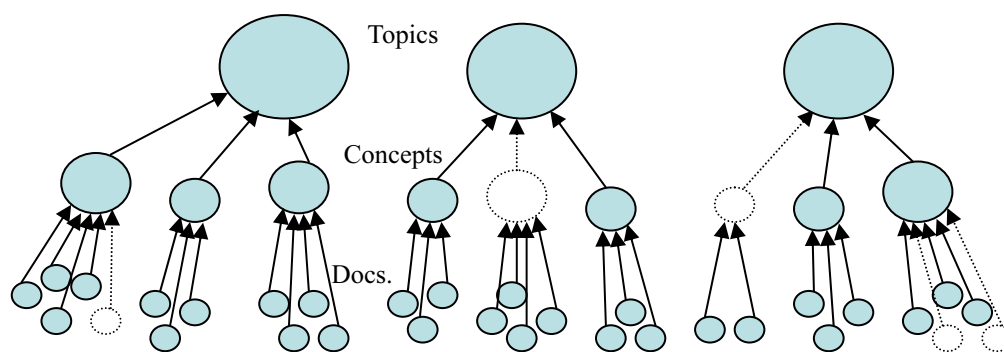


圖2 多階段歸類示意

註：圖中虛線空心圓圈代表低於門檻而在下一階段沒有被歸類的文件或類別 (亦即空心圓圈代表其自成一類)

圖 2 中的示意，僅是一種理想狀況。實務上，兩個明確的概念，即便有一定程度的相似性，歸類合併後可能無法聚焦成一個有意義、可解讀的主題。例如，當「校長領導」與「班級大小」兩個概念被歸類在一起時，將難以解讀出一個清楚的主題。其背後的原因，可能是這兩個概念的文件都參考到相同的教育政策文獻，而形成書目對，或是校長領導的卓越與否，間接影響到班級大小的決定而存在共現字。因此，多階段主題歸類僅是一種類別合併或類別排序的策略（方便管理、辨識類別）。在某一階段無法解讀出類別主題時，可以回溯上一階段去解讀其所包含的概念，而不必受限、拘泥於在同一階段作主題上的判讀。亦即，實務上在進行多階段歸類結果的分析時，分析焦點可能會在不同階段的類別之間來回遊移，以便進行有意義的結果歸納與解讀。

六、類別描述詞擷取

前述歸類結果，僅知類別中包含哪些文件，要解讀類別的內容主旨，至少需瀏覽過文件的篇名，甚至摘要。為協助人員快速解讀類別內容，本文採用 Tseng (2010a) 的作法，在每個歸類階段，自動擷取各類別特徵詞彙作為其描述詞，以便利主題的快速判讀。其作法是以關鍵詞自動擷取方法 (Tseng, 2002) 擷取篇名與摘要中的重要片語或詞彙，再根據表 1 的篇數交叉分析表，計算詞彙與類別的相關係數：

$$Co(T,C) = \frac{(TP \times TN - FN \times FP)}{\sqrt{(TP+FN)(FP+TN)(TP+FP)(FN+TN)}} \quad \text{公式(2)}$$

其中 TP 、 FP 、 FN 、 TN 分別代表詞彙 T 出現在類別 C 中的篇數 (TP)、 T 在其他類別中的篇數 (FP)、 C 不包含 T 的篇數 (FN) 以及其他類別也沒出現詞彙 T 的篇數 (TN)。

表 1 詞彙 T 在類別 C 中的出現篇數交叉分析

		Term T	
		Yes	No
Cluster C	Yes	TP	FN
	No	FP	TN

公式(2)的意義，在於當詞彙 T 只出現在類別 C 中的每一篇文件，而在其他類別都不出現時（亦即 $FP=FN=0$ ），則 C 與 T 為最大正相關，有最高值 +1；反之（ $TP=TN=0$ ）為最大負相關，有最低值 -1。因此，本項係數可反映詞彙 T 在類別 C 中的相關程度。據此，針對每個類別，可以計算每個詞彙的相關係數（或者稱為該詞彙在該類別的代表性係數），再根據此係數由高到低排序詞彙，最後選取其中的前數個詞彙，即可作為該類別的描述詞（即類別的代表性詞彙）。

此相關係數對包含短文件且數量眾多的小類別，可發揮統計優勢而得出不錯的結果。但因沒有考慮到詞彙在文件中的出現次數，對包含長文件且數量不多的類別效果有限，常常擷取出過於特定的狹義詞，造成類別描述詞主題涵蓋面不夠廣泛的結果。為解決此問題，本步驟的實際作法，是將出現篇數過低的詞彙（低於類別篇數的二分之一者）排除，並以乘積法則，將此相關係數與詞彙在該類別中的出現次數相乘，作為新的類別代表性係數，再依此排序，並取前數個詞彙，以獲得較常見、涵蓋面較廣的類別描述詞。因此，若一個類別的描述詞是空的（因篇數過低而被排除掉），表示此類別的文件所使用的詞彙大多不同，因而難有一致的詞彙可被選為類別代表詞。

在實作時，為正確統計篇數，本步驟還對英文詞彙進行大小寫轉換，並修改 Porter (1980) 取詞幹 (stem) 的演算法，只進行複數詞彙轉單數詞彙、動名詞轉原形等處理。例如 size、sizes、sizing 這三種詞彙，都轉化成 size 一詞。

七、多面向資料交叉分析

一旦非結構的主題資訊擷取出來後，可結合其他既有的結構化資料（如發表年代、作者、期刊出處、引用文獻等），建構交叉分析與統計資訊，以提供多樣性的判讀素材。此種交叉分析資訊，比起單獨一項資料的統計，常能提供更豐富的解讀素材。例如，可從中得知重要的作者在不同主題發表的論文篇數（即其專長分布），而不只得知這些作者的生產力（論文總篇數）。甚至研究領域中有哪些重要的作者（Who）在做哪些主題（What）、何時開始（When），以及其熱門趨勢（emerging trends，例如最近幾年才在這個領域出現的主題或是關鍵詞），都可經由多面向的交叉分析獲得相關訊息。

一個領域的研究發展趨勢，經常是各個研究者關心的對象。本文採用 Tseng、Y.-I. Lin、Y.-Y. Lee、Hung 與 C.-H. Lee (2009) 所印證的趨勢指標，來量化各主題的研究趨勢。此指標其實是年代篇數序列的線性迴歸斜率，其計算公式如下：

$$slp = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} \quad \text{其中 } x_i = i - \frac{1}{n} \sum_{i=1}^n i \text{ 且 } y_i = d_i - \frac{1}{n} \sum_{i=1}^n d_i \quad \text{公式(3)}$$

在 n 個時間區段中，假設某主題的第 i 個區段有論文 d_i 篇，則 $[d_1, d_2, \dots, d_n]$ 為該主題的篇數時間序列，依照上述公式計算出來的線性迴歸斜率 slp ，可代表此主題的趨勢指標，其值愈大，代表愈晚近有愈多篇數。

與其他趨勢指標比較，在不同領域、不同規模的文件集、各種時間區間、甚至考量個別期刊有不同重要性的多種情況下，線性迴歸斜率被印證是表現最穩定的優良趨勢指標 (Tseng et al., 2009)。以此指標來比較各主題的研究趨勢，可以提供有用的參考資訊。

八、結果呈現與視覺化

上述主題偵測、多階段歸類的結果，僅是一堆電腦化的數據，需要以視覺化的工具，配合其他資訊一起呈現，才能使其方便解讀，甚至快速瞭解各類別間的關係、觀察到電腦化數據所無法顯露的訊息。本步驟以兩種視覺化方式呈現結果：「主題樹」(Topic Tree)和「主題地圖」(Topic Map)。主題樹是根據前述凝聚歸類與多階段歸類結果，配合類別描述詞自動擷取以及多面向資料交叉分析，將所有結果以 HTML 標示的網頁形式呈現，以方便檢視各個主題群落的細節。主題地圖則是以前述多維縮放技術製作出來的。本步驟特別增強此技術的特性，在地圖中除了以圓圈（而非單一點）來代表一群被歸類在一起的文件，還以圓圈大小來代表該群文件的篇數多寡。相關的結果範例，將於後續詳述。

九、比較與印證

上述的分析方法、流程與結果呈現，能否得出有效或有意義的結果，是值得關切的議題。為此，除了建議以書目對與共現字兩種不同的方法來互相比較外，針對教育評鑑領域的分析結果，本文也參考 Alkin (2004) 的評鑑理論樹 (evaluation theory tree) (可參見網頁電子檔：http://www.sagepub.com/upm-data/5074_Alkin_Chapter_2.pdf)，來進行結果的比對，以印證系統自動偵測主題的可靠程度。

Alkin 為 UCLA 大學 (University of California, Los Angeles) 教育與資訊研究所的名譽教授，他以 40 年的經驗，整理評鑑領域的各種理論、方法與應用，並以評鑑理論樹將各種面向、主題關聯起來。他將著名的學者，依照其研究的重點與貢獻，呈現在該主題樹上，以方便大家追溯各種理論與實務的來龍去脈。Alkin 並邀請了 22 位當代的評鑑學者，描述其自身的評鑑觀點、評論其在評鑑理論樹中的定位、以及探討各種觀點對他們的影響。透過這樣的分類體系與知識關聯，提供評鑑學者與實務者新的啟發、開創新的概念，且得以藉此檢討加強既有的認知 (Alkin, 2004)。

肆、結果與討論

根據前述方法，本節分三小節說明：一、核心期刊的篩選過程與結果，並說明篩選的依據與考量；二、單一相關期刊的探勘經驗，解說主題樹與主題地圖的細節與意涵，其中並包含書目對與共現字方法的互相比較與印證；三、六種核心期刊的主題歸類與交叉分析，並以領域專家的知識來驗證結果的正確性。

一、核心期刊篩選

本研究的一項重點，是經由資料庫搜尋和專家問卷，篩選出教育評鑑領域的核心期刊，

以供後續研究的參考與運用。其篩選步驟分兩階段進行：首先，從數篇教育評鑑領域導論性論文（簡稱種子文獻，詳如後述）的引用參考文獻中，蒐集與教育評鑑內容相關的期刊，並連同其被資料庫收錄、國內師範大學訂閱的情況，作為核心期刊判斷的參考條件，如表 2 所示（最後一欄除外）。其次，根據這份列表，邀請國內相關的專家學者，進行核心期刊的判斷。表 3 臚列受邀學者名單，其判斷結果如表 2 最後一欄所示。在進行表 2 的判斷時，同時也請專家增補重要期刊，並列舉重要學者。表 4 顯示專家們額外推薦的期刊名稱，表 5 則為專家所列舉的重要學者名單。

表 2 教育評鑑領域核心期刊學者專家判斷

編號	候選期刊	資料庫			文獻之 參考 期刊	國內師 範大學 圖書館 收錄之 資訊	專家 判斷 (請勾 選出核 心期 刊)
		EBSCOhost (Professional Development Collection, PDC) (篇數)	Web of Science (WoS) (篇數)	EBSCOhost (ERIC) (篇數)			
1	<i>Educational Assessment</i>	161		111	v	1	3
2	<i>Assessment & Evaluation in Higher Education</i>	683		297	v	1,2,3	8
3	<i>Journal of Education Policy</i>	395	458	464		1,2	1
4	<i>Educational Policy</i>	593	611	449		1,2	1
5	<i>Drugs: Education, Prevention & Policy</i>	362		4		1	0
6	<i>Educational Management Administration & Leadership</i>	133					0
7	<i>Perspectives: Policy & Practice in Higher Education</i>	249				1	0
8	<i>Educational Administration Quarterly</i>	402	528	765		1,2,3	1
9	<i>Journal of Higher Education Policy & Management</i>	206			v	1	2
10	<i>Assessment in Education: Principles, Policy & Practice</i>	341				1	4
11	<i>Educational Research & Evaluation</i>	235		3		1	4
12	<i>Arts Education Policy Review</i>	549		257		1,2,3	0

表 2 (續) 教育評鑑領域核心期刊學者專家判斷

編號	候選期刊	資料庫			文獻之 參考 期刊	國內師 範大學 圖書館 收錄之 資訊	專家 判斷 (請勾 選出核 心期 刊)
		EBSCOhost (Professional Development Collection, PDC) (篇數)	Web of Science (WoS) (篇數)	EBSCOhost (ERIC) (篇數)			
13	<i>Evaluation & Research in Education</i>	64		28		1,2	4
14	<i>Educational Management & Administration</i>	339		287	v	1,2	2
15	<i>Higher Educational Quarterly</i>				v		0
16	<i>Educational Researcher</i>			1,219	v	1,2,3	2
17	<i>Studies in Educational Evaluation</i>			737	v	1,2,3	7
18	<i>Educational Evaluation and Policy Analysis</i>		322	730	v	1,2,3	9
19	<i>American Journal of Evaluation</i>	508	493	255	v	1	7
20	<i>Evaluation Review</i>	413	577	637	v	1,2,3	9
21	<i>Evaluation Practice</i>		140	159	v	1,2,3	6
22	<i>Evaluation and Program Planning</i>		867	537	v	1,2,3	7
23	<i>Evaluation</i>			6,058	v		7
24	<i>Evaluation Comment</i>			14	v		2
25	<i>New Directions for Program Evaluation</i>				v	2	4
26	<i>Evaluation Researcher</i>				v		1
27	<i>New Directions for Evaluation</i>				v	1,2	5
28	<i>Evaluation Studies Review Annual</i>				v	1,2,3	5
29	<i>Journal of Personnel Evaluation in Education</i>		74	293	v	1,2,3	6
30	<i>Evaluation News</i>			6	v		1

表 3 教育評鑑領域核心期刊專家學者問卷名單（依姓氏筆畫排序）

學者專家	服務學校／系所（2007年7月）
丁一顧教授	臺北市立教育大學教育行政與評鑑研究所
王保進教授	臺北市立教育大學教育行政與評鑑研究所
王麗雲教授	國立臺灣師範大學 教育學系／教育評鑑研究中心資料組組長
吳清山教授	臺北市立教育大學教育行政與評鑑研究所 財團法人高等教育評鑑中心基金會執行長
郭昭佑教授	臺灣藝術大學師資培育中心
張德銳教授	臺北市立教育大學教育行政與評鑑研究所
劉春榮教授	臺北市立教育大學教育行政與評鑑研究所
潘慧玲教授	國立臺灣師範大學教育學系／教育評鑑研究中心副主任
蘇錦麗教授	國立新竹教育大學教育學系

表 4 專家推薦增補之重要期刊刊名

項次	推薦增補之期刊名稱	推薦專家	WoS篇數
1	<i>International Journal of Educational Research</i>	蘇錦麗教授	
2	<i>Journal of Educational Research</i>	蘇錦麗教授	661
3	<i>Journal of Multi-disciplinary Evaluation</i>	蘇錦麗教授	
4	<i>Quality in Higher Education</i>	王保進教授	
5	<i>Research Evaluation</i>	王保進教授	174

以下詳細說明表 2 中各欄位內容的意義：

（一）候選期刊

由種子文獻所引用參考的資料中，蒐集與教育評鑑內容相關的期刊，作為候選期刊。

（二）資料庫

從下列資料庫中搜尋候選期刊被收錄的情況（空白者表示沒收錄）：

1. WoS（Web of Science）資料庫

提供理、工、醫、農、社會科學等各學科領域之文獻書目、作者摘要及引用文獻等資料。該系統收錄期刊超過 9,000 種，每年提供超過 110 萬筆書目及 2,300 萬筆引用文獻，其中包

表 5 專家列舉教育評鑑領域重要學者（重要文獻作者共 42 位）

Alkin, M. C.	Davidson, J.	Leviton, L. C.	Preskill, H.	Tyler, R. W.
Boruch, R. F.	Eisner, E. W.	Lincoln, Y. S.	Provus, M.	Weiss, C. H.
Campbell, D.	Fetterman, D. M.	MacDonald, B.	Rossi, P. H.	Wholey, J. S.
Chelimsky, E.	Guba, E. G.	Millman, J.	Sanders, J. R.	Wingate, L. A.
Chen, H.-T.	Hopkins, D.	Nevo, D.	Scriven, M.	Wolf, R. L.
Cook, T. D.	House, E. R.	Owen, J. M.	Shadish, W. R.	Worthen, B. R.
Cousins, J. B.	Kellaghan, T.	Owens, T.	Stake, R.	
Cronbach, L. J.	Kells, H. R.	Patton, M. Q.	Stufflebeam, D. L.	
Darling-Hammond, L.	King, J. A.	Popham, W. J.	Suchman, E.	

註：此份名單為潘慧玲與蘇錦麗教授提供

含 SCI (Science Citation Index)、SSCI (Social Sciences Citation Index) 與 AHCI (Arts & Humanities Citation Index)。

2. PDC (Professional Development Collection) 資料庫

專為教育工作人員設計，收錄了分科細密的五百二十種高品質教育期刊，還包括近三百五十個同儕評鑑標題。本資料庫亦收錄逾 200 份教育報告，堪稱為最專業的教育期刊集合。

3. ERIC (Educational Resource Information Center) 資料庫

收藏逾 1,000 種教育相關的期刊及其他教育參考資料。

(三) 種子文獻之參考期刊

本研究參考的教育評鑑導論性論文，亦即候選期刊的種子文獻，主要為一本教育評鑑模式論文集編，以及數篇教育評鑑概論方面之文章（王保進，2002；吳清山、王湘栗，2004；郭昭佑，2006；黃德祥、薛秀宜，2004；劉春榮，2004；潘慧玲，2005；蘇錦麗譯，2005），其主要作者包括：王保進、吳清山、郭昭佑、黃德祥、劉春榮、潘慧玲、A. Steinmetz、D. L. Stufflebeam、G. F. Madaus、H. S. Bholā、L. J. Cronbach、M. Scriven、R. E. Stake、R. W. Tyler、T. Kellaghan 等（按姓氏筆畫、字母順序排序）。

(四) 國內師範大學圖書館收錄之資訊

指候選期刊在國內北、中、南 3 所師範大學圖書館於 2007 年 6 月被收錄的情形。其中：「1」代表國立臺灣師範大學；「2」代表國立高雄師範大學；「3」代表國立彰化師範大學。

(五) 專家判斷

表 2 的最後一欄數值，為表 3 中 9 位專家判斷該期刊為核心期刊的累計次數。數值愈高，表示愈多專家認為此期刊乃重要的核心期刊。此專家判斷於 2007 年 7 月進行，表 3 中專家的服務單位為當時的現況。

為了決定最後的核心期刊，除了累計專家的判斷外，本文也統計了表 5 中的重要學者在這些期刊中發表論文的篇數。表 6 列出被專家判斷為核心期刊最多次（超過五次）的九種期刊，以及重要學者在其中發表論文的總和篇數（檢索時間為 2007 年 8 月）。從表 6 中，本研究決定了編號 18-22 以及 29 這六種期刊，作為本次分析的核心期刊。其中編號 17 的期刊，與教育評鑑主題非常相關，重要學者的發表數量也多，但由於沒有被 WoS 資料庫收錄而無法進行書目對分析，因此只好將其排除。

表 6 專家判斷累計次數超過五次之期刊與重要學者在其中發表篇數總和統計

編號	候選期刊	資料庫			專家判斷 累計 次數	重要學 者發表 文章篇 數總和
		PDC (篇數)	WoS (篇數)	ERIC (篇數)		
2	<i>Assessment & Evaluation in Higher Education</i>	683		297	8	2
17	<i>Studies in Educational Evaluation</i>			737	7	40
18	<i>Educational Evaluation and Policy Analysis</i>		322	730	9	25
19	<i>American Journal of Evaluation</i>	508	493	255	7	68
20	<i>Evaluation Review</i>	413	577	637	9	23
21	<i>Evaluation Practice</i>		140	159	6	30
22	<i>Evaluation and Program Planning</i>		867	537	7	15
23	<i>Evaluation</i>			6,058	7	1
29	<i>Journal of Personnel Evaluation in Education</i>		74	293	6	17

與教育評鑑領域有關的期刊，可能有三十五種（三十種作者整理的候選期刊，以及五種專家增補的期刊），但在考量資料的可得性與篩選合理性之情況下，最後選擇以下六種：*Educational Evaluation and Policy Analysis* (EEPA)、*American Journal of Evaluation* (AJE)、*Evaluation Review* (ER)、*Evaluation Practice* (EP)、*Evaluation and Program Planning* (EPP)、*Journal of Personnel Evaluation in Education* (JPEE)。至於其他未被本研究選為核心期刊者，上述各種表格的資訊，仍可供個別應用時的參考。

在 Thomson Reuters 的 Journal Citation Report (JCR) 資料庫中，EEPA 屬於 Education & Educational Research 領域，其共包含一百十三種期刊。EEPA 於 2008 年的影響指數

(impact factor) 為 1.133，排名前 25 名。而 AJE、ER、EPP 則屬於 JCR 的 Social Sciences, Interdisciplinary 領域，其包含六十一種期刊，此三種期刊於 2008 年的影響指數分別為：1.104、1.051、0.735，排名各為：11、13、28。至於 EP 只於 1994-1997 年這 4 年間被 WoS 收錄，而 JPEE 也只於 1989-1992 年這 4 年間被 WoS 收錄，因此，在 JCR 中都找不到。

二、單一期刊分析

在進行核心期刊的分析前，本文先對唯一屬於 Education & Educational Research 領域的 EEPA 期刊進行分析，以瞭解在狹義的範圍中，教育評鑑研究主題的特色。EEPA 自 1992 年起，始收錄於 WoS，截至 2005 年底，共有 318 篇論文，其年代篇數資料，如表 7 所示。依表 7 中的年代篇數序列，取其線性迴歸線來預測趨勢，所得回歸線之斜率為-1.1736，顯示整體研究趨勢呈現緩步下滑。為何如此，可能和此期刊採編政策有關，或與研究人口、經費、主題熱門程度有關，需專家進一步瞭解與解讀。

表 7 EEPA 期刊的 318 篇論文年代篇數統計

年代	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05
篇數	30	28	31	29	22	24	21	24	22	18	17	23	17	12

依照上一節研究方法的說明，針對這些論文，以書目對進行三階段歸類的結果，如圖 3 的主題樹所示；而依照共現字三階段歸類的結果，如圖 4 的主題樹所示。在圖 5 中，本研究將兩種歸類結果可解讀的部分（如圖 3、圖 4 中被圈起來的部分），翻譯成中文詞彙，然後將其並列，再把相似的主題連線起來，以方便分析比較。

1. 143 : 46筆 : 0.006 (**leadership: 6.7, school: 5.1, principal: 3.7, teacher: 3.5, reform: 2.6**)
 - 11筆 : 0.025397 (**assessment: 1.5, kentucky: 1.4, wale: 1.4, england: 1.4**)
 - 7筆 : 0.020 ()
 - 4筆 : 0.010 (wale: 1.4, england: 1.4, autonomous: 0.7)
 - 35筆 : 0.040 (**leadership: 9.0, teacher: 3.9, school: 3.3, management: 3.0**)
 - 28筆 : 0.044 (**leadership: 6.4, educational leadership: 2.8, teacher: 2.8**)
 - 15筆 : 0.011 (teacher: 1.9, performance: 1.2, involvement: 1.1, instructional: 0.9, school: 0.7)
 - 13筆 : 0.023 (case: 1.2, model: 0.7, reform: 0.5, education: 0.3)
 - 7筆 : 0.017 (school: 0.7)

圖3 根據書目對分析的歸類結果

2. 189 : 41筆 : 0.004 (**size: 7.8, school: 4.1, resource: 2.1, effect: 2.1, special: 2.0**)
- 37筆 : 0.019 (**size: 9.4, effect: 3.5, school: 3.3, special: 2.4, social: 2.1**)
 - 31筆 : 0.045 (~~class size: 6.9, special: 3.0, achievement: 2.4, tennessee: 2.3~~)
 - 8筆 : 0.005 (title: 1.3, school: 0.9, achievement: 0.5)
 - 23筆 : 0.019 (size: 7.0, class: 6.4, class size: 6.2, effect: 1.6, achievement: 1.4)
 - 6筆 : 0.054 (size: 1.6, high: 1.4, high school: 1.0, schools-within-schools: 0.7)
 - 4筆 : 0.021 (bad: 0.7, religious-education: 0.7, choose: 0.7, drive: 0.7)
3. 116 : 67筆 : 0.009 (**mathematic: 7.1, school: 6.3, student: 4.7, state: 3.4, policy: 3.2**)
- 61筆 : 0.016 (**mathematic:9.0, school:5.2, student:3.7, middle:3.4, high: 3.4**)
 - 53筆 : 0.056 (**mathematic:6.8, high school:5.6, test:3.6, state:3.5, graduation:3.4**)
 - 37筆 : 0.005 (high: 6.0, mathematic: 5.2, school: 2.9, student: 2.6, secondary: 2.1)
 - 16筆 : 0.015 (reform: 4.3, state: 3.3, practice: 2.0, instruction: 1.4, mathematic: 1.2)
 - 8筆 : 0.014 (**retention:2.4, kindergarten: 1.4, sex: 1.4, grade retention: 1.4**)
 - 6筆 : 0.023 (value: 1.4, teacher: 1.2, politic: 0.8, receive: 0.7, aid: 0.7)
4. 168 : 18筆 : 0.005 (**evaluation:3.3, expectation:1.6, program: 1.5, validity: 1.3, outcome: 0.7**)
- 14筆 : 0.022 (**validity:1.6, leigh:1.4, evaluation:1.1, assessment:1.1, outcome: 1.0**)
 - 8筆 : 0.011 ()
 - 6筆 : 0.054 (gender: 1.4, difference: 1.4, science: 1.4, test: 1.0, approximation: 0.7)
 - 4筆 : 0.040 (evaluation: 1.6, research: 0.8, demonstration: 0.7, federally: 0.7, nature: 0.7)
5. 93 : 21筆 : 0.011 (**opportunity to learn: 3.2, international: 3.2, state: 2.6, policy: 1.8, policy-making: 1.6**)
- 17筆 : 0.048880 (**opportunity to learn: 4.0, international: 4.0, mathematic: 1.4, mathematic and science: 1.4, implement: 1.4**)
 - 10筆 : 0.009 (system: 1.0, policy: 0.3)
 - 7筆 : 0.023 (learn: 1.0)
 - 4筆 : 0.037 (examine: 0.7,innovation: 0.7, centralization: 0.7, education: 0.7, prominence: 0.7)
6. 12 : 17筆 : 0.036 (**voucher: 3.0, choice: 2.8, empowerment: 2.4, parent: 2.1, competition: 2.0**)
- 4筆 : 0.015 (ideal:0.7, school-site:0.7, zealand:0.7, principal-agent:0.7, school-based: 0.5)
 - 13筆 : 0.019 (~~choice: 3.4, school: 2.4, public: 1.5, voucher: 1.1, national: 0.6~~)
7. 6 : 19筆 : 0.046 (**certification:7.0, teacher:4.5, teacher certification:4.0, matter:2.0, value: 1.6**)
- 5筆 : 0.076 (~~certification: 4.1, teacher certification: 1.0, certify: 0.7, traditionally: 0.7~~)

圖3 (續) 根據書目對分析的歸類結果

- 14筆 : 0.011 (teacher: 5.0, school: 0.3)
- 8. 132 : 9筆 : 0.007 (**gift: 1.4, employer: 1.4, cost-analysis: 0.7, signal: 0.7, school-to-work: 0.7**)
 - 4筆 : 0.031 (employer: 2.0, academy: 0.7, signal: 0.7, scale: 0.7, ged: 0.7)
 - 5筆 : 0.013 (gift: 1.4, analysi: 1.2, equal: 0.7, cost-analysis: 0.7, education: 0.7)
- 9. 242 : 4筆 : 0.014 (**testimony: 1.0, reframe: 1.0, litigation: 1.0, democratic community: 1.0**)
- 10. 241 : 4筆 : 0.014 (**help: 1.0, diego: 1.0, affirmative: 1.0, attrition: 1.0, increase: 1.0**)
- 11. 239 : 5筆 : 0.011 (**child-care: 2.0, poverty: 1.4, family: 1.4, preschool: 1.4, shadow: 1.0**)
- 12. 240 : 9筆 : 0.013 (**community college: 4.0, remedial education: 2.0, price: 2.0, remedial: 2.0, proprietary school: 2.0**)

圖3 (續) 根據書目對分析的歸類結果

- 1. 179 : 71筆 : 0.147 (**high school: 84.5, teacher: 25.9, policy: 22.3, grade: 19.6, new: 14.8**)
 - 67筆 : 0.232 (**high school: 97.4, size: 41.6, class: 35.5, small: 21.0**)
 - 56筆 : 0.331 (**high school: 79.9, size: 48.9, class: 41.7, teacher: 29.8**)
 - 44筆 : 0.502 (**size: 61.0, high school: 60.5, class: 52.0, small: 32.1**)
 - 26筆 : 0.233 (high school: 84.6, student: 30.5, course: 27.8, math: 22.1, ...)
 - 18筆 : 0.152 (teacher: 18.3, design: 15.7, class: 12.4, effect: 9.7, positive: 7.3)
 - 12筆 : 0.326 (principal: 39.0, leadership: 15.7, school: 6.2, change: 5.2, policy: 4.1)
 - 11筆 : 0.223 (college: 39.0, student: 4.4, education: 2.2)
 - 4筆 : 0.176 (high-stakes: 5.6, skill: 4.2, high-stakes test: 4.2, salary: 3.5, plan: 3.5)
 - 2. 88 : 25筆 : 0.236 (**fund: 22.1, district: 20.6, finance: 16.8, new: 12.8, school: 9.2**)
 - 19筆 : 0.386 (**finance: 21.0, district: 13.9, reform: 11.8, equity: 10.9, school district: 10.4**)
 - 6筆 : 0.342 (equity: 7.7, school finance: 7.7, debate: 4.9, attention: 4.2, system: 3.6)
 - 13筆 : 0.316 (district: 14.3, reform: 10.8, school district: 10.3, spend: 7.2) opportunity
 - 6筆 : 0.443 (federal: 14.0, title: 8.8, fund: 6.2, evaluation: 5.3, schoolwide: 4.9) to learn
 - 3. 67 : 38筆 : 0.263 (**test: 32.2, assessment: 27.4, performance: 20.0, otl: 16.8, accountability: 15.2**)
 - 27筆 : 0.308 (**assessment: 37.5, test: 23.2, otl: 21.0, accountability: 19.0, score: 17.4**)
 - 17筆 : 0.187 (assessment: 41.8, test: 8.4, performance: 7.0, national: 6.2)
 - 10筆 : 0.205 (score: 14.1, test: 6.4, hands-on: 4.9, science: 1.6, high: 1.6)
 - 11筆 : 0.160 (quality: 15.5, work-based: 5.6, teacher: 5.2, learn: 2.3, program: 1.9)

圖4 根據共現字分析的歸類結果

4.	13 : 38筆 : 0.387 (teacher: 35.2, certification: 22.5, practice: 20.5, policy: 20.0, mathematic: 16.4)
	○ 30筆 : 0.492 (teacher: 25.4, instructional: 18.4, knowledge: 16.0, content: 13.2, practice: 12.5)
	■ 11筆 : 0.184 (content: 15.5, leadership: 9.0, subject: 5.0, mathematic: 3.6)
	■ 19筆 : 0.117 (teacher: 19.8, read: 12.9, instructional: 10.4, policy: 9.9, practice: 5.2)
	○ 8筆 : 0.321 (apply: 11.0, model: 3.0, national: 2.9, policy: 2.2, school: 1.7)
5.	207 : 21筆 : 0.129 (social: 16.1, rate: 9.7, education: 8.5, higher: 4.9, system: 4.1)
	○ 16筆 : 0.263 (social: 22.0, indicator: 17.4, capital: 9.7, social capital: 8.3, cohesion: 7.6)
	■ 9筆 : 0.234 (characteristic: 7.9, state: 6.4, level: 5.4, opportunity: 4.0, district: 4.0)
	■ 7筆 : 0.330 (social: 17.6, capital: 8.4, social capital: 7.7, social cohesion: 6.3)
	○ 5筆 : 0.180 (reviewer: 9.8, productivity: 7.0, external reviewer: 6.3, research: 5.7)
6.	8 : 12筆 : 0.441 (segregation: 14.6, group: 13.8, ethnic: 9.0, racial: 7.6, achievement: 6.9)
	○ 4筆 : 0.399 (ability: 8.9, ability group: 5.6, middle: 4.2, political: 3.9, honor: 3.5)
	○ 8筆 : 0.308 (difference: 12.8, california: 4.5, science: 3.3, achievement: 2.5, test: 2.0)
7.	1 : 8筆 : 0.541 (succe: 14.4, massachusetts: 7.6, bonu: 5.6, sign: 4.9, institution: 4.9)
	○ 4筆 : 0.483 (succe: 7.5, institution: 4.2, budget: 3.9, defeate: 2.8, statistically: 2.8)
	○ 4筆 : 0.374 (massachusetts: 8.1, bonu: 4.9, sign: 4.2, enrollment: 4.2, intervention: 3.9)
8.	32 : 16筆 : 0.324 (retention: 46.0, preschool: 20.9, children: 18.6, family: 11.2, service: 10.9)
	○ 12筆 : 0.212 (children: 29.8, preschool: 19.0, special education: 9.7, special: 9.7, lower: 7.3)
	○ 4筆 : 0.329 (voucher: 4.9, family: 4.9, program: 3.1, service: 3.0, educational voucher: 2.8)
9.	39 : 15筆 : 0.308 (evaluation: 18.0, program: 6.9, decision: 6.5, teacher: 6.4, question: 6.1)
	○ 9筆 : 0.261 (question: 10.2, evaluation: 8.3, data: 5.0, research: 3.5, program: 1.7)
	○ 6筆 : 0.269 (participation: 7.7, decision: 5.1, program: 3.8, willingne to participate: 3.5)
10.	83 : 9筆 : 0.242 (systemic: 8.3, change: 8.3, proprietary school: 6.3, problem: 5.6, growth: 4.2)
	○ 4筆 : 0.204 (proprietary: 4.9, proprietary school: 3.9, growth: 3.9, metric: 3.5, sery: 3.5)
	○ 5筆 : 0.499 (systemic: 6.3, problem: 4.9, change: 4.7, frame: 3.5, policy change: 3.5)
11.	12 : 12筆 : 0.398 (implementation: 9.6, case: 8.0, school reform: 7.6, testimony: 6.3, rule: 6.3)
	○ 8筆 : 0.239 (reform: 13.1, school reform: 10.0, adoption: 5.6, cms: 4.9, foundation: 4.9)
	○ 4筆 : 0.383 (testimony: 5.6, rule: 5.6, intermediary: 5.6, expert: 5.6, case: 5.6)
12.	35 : 21筆 : 0.317 (choice: 31.3, resource: 20.5, parent: 13.9, public school: 13.2, charter: 10.4)
	○ 6筆 : 0.343 (resource: 9.3, acce: 7.8, model: 7.2, gift: 5.4, additional: 2.9)
	○ 15筆 : 0.359 (choice: 44.0, school: 12.9, public: 9.2, private: 2.9, education: 2.5)

圖4 (續) 根據共現字分析的歸類結果

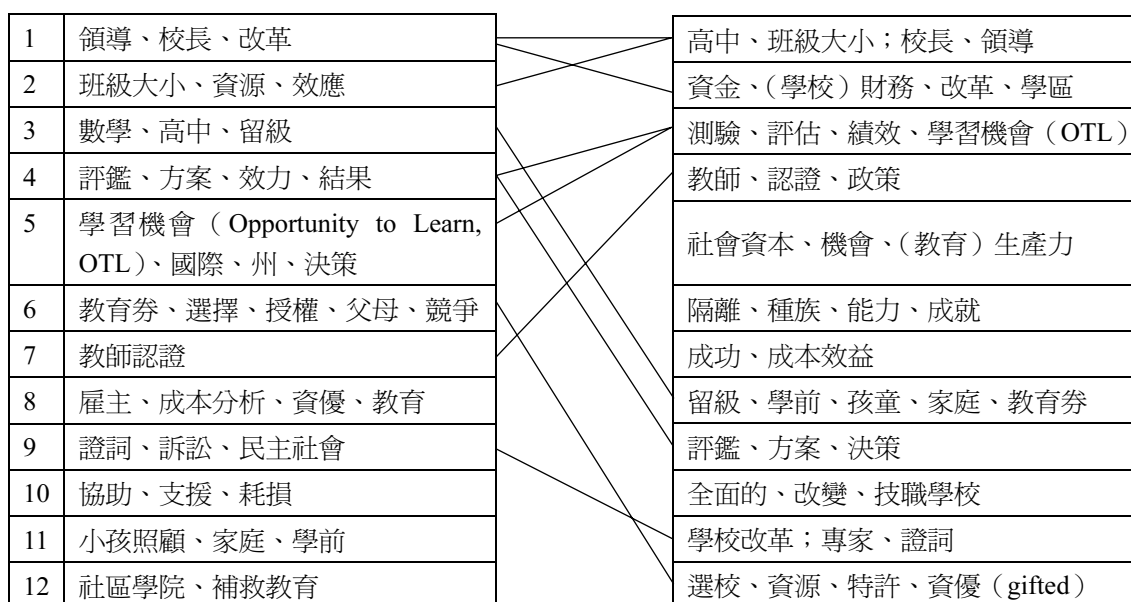


圖5 兩種歸類結果的比較

註：左邊為書目對結果，右邊為共現字結果

這些主題樹所代表的意義，以圖 3 第一棵主題樹為例，其第一行「143：46 筆：0.006 (leadership: 6.7, school: 5.1, principal: 3.7, teacher: 3.5, reform: 2.6)」，分別代表此類別的內部 (程式給的) 編號為 143，共包括 46 篇文章，是由各含 11 篇與 35 篇文章的兩個小主題所組成，且此兩個小主題之間的相似度不低於 0.006。另外，此類別內的 46 篇文章中，最具代表性的關鍵詞彙為 leadership，代表性係數為 6.7，其次為 school，代表性係數為 5.1 等等。

從實務的角度看，這些自動偵測的主題，除印證本研究對教育評鑑領域的一般認知外 (如「教師認證」、「評鑑」、「改革」)，也提供額外的主題項目與數據 (如篇數)，因此這些結果具有相當的參考性。然而，從研究的角度，要如何驗證這些自動化分析結果的正確性或有效性？

從圖 3、4 可知，不同的文件相似度定義 (書目對與共現字)，會得出不同的歸類結果。這些相似度都有其主題近似上的意義，因此這兩種結果，不難看出其共通性，如圖 5 中相同的詞彙：「校長、領導」、「班級大小」、「評鑑」、「改革」、「教師認證」、「學習機會」、「留級」、「教育券」等。這種共通性，顯示這群文件中確實存在這些主題的可能性極高，才有可能經由不同的途徑而獲致相似的結果。

雖然本研究沒有尋求專家的協助，來閱覽、分析、解讀這群文件的主題，以致於無法完全確認上述共通主題類別的存在性 (有效性)，但即便專家也會有主題分析看法不一致的情形，而難以完全互相確認彼此的結果。像這樣透過不同的自動方法而獲得的共同主題，就好像同時問了 2 位看過這些文件的人 (或專家)，他們各自有主題相似度判斷的標準 (各自有不同的專長領域和程度)，可是卻從他們之口中，獲得相同的結果。在類似的研究中 (如資訊檢

索的相關判斷)，這些一致的專家判斷結果，常被視為（較為嚴格的）「標準答案」來看待。

從圖 3、4 的異同，也可觀察到每一種主題歸類的方法，都有其優點與缺點之處。例如，依照書目對來歸類，主要是靠論文之間的引用情況來計算相似度。其明顯的好處是，論文 A 與 B 的用語可以差異極大，即使其共現字極罕見，只要 A 與 B 共同引用足夠多的相同文獻，則 A 與 B 即有足夠高的相似度。反之亦然，若論文 A 與 B 探討相同的主題，卻因為某些因素而沒有相同的參考書目時，A 與 B 就不被書目對的關係對應起來。此時，若 A 與 B 共用一些主題詞彙，則他們就可被共現字的關係對應起來。也就是說，書目對與共現字的分析可以互補，也可以彼此強化對方的結果。因此，單一方式偵測到的主題也值得探究。這包括利用書目對得到的：「補救教育」，以及利用共現字得到的：「社會資本」、「種族教育問題」等主題。

表 8 與表 9 分別顯示書目對與共現字的主題歸類結果與年代交叉分析的篇數資料，其中第一欄為主題編號，第二欄為該主題內的篇數，第三欄則根據第四欄到第十七欄的年代篇數序列，以線性迴歸方式計算的趨勢斜率，此斜率可作為趨勢指標來預測該主題的熱門程度。表中第二列為這 12 大主題的篇數累計，其總和不到 318 篇的原因，乃多階段歸類過程中，取相似度門檻將低相似度者過濾不計造成的結果。此種結果在科學計量中相當普遍，因為雖然有眾多文獻在探討主流議題，仍會有為數不少的文獻在探討獨立的小議題，而無法與其他文獻歸類在一起。此種現象也常可用網路書店銷售統計曲線的長尾理論來理解。

表 8 書目對歸類結果的主題年代交叉分析與趨勢指標

Cluster	df	Trend	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05
Total	260	-0.8659	22	20	28	25	17	23	16	20	14	16	16	20	12	11
1	46	-0.3341	5	4	11	4	2	4	3	0	2	1	1	5	3	1
2	41	0.1692	2	0	2	1	1	4	1	9	6	4	4	4	2	1
3	67	0.0462	6	3	3	5	6	7	3	5	2	6	6	6	3	6
4	18	-0.3165	3	5	2	3	1	2	1	1	0	0	0	0	0	0
5	21	-0.2440	1	3	2	7	3	1	1	1	0	0	0	1	1	0
6	17	-0.0330	3	1	1	1	1	0	1	0	3	0	3	3	0	0
7	19	-0.0813	1	2	3	1	1	1	2	2	1	2	1	1	1	0
8	9	-0.0462	0	0	2	1	1	2	1	0	0	1	0	0	1	0
9	4	0.0044	1	0	1	0	0	0	0	0	0	0	0	0	1	1
10	4	0.0044	0	1	0	0	0	0	1	1	0	0	0	0	0	1
11	5	-0.0154	0	1	0	0	1	0	1	0	0	2	0	0	0	0
12	9	-0.0198	0	0	1	2	0	2	1	1	0	0	1	0	0	1

註：粗體者顯示趨勢下滑之主題

表 9 共現字歸類結果的主題年代交叉分析與趨勢指標

Cluster	df	Trend	'92	'93	'94	'95	'96	'97	'98	'99	'00	'01	'02	'03	'04	'05
Total	286	-0.9714	28	26	27	25	18	20	17	23	21	17	16	21	15	12
1	71	-0.0549	6	4	1	3	8	8	3	10	8	6	4	6	2	2
2	25	-0.1297	1	0	7	3	1	3	2	1	3	0	0	1	2	1
3	38	-0.2813	4	4	2	8	1	3	4	3	2	3	2	1	0	1
4	38	0.0747	2	3	4	2	3	2	1	1	2	4	3	4	5	2
5	21	-0.1604	3	3	1	4	1	1	1	1	1	1	1	2	1	0
6	12	-0.0791	3	1	1	0	0	2	1	1	1	0	0	0	0	2
7	8	0.0264	0	1	2	0	0	0	1	0	0	0	1	0	2	1
8	16	0.0132	2	1	1	1	1	0	2	1	0	2	1	1	0	3
9	15	-0.2264	4	3	3	1	0	1	0	1	0	0	1	0	1	0
10	9	-0.0813	1	2	1	1	1	0	0	1	0	0	0	2	0	0
11	12	0.0308	0	2	0	1	1	0	0	1	3	0	1	1	2	0
12	21	-0.1033	2	2	4	1	1	0	2	2	1	1	2	3	0	0

註：粗體者顯示趨勢下滑之主題

表 8 中顯示，這 12 大類的總體趨勢指標為-0.8659，同前述，顯示整體研究趨勢呈現緩步下滑。其中各主題之間下滑得相對嚴重的類別有 1、4、5 三類（在表 8 中以粗體顯示），特別是第 4 類（評鑑、方案、效力、結果）與第 5 類（學習機會、國際、州、決策），這兩類在 2000 年以後幾乎沒有相關的論文出現。在表 9 中，總體趨勢斜率為-0.9714，各主題之間趨勢下滑相對嚴重（以粗體顯示）的有第 3 類（測驗、評估、績效、學習機會）與第 9 類（評鑑、方案、決策）。此結果顯示，在趨勢預測的結果中，書目對與共現字都一致得出「評鑑、方案、學習機會」是相對趨勢下滑的研究課題。在教育評鑑領域中，這些看似是主要的議題，居然論文數量愈來愈少，實在值得深入探究。不過他們的趨勢斜率都比總體趨勢斜率高，或許並非真正的愈來愈少人研究。無論如何，這些量化數據剛好可以作為研究假說（主要議題是否愈來愈少被探討）的重要支柱。

上述歸類的結果，除了以主題樹呈現外，也以多維縮放技術，計算類別之間在二維空間的相對關係，從而繪製出其主題地圖，如圖 6 所示。在主題地圖中每個有編號的圓圈代表一群被歸類在一起的文件，圓圈的大小代表該群文件的篇數多寡，圓圈的距離代表該群文件與另一群文件的關聯強度，距離愈短表示主題上的關聯性愈高。

此主題地圖與主題樹是互相對應的。在主題地圖中每個圓圈的編號，都對應到主題樹中一個主題的內部編號，如圖 6 書目對編號 189 的圓圈，其所代表的類別，是圖 3 中編號 189 的類別。此兩張圖互相對照，可以將文件之間的主題關係與知識結構呈現出來，方便專家的

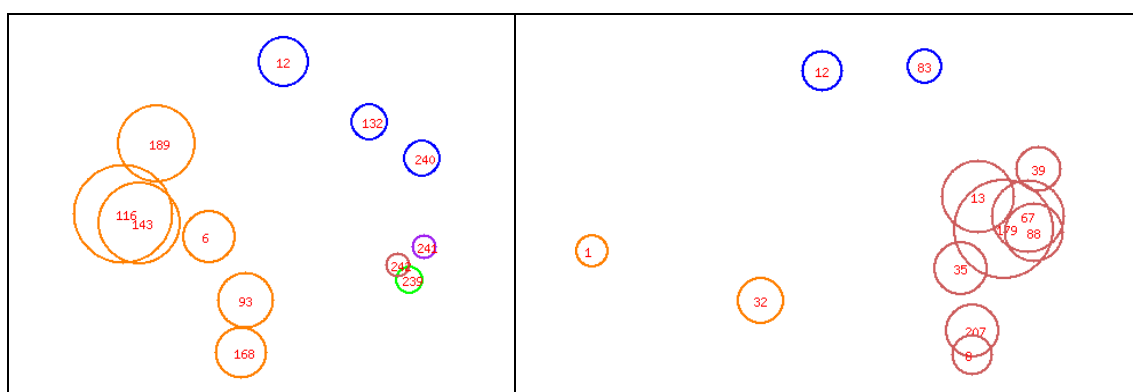


圖6 書目對（左）與共現字（右）的主題地圖（圓圈中的數值對應到主題樹的類別編號）

判別與分析。

例如，主題地圖中，編號 143 的群組（在主題樹中的主題為：領導、校長、改革等），與群組編號 189（主題為：班級大小、資源、效應）距離較近，因此主題上較相似，而這一點可在共現字的結果中得到印證。在共現字的結果中，這兩個議題被放在同一個主題中，如圖 4 第一類所示。又如，編號 168 的群組（在主題樹中的主題為：評鑑、方案、效力、結果、性別、差異等）與群組編號 93（主題為：學習機會、國際、州、決策）距離較近，因此主題上也較相似（學習機會和性別、差異有關，而決策與評鑑、方案有關），如有必要，其在更高階段的歸類，可再合併成更大的類別來一起分析。

比較書目對與共現字的主題地圖，大體上可知，他們都有較為離散的小議題，也都有較為相近的主流議題。不過共現字的主流議題看起來重疊性較多，顯示探討這些主流議題的文章，其用詞重複性高。換句話說，對主流議題而言，共現字不容易將他們歸類、劃分的較為清楚。

三、核心期刊分析

這一節以較廣泛的論文範圍來分析教育評鑑的研究議題，亦即根據前述所得的六種核心期刊，於 2007 年 8 月從 WoS 資料庫下載論文資料，並根據上述研究方法，進行資料的處理與運算。此六種核心期刊自 1990 年到 2006 年之間共出版 1,882 篇論文。這些論文經書目對分析後，可歸納為四百七十個類別。雖然電腦分析所花時間僅十數秒，但其歸類結果的資訊量龐大，無法全部列出，也無法一一探討。附錄中顯示其中一類的結果範例，此類別是關於「美國田納西州學校班級大小效應」的主題，自 1997 年起被討論，1999 年為討論的高峰期（共 7 篇論文），2002、2003 年各有 1 篇尚在討論此議題。在此主題中，Hedges 為發表最多論文的作者之一，有 2 篇，而且 Hedges 較早期的著作也是此主題中被引用最多次的論文之一，其 1994 年的論文被引用五次，1996 年的論文被引用三次。

若深入資料庫中查詢、追蹤，可發現 Hedges 在此類別中所發表的 2 篇論文，都跟 Nye 與 Konstantopoulos (Nye, Hedges, & Konstantopoulos, 1999, 2002) 一起發表，1 篇是 1999 的“The long-term effects of small classes”，另一篇是 2002 年的“Do low-achieving students benefit more from small classes?”，這 2 篇的書目對相似度指標為 0.4746，比其他書目對都還高，因此在歸類結果中被放在一起，如附錄中第一欄的結果。

圖 7 顯示 Alkin 整理的評鑑理論樹，其將評鑑理論大致分為使用、方法以及評價三方面，並將各方面代表性的學者依照時代列於各樹幹中。比較表 5 與圖 7 中的專家名單，Hedges 並沒有列名其中。再回頭查詢資料庫，Hedges 也只在 2004 年多發表了 1 篇：“How large are teacher effects?”，的確稱不上重量級的人物，但在班級大小的相關議題中，以其論文的發表量與被引用量而言，Hedges 是值得注意的人物。此訊息在 Alkin 的書中與國內 9 位專家提供的資料中並沒有顯示，但透過本研究分析方法稍加留意後，可以輕易觀察出來。顯示自動化的方法，可以輔助專家，歸納整理出更多後續值得觀察的人物與現象。

此四百七十個類別中，最大的一類有 35 篇論文，其分析結果摘要如表 10 所示。由系統自動擷取的描述詞中，可看出此類最常出現 evaluation、theory、practice、performance 等詞，因此可看成是關於評鑑理論的主題。相較於其他四百多個類別而言，此主題年代分布長，最後幾年篇數多，其趨勢指標 0.3333 大於整體的 0.2353，因此可視為是趨勢漸增之主題。另外，從其作者與引文的作者名單，可看到此主題的學者（粗黑體字者），大多為圖 7 中的重要學者，而且落在評鑑理論樹中最左邊的枝幹上，如圖 7 中圓短虛線所示。此訊息可驗證此類別的確是關於評鑑理論或其基礎研究的主題。而從期刊出處，大多數論文出自 *American Journal of Evaluation* 此一期刊，顯示此期刊在評鑑理論中的核心地位。

表 11 與表 12 分別為第二及第三大類的資訊摘要表，雖然兩者的類別描述詞一樣，均為方案評鑑，但其所引用的主要文獻卻不同。表 11 的類別中，每一篇都有引用到 Cook 在 1979 年的文章，此外 Compbell 的論文被引用的次數合計也有六次，而這兩人在圖 7 中（細實線者）被 Alkin 分類為評鑑方法的學者。因此本類可視為是方案評鑑方法的主題，其年代分布顯示相關的研究歷久彌新，是可以持續關注的主題。

表 12 的主題中，其 17 篇論文最常引用的是 Chen 的文章，其次是 Weiss、Cook 與 Rossi。同樣地，這些人在圖 7（長虛線）的評鑑理論樹中，也都屬於評鑑方法的領域，不過是在樹幹中較高的位置。從其年代分布中也可看出，此主題較表 11 中的主題發生較晚，可能在探討較新的評鑑方法。

這兩類的結果顯示，雖然其探討的主題非常接近（主要的用詞相同），但透過書目對的歸類，仍可細微區別其差異。然而，此結果同時也顯示，若非有 Alkin 的專家知識輔助，外行人實難理解自動化系統為何會分出類別名稱一樣的這兩類，因而可能會降低對系統的信心。

雖然不瞭解 Alkin 是如何發展出其評鑑理論樹的，但從上述的觀察與討論過程中，我們可

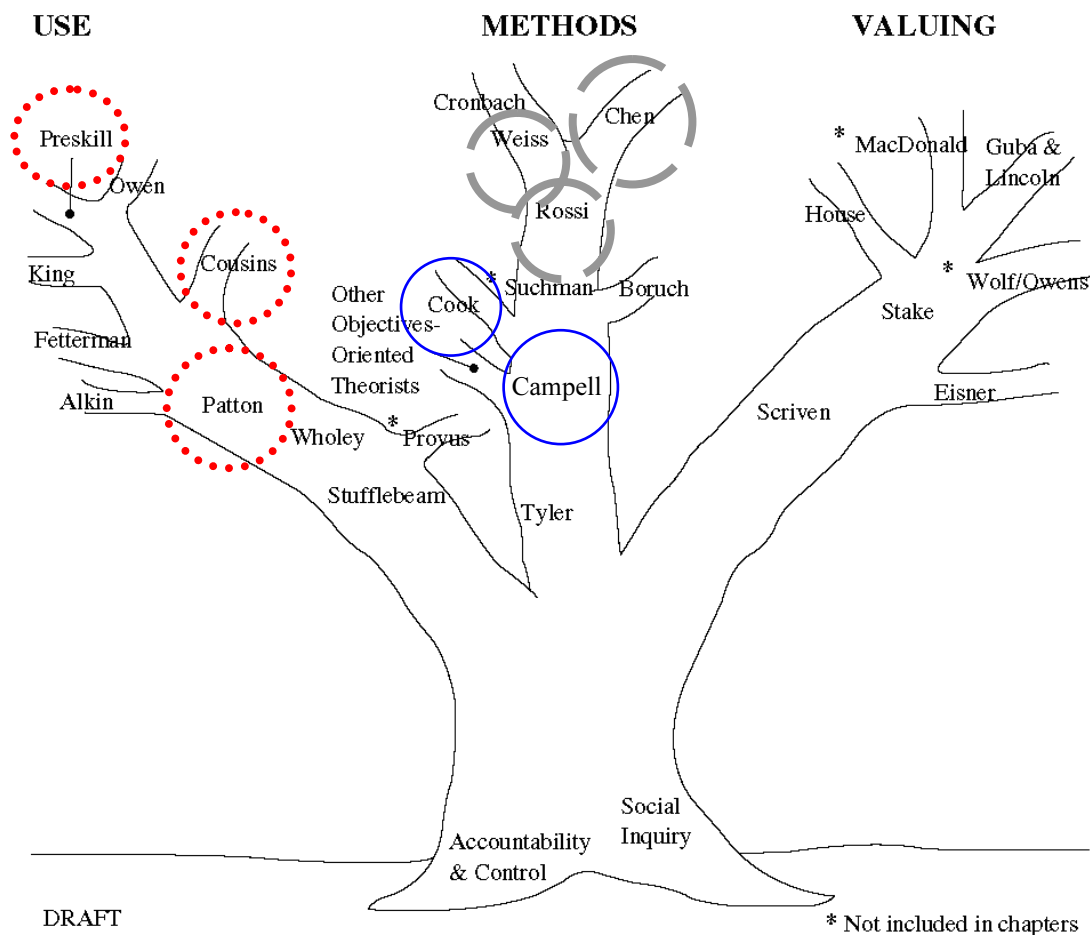


圖7 Alkin的評鑑理論樹

註：此圖中被同形式圓圈圈起來的作者，乃被本系統歸在同一類者

資料來源：Alkin 與 Christie (2004, p. 13)

以感受到，若要将此理論樹繼續擴展、茁壯，上述的結果資訊與分析過程應可提供有價值的材料與線索。例如，持續觀察表 10 中尚未被列名於圖 7 中的學者，深入查詢、追蹤其著作與研究實績後，也許可以在圖 7 中安置新的學者在更高的樹幹上。如此，此理論樹或主題樹，在領域專家們因其他工作負荷一時難以支援時，仍得以透過系統的輔助，由圖書資訊編目者持續地擴展、演變，繼續造福後來的學者。

表 10 書目對歸類結果，第一大類類別資訊分析

類別篇數與描述詞：35筆：theory, evaluation, practice			
年代分布	期刊出處	作者篇數	引文次數
1. 1996: 1	1. AMERICAN	1. Patton, MQ: 4	1. PATTON MQ, 1997, UTILIZATION
2. 1998: 3	JOURNAL OF	2. Fishman, DB: 3	FOCUSED: 35
3. 1999: 3	EVALUATION:	3. Cousins, JB: 2	2. SHULHA LM, 1997, EVAL PRACT,
4. 2000: 3	21	4. Neigher, WD: 2	V18, P195: 6
5. 2001: 5	2. EVALUATION	5. Michalski, GV: 2	3. PATTON MQ, 2002,
6. 2002: 3	AND	6. Preskill, H: 2	QUALITATIVE RES EVAL: 6
7. 2003: 4	PROGRAM	7. Renger, R: 2	4. ROSSI PH, 1993, EVALUATION
8. 2004: 9	PLANNING: 12	8. Kaltreider, DL: 1	SYSTEMATI: 5
9. 2005: 4	3. EVALUATION	9. Fort, L: 1	5. SHADISH WR, 1991, FDN
	PRACTICE: 1	10. Perrin, B: 1	PROGRAM EVALUATI: 5
	4. EVALUATION	11. Church, M: 1	6. COUSINS JB, 1992, EDUC EVAL
	REVIEW: 1	12. Rowan, MS: 1	POLICY AN, V14, P397: 5
		13. Shadish, WR: 1	7. LEVITON LC, 1981, EVALUATION
		14. Klein, BM: 1	REV, V5, P525: 4
		15. Stevenson, JF: 1	8. GUBA EG, 1989, 4 GENERATION
		16. Egherman, T: 1	EVALUAT: 4
		17. Rugh, J: 1	9. PRESKILL H, 1999, EVALUATIVE
		18. Chelimsky, E: 1	INQUIRY L: 4
		19. Cherniss, C: 1	10. CHELIMSKY E, 1997,
		(more are omitted)	EVALUATION 21 CENTUR: 4
			11. PRESKILL H, 1997, EVAL PRACT,
			V18, P209: 4
			12. WEISS CH, 1998, EVALUATION: 3
			13. GUBA E, 1989, 4 GENERATION
			EVALUAT: 3
			14. SCRIVEN M, 1991, EVALUATION
			THESAURUS: 3
			15. STAKE RE, 1995, ART CASE
			STUDY RES: 3
			16. PRESKILL H, 2000, NEW
			DIRECTIONS EVALU, V88, P25:3
			17. COUSINS JB, 1995,
			PARTICIPATORY EVALUA: 3
			18. ALKIN MC, 1998, ADV EDUC PR,
			V7, P87: 3
			19. COUSINS JB, 1998, NEW
			DIRECTIONS EVALU, V80, P5: 3
			(more are omitted)

表 11 書目對歸類結果，第二大類類別資訊分析

類別篇數與描述詞：22筆：program, evaluation			
年代分布	期刊出處	作者篇數	引文次數
1. 1991: 1	1. EVALUATION AND PROGRAM PLANNING: 12	1. Devine, JA: 2	1. COOK TD, 1979, QUASIEXPERIMENTATION: 22
2. 1992: 1		2. Brody, CJ: 2	
3. 1993: 2		3. Wright, JD: 2	
4. 1994: 1		4. Michael, PG: 1	
5. 1995: 3		5. Green, BL: 1	
6. 1996: 2		6. Zhu, SH: 1	
7. 1997: 3		7. PERRY, PD: 1	
8. 1998: 1		8. Collins, ME: 1	
9. 1999: 3		9. Campbell, DT: 1	
10. 2000: 1		10. CAMPBELL, DT: 1	
11. 2002: 2		(more are omitted)	
12. 2003: 1			
13. 2005: 1			

表 12 書目對歸類結果，第三大類類別資訊分析

類別篇數與描述詞：17筆：program, evaluation				
年代分布	期刊出處	作者篇數	引文次數	
1. 1996: 2	1. EVALUATION AND PROGRAM PLANNING: 7	1. Chen, HT: 5	1. CHEN HT, 1990, THEORY DRIVEN EVALUA: 17	
2. 1997: 3		2. Donaldson, SI: 2		
3. 1998: 2		3. Lin, LH: 1		
4. 1999: 3		4. Bledsoe, KL: 1		
5. 2000: 2		5. Mead, JP: 1		
6. 2001: 3		6. Sechrest, L: 1		
7. 2003: 1		7. Orwin, RG: 1		
8. 2005: 1		8. Mercier, C: 1		
		9. Gooler, LE: 1		
		10. Cordray, DS: 1		
		(more are omitted)		
				7. ROSSI PH, 1993, EVALUATION SYSTEMATI: 3
				8. WEISS CH, 1997, EVALUATION REV, V21, P501: 3
		9. WEISS CH, 1972, EVALUATION RES METHO: 3		
		10. CHEN HT, 1987, EVAL PROGRAM PLANN, V10, P95: 3 (more are omitted)		

另外，從四百七十類的類別描述詞中，也可以觀察到一些有趣的詞彙，如 Daniel empowerment evaluation、developmental evaluation、diffusion of innovation theory、evaluator competency、group evaluation、stakeholder-based evaluation 等，這些詞彙在 Scriven (1991) 的《評鑑索引典》中（含 847 詞）均未提及。因此，本研究自動從歸類結果中抽取類別特徵詞的方法與結果，也可用於未來相關索引典編製時的參考。

將上述書目對的四百七十個類別，經四次多階段歸類後，可以得到如圖 8 左圖的主題地圖。若以共現字進行多階段歸類，其主題地圖如圖 8 右圖所示。比較這兩個圖可知，教育評鑑（或評鑑）領域包含跨領域、多樣性的主題，但其用詞卻非常相似，以致於共現字方法幾乎把所有的文件放在同一大類中，然而，書目對方法卻因其引用文獻來源的不同，而將文件分配到各個類別裡。因此，針對教育評鑑領域的後續深入分析，建議以書目對方法為主。

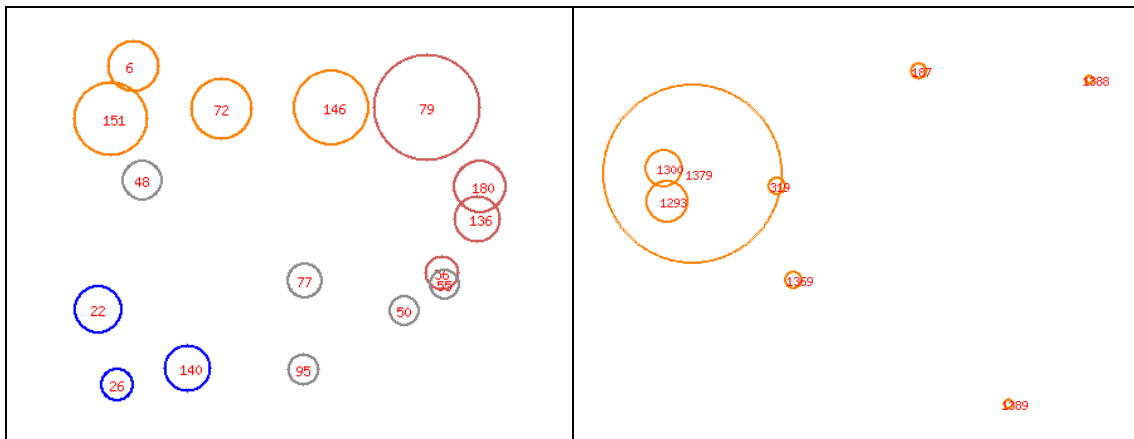


圖8 書目對（左）與共現字（右）的主題地圖

伍、結論與建議

本研究試圖從教育評鑑的研究文獻中找出核心期刊，並以引文分析與文字探勘等方法為基礎，提出一套文獻探勘模式，以探索教育評鑑領域中隱而未顯、潛在有用的資訊或知識。此探勘模式可以分析教育評鑑領域中的 What（哪些主題）、When（何時開始）、Who（有哪些重要的作者在進行研究），因而可以快速掌握教育評鑑各主題領域，以及其歷年來的發展軌跡。透過書目對與共現字方法的互相驗證，並與 Alkin 的評鑑理論樹比對後，驗證本研究文獻探勘模式與方法應用於教育評鑑領域之有效程度。

經由文獻分析、資料庫檢索與專家判斷的交相確認後，本研究歸納出 *Educational Evaluation and Policy Analysis*、*American Journal of Evaluation*、*Evaluation Review*、*Evaluation Practice*、*Evaluation and Program Planning* 這六種期刊可為評鑑領域核心代表期刊。至於沒有

列在本研究的核心期刊者，對不同的應用而言，某些期刊可能才是核心所在。無論如何，本研究的核心期刊確認方式與結果，可供後續相關的研究參考、比較。

透過書目對與共現字的相互印證，在 *Educational Evaluation and Policy Analysis* 期刊中，本研究找出了「校長、領導」、「班級大小」、「評鑑」、「改革」、「教師認證」、「學習機會」、「留級」、「教育券」等是較為明確的研究主題。透過主題趨勢分析，則發現教育評鑑裡的主要議題，其論文數量似乎有愈來愈少的趨勢。而在主題地圖中，本研究指出某些主題可能的關聯，從共現字的主題地圖中也發現，此領域使用的詞彙有限，不管研究什麼議題，主要的詞彙重疊性（重複性）頗高，亦即大家關心的內容相當集中在某些概念上。

針對六種核心期刊分析的結果，發現在從事方案評鑑（program evaluation）或評鑑方法（Methods）的學者中，至少包括了：Cook、Compbell、Chen、Weiss 與 Rossi 等重要學者，而且在評鑑方法的主題中，Chen、Weiss 與 Rossi 討論的方法，可能比 Cook、Compbell 的還新。從 Alkin 在主題樹中對專家的擺設位置，也可觀察到相同的現象。除此之外，Patton、Cousins 與 Preskill 等研究者，被本研究的方法歸類在一起，同時也被 Alkin 歸類在一起。此種一致性，顯示本研究的方法具有一定程度的效力。然而，由於此種資料與結果不具結構性，因此無法以統計考驗的方式，來驗證其統計上的意義。只能從輔助人工作業、降低分析工作成本的觀點來說明其有效性。

任何一個學科的發展，都有一些重要的研究主題、重要的研究者，以及研究發展的趨勢和脈絡。學科領域專家或許因為有多年的學術涵養，對該領域的研究主題演變可以了然於胸，如同 Alkin 可以畫出一棵引領後進的評鑑理論樹一樣。然而，並非每一個領域都能輕易找到這樣的專家。本研究方法的分析探勘結果，與 Alkin 的評鑑理論樹對照，可看出有一定的效果。因此，此方法應該也可以用在不同學科領域，繪製出該領域的主題樹或主題演進圖，以瞭解該學科有哪些重要的研究主題、研究脈絡，甚至標示出重要學者的位置，從而形成以證據為基礎的（evidence-based）假說，提供進一步探索、研究的素材。

而即使達不到上述效果，其所提示的資訊也具有增補專家既有知識之功效。例如，較新的專有名詞：Daniel empowerment evaluation、developmental evaluation、diffusion of innovation theory、evaluator competency、group evaluation、stakeholder-based evaluation 等，可用於修訂該領域的索引典。又如，Hedges（或 Nye、Konstantopoulos）在班級大小的相關議題中，也許是值得注意的人物，因而可以提供 Alkin 的理論樹，針對較晚進的議題、人物，進行擴展、茁壯的材料與線索。

本研究方法（第三節）的步驟二到步驟八，已寫成一套名為 CATAR（Content Analysis Toolkit for Academic Research）的工具程式（Tseng, 2010b），其先前版本，已陸續應用於谷佳臻（2008）、許育聞（2009）、曾元顯（2007, 2009）、Y.-H. Chang、C.-Y. Chang 與 Tseng（2010）等人的研究或計畫之中，以便從中獲得各種試驗、印證、回饋與修正。本研究乃此分析技術

最成熟、最完整的呈現，以及全新的應用。未來將繼續擴大應用範圍，以深入瞭解其實務應用上的各種效應，從中累積經驗，並尋求精進改善之處，以提供各領域的實務者與研究者更為有效地運用。

誌謝

本文由國立臺灣師範大學教育研究與評鑑中心（計畫編號：96E0012-04-03）以及行政院國家科學委員會補助部分經費（計畫編號：NSC96-2221-E-003-017-、NSC99-2631-S-003-002-），並由丁一顧、王保進、王麗雲、吳清山、郭昭佑、張德銳、劉春榮、潘慧玲、蘇錦麗等教授（依姓氏筆畫排序）協助，才得以完成，謹此誌謝。

參考文獻

一、中文文獻

- 王保進 (2002, 10 月)。國民中小學校務評鑑現況與重要議題之省思。論文發表於國立教育資料館、臺北市立師範學院聯合舉辦之「現代教育論壇－校務評鑑」學術研討會，臺北市。
- 【Wang, B.-C. (2002, October). *Current status of evaluation of elementary and junior high school and reflection on important issues*. Paper presented at the meeting of Modern Education Forum, Taipei, Taiwan.】
- 村田忠禧 (2002)。從《人民日報》元旦社論看中華人民共和國的歷史。2007 年 7 月 22 日，取自 <http://www.japanresearch.org.tw/forum-12.asp>
- 【Murata, T. (2002). *A perspective on the PRC's history based on the People's Daily editorials on the New Year's Days*. Retrieved July 22, 2007, from <http://www.japanresearch.org.tw/forum-12.asp>】
- 何光國 (1994)。文獻計量學導論。臺北市：三民。
- 【He, K.-K. (1994). *Introduction to bibliometrics*. Taipei, Taiwan: San Min Book.】
- 谷佳臻 (2008)。電腦輔助分析軟體運用於質性研究訪談稿內容分析之探討。國立臺灣師範大學圖書資訊學研究所碩士論文，未出版，臺北市。
- 【Ku, C.-C. (2008). *A study on the application of computer assisted qualitative data analysis software to qualitative research interview manuscript content analysis*. Unpublished master's thesis, National Taiwan Normal University, Taipei, Taiwan.】
- 吳清山、王湘栗 (2004)。教育評鑑的概念與發展。教育資料集刊，29，1-25。
- 【Wu, C.-S., & Wang, S.-L. (2004). The conception and development of educational evaluation. *Bulletin of National Institute of Education Resources and Research*, 29, 1-25.】
- 郭昭佑 (2006)。當評鑑遇上教育－教育評鑑意涵探究。教育行政與評鑑學刊，2，19-42。
- 【Kau, C.-Y. (2006). When evaluation meets education: The exploration of the meaning for educational evaluation. *Journal of Educational Administration and Evaluation*, 2, 19-42.】
- 許育聞 (2009)。會議與期刊文獻對預測主題趨勢之比較研究－以「資訊檢索」領域為例。國立臺灣師範大學圖書資訊學研究所碩士論文，未出版，臺北市。
- 【Hsu, Y.-W. (2009). *A comparison study on conference papers and journal articles for predicting topic trends: Using "Information Retrieval" as an example*. Unpublished master's thesis, National Taiwan Normal University, Taipei, Taiwan.】
- 曾元顯 (2007)。專利主題萃取之研究開發。財團法人工業技術研究院委託報告。臺北市：國立臺灣師範大學。
- 【Tseng, Y.-H. (2007). *Research and development of topic extraction from patents*. Unpublished research report for Industrial Technology Research Institute. Taipei, Taiwan: National Taiwan Normal University.】
- 曾元顯 (2009)。科學發展趨勢調查－科學地圖製作。財團法人國家實驗研究院科技政策研究與資訊中心委託報告。臺北市：國立臺灣師範大學。
- 【Tseng, Y.-H. (2009). *Survey of scientific development: A science mapping approach*. Unpublished research report for the Science & Technology Policy Research and Information Center of National Applied Research Laboratories. Taipei, Taiwan: National Taiwan Normal University.】

- 黃德祥、薛秀宜 (2004)。教師評鑑的模式與發展趨勢。《教育研究月刊》，127，18-32。
【Huang, D.-H., & Hsueh, H.-Y. (2004). Models and development of teachers' evaluation. *Journal of Education Research*, 127, 18-32.】
- 劉春榮 (2004)。教育評鑑方案之建構。《教育資料集刊：教育評鑑專輯》，29，119-128。
【Liu, C.-R. (2004). Construction of educational evaluation program. *Bulletin of National Institute of Educational Resources and Research: Special Issue of Educational Evaluation*, 29, 119-128.】
- 潘慧玲 (主編) (2005)。《教育評鑑的回顧與展望》。臺北市：心理。
【Pan, H.-L. (Ed.). (2005). *Retrospection and prospection of educational evaluation*. Taipei, Taiwan: Psychological Publishing.】
- 蔡明月 (2003)。《資訊計量學與文獻特性》。臺北市：華泰文化。
【Tsai, M.-Y. (2003). *Informetrics and literature characteristics*. Taipei, Taiwan: HuaTai Cultural.】
- 蘇錦麗 (主編) (譯) (2005)。《評鑑模式：教育及人力服務的評鑑觀點》(D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan 著, *Evaluation models: Viewpoints on educational and human services evaluation*, 2nd ed.)。臺北市：高等教育。(原著出版於1983年)
【Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (Eds.). (2005). *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed.) (G.-L. Su, Trans., Ed.). Taipei, Taiwan: Higher Education. (Original work published 1983)】

二、外文文獻

- Alkin, M. C. (2004). *Evaluation roots: Tracing theorists' views and influences*. London: Sage.
- Alkin, M. C., & Christie, C. A. (2004). *An evaluation theory tree*. Retrieved October 21, 2009, from http://www.sagepub.com/upm-data/5074_Alkin_Chapter_2.pdf
- Andres, A. (2009). *Measuring academic research: How to undertake a bibliometric study*. New York: Neal-Schuman.
- Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3-17.
- Chang, Y.-H., Chang, C.-Y., & Tseng, Y.-H. (2010). Trends of science education research: An automatic content analysis. *Journal of Science Education and Technology*, 19(4), 315-331.
- Feldman, R., & Sanger, J. (2006). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge, UK: Cambridge University Press.
- Golub, G. H., & Van Loan, C. F. (1996). *Matrix computations* (3rd ed.). Baltimore, MD: Johns Hopkins.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). London: Psychology Press.
- Hearst, M. (2003). *What is text mining?* Retrieved January 15, 2010, from <http://people.ischool.berkeley.edu/~hearst/text-mining.html>
- Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Boston, MA:

- Addison-Wesley.
- Jajuga, K., Sokolowski, A., & Bock, H. H. (2002). *Classification, clustering, and data analysis: Recent advances and applications (Studies in classification, data analysis, and knowledge organization)*. New York: Springer.
- Kohonen, T. (2000). *Self-organizing maps* (3rd ed.). New York: Springer.
- Krippendorff, K. H. (2003). *Content analysis: An introduction to its methodology* (2nd ed.). London: Sage.
- Kruskal, J. B. (1977). Multidimensional scaling and other methods for discovering structure. In K. Enslein, A. Ralston, & H. S. Wilf (Eds.), *Statistical methods for digital computers* (pp. 296-339). New York: Wiley.
- Lasswell, H. D. (1964). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas* (pp. 37-51). New York: Cooper Square.
- Lent, B., Agrawal, R., & Srikant, R. (1997, August). *Discovering trends in text databases*. Proceedings of the 3rd International Conference of Knowledge Discovery and Data Mining, California.
- Leydesdorff, L. (2001). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications* (2nd ed.). Boca Raton, FL: Universal.
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht, the Netherlands: Springer.
- Noyons, E. C. M., & van Raan, A. F. J., (1998). *Mapping scientometrics, informetrics, and bibliometrics. CWTS Working papers*. Retrieved January 2, 2008, from <http://shara.fsw.leidenuniv.nl/ed/sib/home.html>
- Nye, B., Hedges, L. V., & Konstantopoulos, S. (1999). The long-term effects of small classes: A five-year follow-up of the Tennessee class size experiment. *Educational Evaluation and Policy Analysis, 21*(2), 127-142.
- Nye, B., Hedges, L. V., & Konstantopoulos, S. (2002). Do low-achieving students benefit more from small classes? Evidence from the Tennessee class size experiment. *Educational Evaluation and Policy Analysis, 24*(3), 201-217.
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14*(3), 130-137.
- Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Boston, MA: Addison-Wesley.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.
- Tseng, Y.-H. (2002). Automatic thesaurus generation for Chinese documents. *Journal of the American Society for Information Science and Technology, 53*(13), 1130-1138.

- Tseng, Y.-H. (2010a). Generic title labeling for clustered documents. *Expert Systems With Applications*, 37(3), 2247-2254.
- Tseng, Y.-H. (2010b). *Content Analysis Toolkit for Academic Research (CATAR)*. Retrieved August 10, 2010, from <http://web.ntnu.edu.tw/~samtseng/CATAR/>
- Tseng, Y.-H., Lin, C.-J., & Lin, Y.-I. (2007). Text mining techniques for patent analysis. *Information Processing and Management*, 43(5), 1216-1247.
- Tseng, Y.-H., Lin, Y.-I., Lee, Y.-Y., Hung, W.-C., & Lee, C.-H. (2009). A comparison of methods for detecting hot topics. *Scientometrics*, 81(1), 73-90.

附錄 類別資訊範例

下表中第一欄為根據書目對計算論文相似度所得出的歸類結果，共 10 篇論文被歸類在一起，底線字為論文的標題，粗黑體字為系統從論文標題與摘要中自動擷取的類別描述詞，用來幫助讀者瞭解該類別的主題內容；第二欄為此 10 篇論文發表的年代分布統計，其中 1997 年占了 7 篇；第三欄為發表的期刊統計，此 10 篇全部發表在同一期刊上；第四欄為論文作者的篇數統計，其中 Hedges 在此 10 篇中發表了 2 篇；第五欄為引用文獻的次數統計，其中第一篇文獻在此 10 篇論文中共被引用了八次。

書目對類別	年代分布	期刊出處	作者篇數	引文次數
<ul style="list-style-type: none"> ● 24171 : 10筆 : 0.0244 (class: 10.3, tennessee: 8.6, size: 7.8, school: 2.6, effect: 1.4) ○ 7021 : 5筆 : 0.0488 (tennessee: 5.4, class: 3.4, size: 2.6, experiment: 1.4, school: 0.4) ■ 179 : 2筆 : 0.2295 (reflection on analysis of the economic of school: 1.4, independent: 0.8, school resource: 0.7, star: 0.7, belief: 0.5) ■ <u>17: ISI: 000072404500002: 1997: Outcomes, incentives, and beliefs: Reflections on analysis of the economics of schools</u> ■ <u>172: ISI: 000082043900005: 1999: Some findings from an independent investigation of the Tennessee STAR experiment and from...</u> ■ 655 : 3筆 : 0.1311 (tennessee: 3.9, class: 2.6, class: 2.1, small: 2.1, size: 1.9) ■ 19 : 2筆 : 0.4746 (class: 2.5, small: 2.5, tennessee: 2.1, do low- achieving student benefit more from small class: 1.4, evidence from the tennessee class size experiment: 1.4) 	1. 1997: 1 2. 1999: 7 3. 2002: 1 4. 2003: 1	1. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 2. EVALUATION AND POLICY ANALYSIS: 10 3. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 4. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 5. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 6. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 7. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 8. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 9. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 10. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 11. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 12. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 13. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 14. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 15. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10 16. EDUCATIONAL EVALUATION AND POLICY ANALYSIS: 10	1. Hedges, LV: 2 2. Konstantopoulos, S: 2 3. Nye, B: 2 4. Hanushek, EA: 2 5. Ritter, GW: 1 6. Grissmer, D: 1 7. Finn, JD: 1 8. Cohen, DK: 1 9. Shkolnik, JL: 1 10. Ball, DL: 1 11. Boruch, RF: 1 12. Betts, JR: 1 13. Bassi, LJ: 1 14. Raudenbush, SW: 1 15. Achilles, CM: 1 16. Ludwig, J: 1	1. FINN JD, 1990, AM EDUC RES J, V27, P557: 8 2. MOSTELLER F, 1995, FUTURE CHILD, V5, P113: 8 3. HANUSHEK EA, 1996, DOES MONEY MATTER EF, P43: 5 4. HEDGES LV, 1994, EDUC RES, V23, P5: 5 5. WORD E, 1990, STUDENT TEACHER ACHI: 5 6. GLASS GV, 1979, EDUC EVAL POLICY AN, V1, P2: 4 7. FERGUSON RF, 1996, HOLDING SCH ACCOUNTA, P265: 3 8. *ED RES SERV, 1980, CLASS SIZ RES CRIT R: 3 9. MOSTELLER F, 1996, HARVARD EDUC REV, V66,

(續)

書目對類別	年代分布	期刊出處	作者篇數	引文次數
<u>of small classes: A five-year follow-up of the Tennessee class size...</u>				P797: 3
■ <u>583: ISI: 000180508300003: 2002: Do low-achieving students benefit more from small classes? Evidence from the Tennessee...</u>				10. COLEMAN JS, 1966, EQUALITY ED OPPORTUN: 3
■ <u>170: ISI: 000082043900003: 1999: The political and institutional origins of a randomized controlled trial on elementary...</u>				11. NYE BA, 1994, LASTING BENEFITS STU: 3
○ 10655 : 5筆 : 0.0404 (class: 3.8, size: 2.9, resource: 2.2, school: 1.6, effect: 0.7)				12. HEDGES LV, 1996, DOES MONEY MATTER EF, P74: 3
■ 3058 : 4筆 : 0.0708 (class: 3.0, size: 2.2, school: 1.1, resource: 0.9, implication: 0.7)				13. GLASS GV, 1978, METAANALYSIS RES REL: 3
■ 808 : 2筆 : 0.1212 (assess the evidence: 1.4, misconception: 1.4, class size effect: 1.4, implication: 1.0, agenda: 0.8)				14. HANUSHEK EA, 1997, EDUC EVAL POLICY AN, V19, P141: 3
■ <u>169: ISI: 000082043900002: 1999: Tennessee's class size study: Findings, implications, misconceptions</u>				15. GREENWALD R, 1996, REV EDUC RES, V66, P361: 3
■ <u>177: ISI: 000082043900010: 1999: Class size effects: Assessing the evidence, its policy implications, and future research...</u>				16. GOLDSTEIN H, 1998, BRIT EDUC RES J, V24, P255: 3
■ 1847 : 2筆 : 0.0870 (puzzle: 1.4, class: 0.9, specification test: 0.7, size: 0.7, case: 0.6)				(More are omitted)
■ <u>175: ISI: 000082043900008: 1999: The behavioral effects of variations in class size: The case of math teachers</u>				
■ <u>246: ISI: 000085247000003: 1999: The puzzling case of school resources and student achievement</u>				
■ <u>666: ISI: 000185750800001: 2003: Resources, instruction, and research</u>				

Journal of Research in Education Sciences

2011, 56(1), 129-166

The Application of Content Mining Techniques to the Analysis of Educational Evaluation Research Trends

Yuen-Hsien Tseng

Information Technology Center,
National Taiwan Normal University
Research Fellow

Yu-I Lin

Department of Education,
Taipei Municipal University of Education
Graduate Student

Abstract

Due to the mass accumulation of digital documents in nowadays research environment, related statistical or content mining methods for analyzing, organizing, and accessing information from the document collections are in great need. This research aimed to: (1) identify the core journals in the field of educational evaluation research through cross-verification of questionnaire survey and database searching; (2) find out major research topics and their evolutions by analyzing the publications in the core journals based on scientometrics methods such as citation analysis and text mining techniques. Based on the findings, the major salient research topics include principal's leadership, class size, evaluation, reform, teacher certification, opportunity to learn, retention, and voucher. And as to the issue of the evaluation methods, the studies conducted by Chen, Weiss, and Rossi are identified as newer ones than those by Cook and Compbell, which coincides with the fact shown in the evaluation theory tree (a sort of topical tree) organized by Marvin Alkin, an experienced expert in the field. Besides, with this topic trend analysis, it is also found that the number of papers dealing with the major research topics seems to be decreasing. However, all these results need to be interpreted by domain experts to reveal their true meanings. The methodology presented in this research may be applied not only to improve the efficiency of content analysis, but also to facilitate regular tracing of topic evolution, to spot the positions of important researchers in the field, and therefore to provide relevant clues or materials for further studies.

Keywords: topic detection, scientometrics, core journals, evaluation theory tree, trend analysis