

國內教育學門系統性文獻評閱策略及 後設分析發展現況與建議

謝進昌¹

國家教育研究院籌備處助理研究員

摘要

以證據為本的教育理念，逐漸受到國內外許多學者的推廣，而建立此理念的最佳方法之一，是屬援用系統性文獻評閱策略；有鑑於本策略在國內教育學門或期刊很少被論述，因此，本研究目的在介紹系統性評閱策略及後設分析的意涵，並分析本策略在國內教育學門的發展與可能面臨困境；接續，借鏡國外教育與醫學領域所延伸建置之系統性評閱合作組織，提出因應建議。最後，在引用系統性評閱策略以建立證據為本教育理念時，存在許多值得省思與探討之議題，亦會於文末提出，以期作為國內實務與研究者之參考。

關鍵字：系統性文獻評閱、系統性評閱組織、後設分析、證據為本的教育

¹. 本文作者要特別感謝國科會補助本專題研究案，編號：98-2511-S-656-001-MY3。任何有關本論文的接洽，請與本研究案通訊作者聯繫：國家教育研究院籌備處助理研究員
謝進昌 E-mail: jin@mail.naer.edu.tw

壹、前言

隨著網路、科技的發展，訊息流通與知識累積的速度不斷加快，雖有利於知識傳遞與教育推展，然而一般民眾、教育實務工作者、學者，甚至是教育決策者，亦往往容易被大量、複雜的資訊所淹沒，而無所適從，若面臨重大決策時，只好逕自根據片面證據或當下的壓力，進行重大議題的決策，其正當性及可行性則易受到質疑；因此，講求以證據為本的教育(evidence-based education)理念逐漸受到許多國內外學者的鼓吹(馬信行，2002；Davies, 1999；Hargreaves, 1996, 1997)。

以證據為本位政策聯盟(Coalition for Evidence-Based Policy, 2002)早先提出的報告就曾表示，美國教育政策的制訂經常是在沒有嚴格證據的情況下進行，教育的實踐有如隨著流行喜好而改變，但矛盾的是，這喜好又常缺乏嚴謹證據與有說服力的研究支持，同時，實踐與政策的不相融，導致在過去三十年間，根據 NAEP(National Assessment of Educational Progress)的評測結果，發現儘管在公立學校每位學童有將近 90%經費支出的增長，但自幼稚園至十二年級，學生的數學、閱讀和科學成就幾乎沒有顯著的進步。隨著警訊的提出，促使著能對教育政策發揮影響力的科學研究需求，亦與日遽增。No Child Left Behind Act (U.S., Department of Education, 2002a)認為可以考慮以科學為本位的研究(scientifically-based research)來形塑方案或實務政策；而 Institute for Education Sciences (U.S. Department of Education, 2002b)支持使用嚴謹的方法，並以其研究結果導引政策與實務的執行方針；在美洲大陸的彼岸，以證據為本教育政策的發展，重要的是什麼才是有效的(What counts is what works)(Davies, Nutley, & Smith, 2000)的口號就明顯的表達出對此理念的支持主張。

在政府對於責任績效(accountability)的重視與學術界對證據的倡導外，加諸社會大眾知識水平的提高、訊息隨資訊科技的進步而更容易獲得、研究社群規模與素質的成長等因素影響下(Davies, Nutley, & Smith, 2000)，都直接或間接促進以證據為本風潮的興起，但此現象已引起許多學者對現行教育研究的批評，他們認為不論是在美國或英國，這個領域已經包含太多與實務關聯不足、品質低落或僅是學術社群討論，無法分享或被其它常需要這些證據的人所獲得的研究(Gough, 2004; Hargreaves, 1996)。對此批評，逐漸喚起許多決策者對於能回答各項研究議題的證據需求。但什麼樣的證據才是值得信賴的，前述不斷強調的都是實證主義下的隨機控制實驗(randomized controlled trials)(Coalition for Evidence-Based Policy, 2002; Oakley, 2001)，強調什麼才是有效(what works)的策略或教學方法，但證據的來源是多元的，在教育情境，更重視的是脈絡、過程與決策者在此現況下，對於證據的判斷智慧，這類證據也可能是嚴謹的質性研究，這就如同 Caldwell(1994)所述，實證主義太偏重科學的客觀面，容易錯失另一半美麗、複雜的故事，只有透過方法論的多元主義才能挽回(1994, p.244)。

在面臨各式的訊息或個別證據，系統性評閱策略的內涵所強調的不僅是對既存研究結果的收集、評估、整合與當下提供決策運用，更期望長遠時，能系統性、不斷更新的建立具高品質與教育實務相關的研究或證據資料庫；此策略也許不是快速解決當前教育問題的特效藥(panacea)，但它卻能夠作為教育人員在面對各式教育政策或解決實務問題

時，作出專業判斷的原則與方針(Davies, 1999)。簡而言之，系統性評閱策略的推動即是欲提供整合性、長期性之研究證據，而系統性評閱策略的本質是能同時包含質性與量化的研究整合與評論，但為聚焦主題，本文的主軸是以量化的系統性評閱策略及其統計技術-後設分析為論述焦點，而研究者必須強調這並不意味著證據的來源是唯一量化導向的，鑑於現行存在的有效證據確是非常的稀少，造成教師或決策者無法將這些證據視為是一種資源，此策略若能系統性的推動將能提供實務工務者一種可行的證據選項，以引領教學實務或政策的推動。

在過去數十年間，國內對於以證據為本的教育或政策理念、主張，已有為數不少的文章發表(如葉連祺、張盈霏，2001；賴志峰，2004)，但對於國外曾被廣泛推薦，以建立此理念的系統性評閱策略，在國內教育學門或期刊則很少被完整論述。據此，本研究目的在介紹系統性評閱策略及後設分析意涵，並分析國內教育學門對此策略的發展與可能面臨困境，進而，期望借鏡國外教育與醫學領域的進步，提出因應之建議。最後，在援用系統性評閱策略以建立證據為本教育理念上，仍有許多值得省思與探討的議題，亦會於文末提出，以期做為國內實務與研究工作者之參考。整體而言，作者希望能有效提供現行國內後設分析學者所面臨難題之因應策略，雖然，借鏡於國外經驗是有其實務推動所必須面臨的課題，但經由本文的倡導，若能成功踏出第一步，依相關理念基礎，逐步建立國內教育領域系統性評閱組織的雛形，則將是此領域很大的進展，不僅能提昇後設分析研究的能見度及社群，更能提供國內教育決策更多長期性與系統性的實務證據。

貳、系統性評閱與後設分析策略之意涵

茲針對系統性評閱與後設分析策略之歷史演變及其概念、流程，逐一陳述如下：

一、系統性評閱與後設分析策略之歷史演變與概念

二十世紀中以前，後設分析或稱整合分析與系統性評閱其實是兩個類似，但卻又不相同的概念。直到 Glass(1976)擔任美國教育研究學會主席時，始提出後設分析(meta analysis)一詞，意指為了整合研究發現而蒐集整理個別研究主題中的相關各式分析結果以進行統計分析的過程(p. 3)。往後的發展中，後設分析亦多集中在統計分析技術的精進或於各實務領域的運用。

系統性評閱(systematic review)可溯自 Feldman(1971)強調對文獻的系統性評閱與整合(systematic reviewing and integrating)的概念，並提出評閱過程的四個主要步驟。隨後，Taveggia(1974)更提出文獻評閱中可能產生的六個關鍵議題，分別是主題選擇，研究檢索、索引與編碼、分析研究結果的可比較性、累積可比較的研究發現、分析研究結果的分配、與結果呈現。

直到 1980 年代初期，Jackson(1980)與 Cooper(1982)兩篇研究的貢獻，才促使後設分析與將文獻評閱視為研究(review-as-research)的觀點加以融合運用。而就後設分析與系統性評閱的區分，Cooper 與 Hedge(1994, p.5)認為，後設分析僅是整個系統性評閱過程中，屬於統計分析的部份，並非代表整體。此外，Egger, Smith 與 Altman(2001)也有類似的想法，認為系統性評閱是針對特定主題、整合所有相關研究，並運用各項策略以減少研

究蒐集的偏誤，但後設分析並非必要條件(pp. 176-177)。

綜合上述，系統性評閱一詞，本身就包含質性與量化兩種，雖然其執行細節不盡相同，但核心概念都是強調系統性文獻的搜尋、選擇、評論與結果累積等，若研究者納入後設分析的步驟，才會是屬於量化系統性評閱，而由於本文架構與基礎，完全是在量化研究的脈絡下進行陳述，因此，筆者是將後設分析視為系統性評閱過程的一環。此外，對質性系統性評閱有興趣的讀者，可考慮參考 Dixon-Woods, Bonas, Booth, Jones, Miller, Sutton, Shaw, Smith, & Young(2006)的著作，該文對整個流程有著初步概述。

二、量化研究之描述性文獻回顧、系統性文獻評閱與深度批判文獻

傳統的量化研究中，在進行文獻評閱，以提出研究結論或政策方針時，大部分研究者在文獻探討時，會採用描述性文獻回顧，但此法太過依賴個人的主觀意見與選擇、缺乏外顯性(lack of explicitness)(Petrosino, Boruch, Soydan, Duggan, & Sanchez-Meca, 2001)，尤其當各研究間結果有衝突或分歧時，便可能因主觀取捨而造成偏誤。進一步來說，此情況通常會發生於某些研究者在進行文獻回顧時，對於證據或研究結果之選擇、比較時，通常會有意或無意的挑選適合本身預期之研究結果，作為回顧的焦點，而易忽略某些與預期結果不同調之文獻，例如：鍾鳳嬌、王國川、陳永朗(2006)在比較不同國籍與家庭背景之本、外籍配偶子女之語文、心智能力發展與學習行為時，文獻探討中曾納入陳湘淇(2004)、謝慶皇(2004)等相關研究，但卻未回顧林璣萍(2003)、吳宜鎔(2005)、柯淑慧(2005)等同樣曾針對外籍配偶子女語文或家庭背景(社經地位)進行分析之研究。而對於這種情形，Petrosino et al.是認為研究者在文獻探討時，對於佐證之文獻選擇，主要是由作者自行判斷要選擇哪篇研究、哪些文章等，缺乏外顯選擇過程，容易造成讀者閱讀時認知的偏誤，以為文獻回顧結果皆偏屬於正向或負向。

針對此策略所面臨的困境及批評，學者們是提出系統性評閱(systematic review)的概念，強調此策略是透過嚴謹的文獻搜集過程，能有效的整合訊息，並藉由後設分析(meta-analysis)技術，濃縮粹取出可靠的結論，以提供決策者合理決策或判斷的基礎與證據(Antman, Lau, Kupelnick, Mosteller, & Chalmers, 1992; Hoffert, 1997; Hunter & Schmidt, 1996; Mulrow, 1994)，而實際執行系統性評閱策略時，研究者是必須清楚表示研究選擇與定位的條件，如此，讀者即可明瞭在何種情況或條件下，才會形成目前的整合結果。

另一方面，文獻的評閱研究中，仍有像深度批判或研究議題式討論的文章(research issues: review article or critical review)，例如：黃毅志與陳怡靖(2005)、Lin(1999)，它們往往是從某個派典的重要理論開始評閱，接續，再檢證這些理論延伸的相關研究，並從中檢討可能產生的困境或缺失，進而提出自己的論點或對政策制定提出建議。這類深度批判文獻，並非屬於上述指稱之情況，這類文獻多是在其評閱脈絡下，形成焦點，再進行研究選取、評閱，可謂屬質性的系統性評閱研究，讀者同樣能明瞭其研究選取條件，同時，其關切的核心多強調的是新論點的合理性與突破性。

三、系統性評閱流程

近期發展中，後設分析多被視為隸屬於系統性評閱流程的一環，而本研究所指稱之

系統性評閱策略亦是如此；整個評閱流程與內涵係包含主題形成(problem formulation)、研究選擇與定位(selecting and locating studies)、研究品質的評估(quality assessment of studies)、資料收集與編碼(data collection and coding)、後設分析、結果呈現與詮釋(meta analysis, result presentation and interpretation)、改善與資料更新(improving and updating reviews)等步驟進行(Campbell Collaboration Web Site, 2008; Cooper & Hedge, 1994; Cuijpers, Straten, Warmerdam, & Anderson, 2008; Higgins & Greens, 2006)。茲依序簡述之：

(一)主題設定

主題設定並給與明確操作型定義，在系統性評閱流程中扮演的角色，可謂是火車頭，它影響著後續要如何搜集研究、如何分析等；若依效果量的組成，可分為三類，分別是差異(d)、相關(r)與勝算比(odds ratio)(Lipsey, & Wilson, 2001)，第一種多指稱該主題是以比較兩組間差異為主，以實驗為研究設計，例如合作學習對學習效果影響之後設分析研究(黃寶園、林世華，2002)；第二種則指稱該主題是以探討變項間關聯，多採問卷調查、相關分析為主，例如校長轉型、交易領導與教師工作滿意關係之後設分析(秦夢群、吳勁甫、濮世緯，2007)；最後一類，其主題選擇多是以比較經實驗或治療後，不同類別人數之成功/失敗比，而此主題多屬醫學領域，較少見諸於教育學門。

(二)研究選擇與定位

後設分析的結果是由搜尋所得的主要研究(primary studies)所組成，而透過廣泛的搜尋策略不僅是為確保詳盡分析所有相關研究，更為減低研究選擇偏誤(selection bias)影響；在進行文獻搜尋前，具體的界定納入與排除準則(inclusion and exclusion criteria)是為首要，以便能精確聚焦欲搜尋與推論的範圍。

1. 主要研究的納入與排除準則

系統性評閱流程中，對於納入與排除準則的界定，不僅是為促進文獻搜尋之效，更是決定整個研究外在效度(external validity)的可推論性(generalization)(如可推論至哪些受試者)、介入策略與結果變項的因果關係等。舉例來說，若某評閱主題是為探討概念構圖教學法對學習成就之影響，則研究者必須明確界定何謂概念構圖教學法與學習成就之操作型定義；此外，其分析之研究樣本範圍，是否包含國中小、亦或是大學生等；在學習成就之定義上，是否納入自然、閱讀等，亦或僅限定於某一學門的某一領域(如自然中的生物)；最後，研究設計(真實驗或準實驗設計)、語言(繁體中文、英文、其它語言)、文獻研究時限等項目的選擇，亦須清楚界定。

2. 主要研究的搜尋策略

在文獻搜尋時，強調的是充分搜尋策略(exhaustive search strategy)(Cooper & Hedge, 1994)，不論是出版或未出版文獻，皆為研究搜集範圍。實際操作上，首先，會根據主題條列出適合的中英文關鍵字，選擇與該主題相符的國內外教育文獻檢索系統，例如：中華民國期刊論文索引、全國博碩士學位論文資料庫、行政院國家科學委員會科學技術資料中心、ERIC、ProQuest Dissertations and Theses、OCLC 全球會議論文索引等進行搜尋，以期遍及期刊、博碩士論文、書目章節、會議論文與研究報告；其次，為避免遺漏

重要文獻，可搭配期刊手動搜尋(journal handsearching)策略，針對與主題密切的期刊、曾發表相關主題之後設分析研究參考文獻，進行逐篇檢視；最後，可考慮與熟悉該主題之關鍵學者進行聯繫，以期取得如灰色文獻或不曾發表之研究結果。

(三)主要研究的品質評估

在系統性評閱研究中，針對納入之主要研究品質進行評估，亦扮演著很重要角色，試想若納入之研究品質不佳，此舉將會嚴重影響整合後結果；但此議題一如 Moher, Jadad, Nichol, Penman, Tugwell 與 Walsh(1995)研究表示，他們至少已辨識出高達 25 種量表與 9 種檢核表，曾用來評估隨機控制實驗的效度與品質，而評量指標是從 3-57 題不等，面對如此紛雜的工具，是很難決定哪一評量工具才是最佳的；而此難題就如 Higgins 與 Greens(2006, p.83)認為，並沒有任何一項最佳標準，足以評斷某項試驗的真實效度，每項指標皆有其限制性，但至少我們能作的是依循測量的基本原則，發展研究品質的評估量表。

(四)資料收集與編碼

在資料收集與編碼方面，研究者通常會依據各自主題需求，制訂編碼簿(codebook)，收集的訊息多包含主要研究的背景特徵(如作者、出版年代、文獻屬性)、效果量計算訊息、研究品質評估資訊，此外，還會納入可能作為調節變項與評量工具品質評估指標(如信、效度)(Harwell & Maeda, 2008; Lipsey, 1994; Wortman, 1994)等。接續，實際執行時，為提昇資料收集信度，可經由交叉驗證的方式，重新檢驗與修訂，以期編碼之後設分析資料達百分之百的一致與準確性。

(五)後設分析、結果呈現與詮釋

後設分析是為系統性評閱中，有關統計技術部份，其論述的內容十分廣泛，亦是各流程中被廣為討論與熟知的，詳細可參考馬信行(2007a)、Hedge 與 Olkin(1985)、Lipsey 與 Wilson(2001)等學者著作。因此，以下僅簡要陳述。

後設分析流程，大致可匯整如下圖 1 所示，首先，研究者必須由上述收集所得主要研究，粹取出可相互比較之效果量(effect size)，但由於各研究所運用之統計分析技術可能不一，其也許是共變數分析、多層次分析(multilevel analysis)或多元迴歸分析等，因此，用以計算效果量公式、面臨的技術問題亦各有差異。但不論其公式為何，轉換後之效果量，其概念所代表意涵都是相同的，亦可透過特定機制互相轉換；接續，鑑於各主要研究所代表的重要性或母群推論性多不一致，因此，必須針對各效果量進行加權，而最早多是以總樣本數作為判斷的指標，但 Hedges 與 Olkin(1985, p.110)則認為最佳加權方式，是為效果量變異數的倒數(inverse variance)，他們認為當變異數愈小時，代表效果量的估計愈準確，相對應具備較大的加權，反之，其加權量則應較小。

在進行效果量(加權)的整合(pooling)前，如同 Hedges 與 Olkin(1985, p.122)認為必須先確定各研究間是否服從著一個共同的效果量(common effect size)，如此，才能以單一平均數來代表整體的表現狀況，而執行研究間效果量異質性的評估(between-study heterogeneity assessment)則為確認此目標是否達成的主要工作，通常研究者可採 Cochran (1954)的 Q 統計量或 Higgins 與 Thompson(2002)的 I^2 指標進行分析，據此結果，以確定各研究之效果量是否來自於同一母群、亦或是存在隨機效果；接續，根據不同的情況，

估計模式大致可分為固定效果(fixed effect)、隨機效果(random effect)、混合效果(mixed effect)模式與調節變項(moderator variable)之分析(Lipsey & Wilson, 2001)，如此，以進行總平均效果量的顯著考驗。最後，在結果呈現與詮釋上，除以傳統的總平均效果量進行陳述外，亦可考慮採用森林圖(forest plot)以描繪各研究效果量的分佈或運用累積後設分析(cumulative meta-analysis)能展現出隨著時間變化，研究效果量的改變情況(Antman et al., 1992)，以表達研究整合的成果。

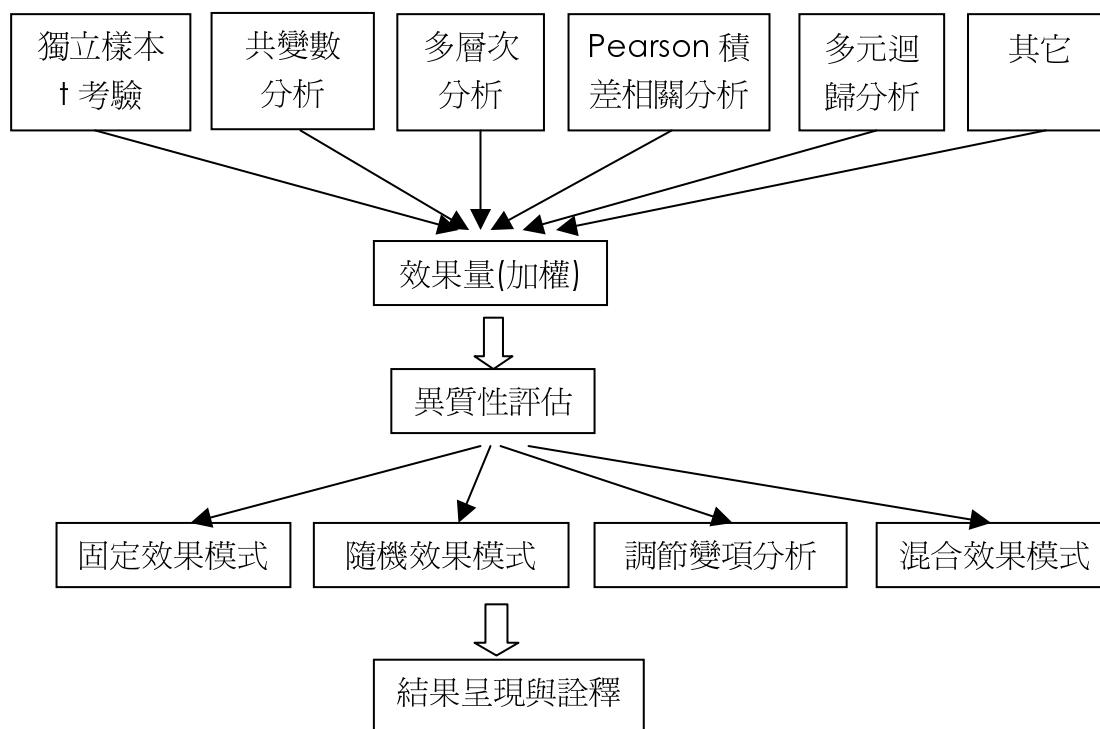


圖 1 後設分析、結果呈現與詮釋流程圖

(六)改善與資料更新

某教學策略、介入方案或變項間關聯可能會因外在環境的變化、時間的更迭而產生不同結果，甚至出現相反或有害的影響，例如：當整體社經環境隨時間變化而提昇時，是否會連帶影響社經地位與學習成就間關聯。諸如此類的議題，所強調的是資料更新，因此，不斷的更新、改善資料則成為研究中必需考量的課題之一，但太過頻繁的更新，有時又會過於浪費資源。傳統上，多是以固定時間為準則，例如 Cochrane 合作聯盟 (Cochrane Collaboration Web Site, 2008) 是以不超過 2 年為原則，進行評閱結果的更新。但 Soll(2008)認為此法是過於簡單化，而在參照 Moher, Tsertsvadze, Tricco, Eccles, Grimshaw, Sampson 與 Barrowman(2007)評閱過去的更新策略，認為在評估更新的需求時，可以考慮該主題的重要性、上次評閱的時間、近期新增研究的可取得性(availability)或透過評閱小組的共識及參考如累積後設分析結果的各種指標等，以決定更新的時程與必要性。

參、國內教育學門系統性評閱與後設分析策略

發展現況與困境

國內教育學門後設分析研究之發展，最早曾由林邦傑教授(1987)針對此方法作深入剖析，時至今日(2009年1月)，其應用層面已可見諸於各教育領域，包含教育行政(如秦夢群、吳勁甫、濮世緯，2007；葉連祺，1998)、諮商輔導(張高賓、戴嘉南，2005)、教學或學習方法(黃寶園、林世華，2002；廖遠光、黃淑敏，2003)、教育相關變項或議題(馬信行，2007a, 2007b)、與衛生教育(郭鐘隆、黃久美、蘇鳳足、邱靜如、劉貴雲、劉潔心、張婷婷，2003)、國外長期教育追蹤調查結果之綜合分析(馬信行，2008)與以統計議題為主軸執行的後設分析(章勝傑、林烘煜，2009)等等，應用範圍可謂十分廣泛，但就研究數量而言，相較於國內外其它統計技術或學門，仍是相對弱勢。例如研究者以後設分析、整合分析、meta analysis、文獻整合、research synthesis、systematic review 等關鍵詞，於中華民國期刊論文索引、全國博碩士論文資訊網進行搜尋，至多只能發現近百篇教育相關之期刊與博碩士論文，相較 Sutton 與 Higgins(2008)以 meta-analyses 為關鍵字在 PubMed 資料庫進行搜尋，發現自 1990 年至 2006 年，皆是以數百篇的速度，逐年快速增長，其中，在 2005、2006 年，更是各自有超過 2000 多篇相關研究、文獻，兩者相差可謂十分懸殊。此外，若單以 systematic review 為關鍵字，針對上述國內資料檢索系統進行搜索時，研究者除確認數十篇醫療相關文獻外，未能發現任何一篇教育相關之文獻發表。綜合上述，隱約顯示出系統性評閱與後設分析策略於國內教育學門發展與應用之弱勢現況。

在援以系統性評閱與後設分析策略，建立以實證為本教育理念上，是有其重要性存在，但若深入探討時，可發現國內外教育學門中對於此策略之發展，仍存在些許困難和限制，當然，除了國內研究社群有限，造成影響力無法充分發揮外，而以下條列的可能面臨困境，不僅造成研究發展上的相對緩慢，更阻礙建立以實證為本教育理念的進步，陳述如下：

一、資料搜集費時、費力，無法提供足夠的證據

在進行系統性評閱與後設分析研究中，其主要步驟之一，是為廣泛的搜集與該主題相關之研究、文獻，同時，在考量避免可能的出版偏誤影響研究整合的品質下，研究者必須詳細羅列有關的關鍵詞，並訂立主要研究的納入/排除標準，再透過教育文獻檢索系統或手動搜尋期刊等方式，詳細搜尋、取得各文獻報告；接續，再由一位或兩位以上研究者逐一檢視文獻、進行資料編碼等(Cooper & Hedges, 1994； Higgins & Greens, 2006)。由此可見，完成一份嚴謹的系統性評閱研究，不僅是時間成本的支出，其背後更是巨大人力、心力的付出；據此，亦不難想像為何國內有關後設分析之研究總數，只有近百篇的文獻。

單就目前國內系統性評閱與後設分析的研究成果而言，若欲以此建立國內以證據為本的教育理念，就如同前述 Hargreaves(1996)、Gough(2004)等人對教育研究的批評，所

引起學者們的反思，認為我們實在是缺乏適當、足夠的證據，以回答各項研究或實務工作中，可能面臨的挑戰。

二、後設分析研究品質不一，降低證據之可信度

系統性評閱策略所強調的是外顯、可重覆驗證的特色，因此，其執行過程都必須詳實逐一執行、記錄；反觀國內研究，多重視後設分析統計技術之運用(如孫旻儀、蔡明學(2007)、秦夢群、吳勁甫、濮世緯(2007)、黃寶園、林世華(2002, 2007))，而易忽略系統性評閱中的其它步驟(註：本文強調重點是國內研究易忽視系統性評閱的其它流程，並非認為後設分析技術不重要)，此點就如同 Harwell 與 Maeda(2008)的文獻所分析，執行後設分析的研究者所呈現的資訊，往往並不足夠，例如：並未說明該整合研究之母群的推論為何？、未清楚定義研究納入或排除的準則或未說明有多少比率之研究，因無法計算效果量而被排除、編碼變項的合理性等等，如此都易影響到後續研究者或實務工作者判斷後設分析結果之可信度(credibility)。

除評閱流程訊息提供的不足外，承續前項有關資料的搜集困難議題，研究者最常見的變通方式，則是縮小研究的推論範圍，如僅針對臺灣地區(或某特定區域)的文獻、研究進行搜尋，但延伸的問題之一，在於會影響文獻的搜集數量，例如黃瑛琪、戴嘉南、張高賓、連廷嘉(2005)以 14 篇文獻進行研究整合，而彭瓊慧(2002)、鄒浮安(1994)等人，進行的後設分析研究中，最後各僅以 15 篇主要研究進行分析。研究數量過少的隱憂，在於進行總平均效果參數之統計顯著檢定時，容易對統計檢定力(power)產生質疑，致使整體分析結果產生偏誤。

三、連續性研究資料累積的困難，無法提供即時更新的證據

系統性評閱與後設分析的目的或出發點，在於期望能有效整合過去對於單一議題之研究成果，經篩選、分析後，提出最後的一般性結論(Glass, 1982)，進而，提供實務工作者或主政者做為決策的證據。就實務面而言，當然希望此整合結果的時效性，是愈新愈好，最好此結果能依主題特性，定期的改善與更新；而在觀察目前教育相關研究，亦可發現有類似的趨向表現出來。單一主題經後設分析整合後，仍會有類似的研究於後期產生，例如 White(1982)早期曾針對社經地位與學業成就相關之研究，進行後設分析，接續，Sirin(2005)承續 White 的研究，針對 1990-2000 年文獻進行搜尋與整合，進行兩變項關聯之後設分析探討；Prout 與 DeMartino(1986)引用 33 個發表於 1984 年前的研究，探討以學校為範疇的諮商與心理介入治療效果，而 Prout 與 Prout(1998)則是再次更新 Prout 與 DeMartino 研究，評閱了 1985 至 1994 年相關的心理諮商治療介入效果；在國內，鄒浮安(1994)曾針對家庭社經地位與學業成就關係，進行後設分析，接續，許崇憲(2002)也有類似的研究發表，以至於孫旻儀、蔡明學(2007)時，亦曾針對 1994 年後新增之研究文獻，進行整合。

諸如此類的研究，顯示出不僅在實務上有更新證據的需求，研究上亦有類似的回應，但實際上，真正屬於連續性、有效更新的研究多屬難得，例如上述研究嚴格說來，由於橫跨時間過長(如 White(1982)-Sirin(2005))，仍是屬於在間斷的時間點，進行後設分

析所下之結論，並未能提供即時證據；此外，承續的研究有時並不是同一作者，能完整、準確依循前面研究之準則，進行後設分析的可能性則相對變低，更甚者如上述 Harwell 與 Maeda(2008)研究所發現，後設分析研究者所呈現的資訊往往不足，易導致先後兩研究間比較的結果，存在著疑問。

四、後設分析研究中的出版偏誤與語言偏誤

影響系統性評閱與後設分析結果的因素中，是以出版偏誤(publication bias)與語言偏誤(language bias)，較被學者廣為討論，前者意指著後設分析研究中所引用之文獻，是系統性的未能代表整個研究母群狀況(systematically unrepresentative of the population of completed studies)(Rothstein, Sutton, & Borenstein, 2005, p.1)，換言之，即是研究中一致性的缺漏了某些重要或顯著的文獻，進而，所得之整合結果，其效度是會受到質疑，而消極的作法通常是會在研究中提供檢定的證據，例如提供 Rosenthal(1979)的 Fail-safe N；但積極的作法則是建立試驗註冊(trials register)資料庫，其概念是預先收集與建立各主題相關之實驗試驗檔案(Berlin, & Ghersi, 2005)，以供研究者在進行文獻搜尋時參考。反觀國內，對於出版偏誤議題的重視仍未普及，許多研究(例如孫旻儀、蔡明學(2007)、張高賓、戴嘉南(2005)、黃淑滿、周麗端、葉明芬(2008)、黃瑛琪、戴嘉南、張高賓、連廷嘉(2005)等)，甚至連消極的檢定作法都未考量在其研究中。

除出版偏誤外，另一項與此類似且受關注的議題即是語言偏誤。由於不同國家間存在的語言隔閡，致使多數後設分析研究會限制以英語為出版之文獻為研究搜尋標的，如此，會造成 Grégoire, Derderian 與 LeLorier(1995)所認為的，一個非英語為母語的國家若出版了一篇負向結果的臨床試驗，多會傾向於將研究結果發表在當地國家的期刊，而不是發表在國際中能快速散播、以英語為主的期刊中，因此，會形成正向結果多發表在英語期刊中，而負向結果多發表在研究者當地期刊中。就國內的研究現況而言，可以同時納入中、英文研究於後設分析中，仍屬少數，但即使如此，語言偏誤可能的影響仍然是存在(如未納入相關主題之日文、德文...文獻等)。對此議題，目前仍無有效、具體的因應之道。最常見的作法，乃是廣泛羅列與研究主題相關之各語言文獻，而對於非母語之研究，則是聘請專業翻譯人員進行譯釋，但此舉的成本所需不斐；另一項逐漸被重視的因應之道，則是成立系統性評閱的合作聯盟，在醫學領域中早已是發展完備，但在國內教育領域則仍是屬於待積極重視、發展的一環。

肆、國際發展與醫學領域之借鏡

就目前國內教育學門在系統性評閱與後設分析發展之現況，可發現我們缺乏的是專責的文獻評閱機構，由此單位制定可重覆驗證的系統性評閱流程，以提供可靠且時刻更新的證據；更可以此單位為基礎，廣泛搜集、評閱與分享各項主題之研究結果，建立系統性評閱資料庫；最後，進一步建立跨國的合作聯盟關係。此舉不僅能減輕語言偏誤干擾，更能提供後設分析學者許多即時、可整合的資料。而此構想，就如同 Davies(1999)、Petrosino 等人(2001)認為，面對不斷累積、增加的資訊，要如何能有效、快速更新，並讓決策者輕易獲得，是可依靠建立系統性評閱資料庫來因應，而此點亦早已被國外許多

研究機構所實踐。其中，尤以醫學領域的發展最為完善。以下茲簡述國外醫學、教育領域中，有系統性評閱組織的發展與建置之系統性評閱資料庫，以供國內借鏡。

一、醫學領域系統性評閱組織之發展

國際間系統性評閱資料庫的建置，最成功者當以醫學領域的 Cochrane 合作聯盟為首。該聯盟的前身最早是為 1992 年底在英國成立之 Cochrane 中心，自此該中心即不斷擴充其研究主題的範圍並尋求跨國合作，逐漸建立起橫跨健康、醫療照護(health care)介入成效之系統性評閱的國際合作組織，而於 1993 年底，國際 Cochrane 合作聯盟(Cochrane collaboration)首度正式成立(Bero & Rennie, 1995)。目前 Cochrane 合作聯盟已遍及全球各地，包含澳洲、巴西、加拿大、中國、丹麥、德國、義大利、荷蘭、南非、西班牙、英國與美國等；長久以來，該聯盟不斷為建立證據為本的醫學與健康照護(evidence-based medicine and health care)而努力，而 Cochrane 資料庫(Cochrane library)亦提供醫學領域許多高品質、最新的醫療介入相關研究報告。

Cochrane 合作聯盟的主要任務是由超過 50 個國際合作的文獻評閱小組(collaborative review groups)進行肺癌、口腔健康、耳鼻喉失調等多項與健康照護主題相關之醫療介入成效的系統性評閱研究，評閱報告與結果會收錄於 Cochrane 系統性評閱資料庫(Cochrane Database of Systematic Reviews)中。而在根據 Jadad, Cook, Jones, Klassen, Tugwell, Moher 與 Moher(1998)對該資料庫品質的分析，發現相較於其它收錄在 MEDLINE 資料庫之期刊所發表的評閱研究，Cochrane 的系統性評閱研究不僅方法較為嚴謹性，資料更新速度更為即時。

Cochrane 合作聯盟更規畫有專門團隊廣泛、系統性收錄有關醫學領域中各項介入方案的隨機實驗資料，納入 Cochrane 臨床試驗資料庫(Cochrane Central Register of Controlled Trials, CENTRAL)中，而此舉亦成為後設分析研究者搜尋相關醫學文獻，最佳的檢索系統。截至 2008 年 9 月為止，估計已納入超過五十萬筆隨機實驗報告索引，同時期，亦有許多國家諸如澳洲、中國、巴西等紛紛宣佈投入系統性文獻搜集工作，為 CENTRAL 共盡心力(Cochrane Collaboration Web Site, 2008; Dickersin, Manheimer, Wieland, Robinson, Lefebvre, & McDonald, 2002)。

二、國際教育領域系統性評閱組織之發展

教育學門中有系統性評閱組織的發展，雖不似 Cochrane 合作聯盟發展的如此緊密與完備，但目前歐美等國各自成立之教育領域系統性評閱組織亦行之有年，皆具備基礎的架構。

美國教育部曾於 2002 年成立 What Works Clearinghouse(WWC)，透過發展與運用標準程序於評閱和整合教育相關研究，以建立教育中主要且可靠的科學證據來源，其範圍包括初始閱讀(beginning reading)、中小學數學、幼兒教育、英語學習方案、輟學預防、品格教育等主題(What Works Clearinghouse, 2008)。

另一方面，同樣是自美國發起的國際非營利組織—Campbell 合作組織(Campbell Collaboration, C2)—成立於西元 2000 年，其發展主要參考 Cochrane 模式，集合一群對

證據為本教育理念與系統性評閱有興趣的學者、政策制訂者、實務工作者與學生，針對教育、犯罪與正義、社會福利等研究領域進行系統性評閱，期望提供該領域相關決策之實證基礎；而評閱的結果也能在 Campbell 系統性評閱資料庫(Campbell Library of Systematic Reviews)中取得(Boruch, Petrosino, & Chalmers, 1999; Campbell Collaboration Web Site, 2008)。此外，Campbell 與 Cochrane 合作聯盟共同成立了社會、心理、教育與犯罪的試驗資料庫(Social, Psychological, Educational and Criminal Trials Register, SPECTR)，系統性地針對教育、心理與犯罪等相關介入方案進行文獻的搜集，以期提供系統性評閱工作者更佳的研究資源平台(Petrosino, Boruch, Rounding, McDonald, & Chalmers, 2000)。

在美洲大陸的彼岸，由英國倫敦大學所主導的 Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)成立於 1993 年，針對社會科學(含教育)與公共政策等領域執行研究整合(research synthesis)與發展評閱技術，目的在貢獻一己之力，提供可靠的研究發現給政策制訂者或實務工作者運用。EPPI-Centre 網站的證據資料庫(evidence library)亦能取得教育、健康促進(health promotion)、公共衛生(public health)等主題的全文電子檔；而 EPPI-Centre 實證資料庫中各篇研究所引用之參考文獻或書目索引，則能以關鍵詞在 EPPI-Centre 教育研究資料庫(EPPI-Centre database of education research)中搜尋取得。據統計，目前(2008 年 9 月) EPPI-Centre 已有超過 70 篇系統性評閱研究，而所引用之文獻累積將近三千筆研究資料(EPPI-Centre Web Site, 2008)。

系統性評閱組織的發展早先是自醫學領域發起，而隨後逐漸被心理、教育等社會科學領域的學者所仿效，並逐步建立起類似的組織，以從事相似的工作，但畢竟完全的經驗轉移，是有其困難及侷限性，例如：醫學學門的實驗研究，強調的是藥物、治療的有效性，而教育學門強調的不是僅教學的結果，更著重教學的過程，致使教育學門在短短數年間發展相關組織過程時，面臨許多待克服的難題與因應策略(例如：鍵檔的資訊要包含哪些內容)，而在碰觸這些實務或細節面向的議題前，研究者更希望能預先就某些上層的觀念與疑慮進行說明，以便後續國內研究者在架構相關組織時，是有其核心思想所導引，此點是於後續進階議題中進行討論。

伍、進階議題的討論與分析

有關進階議題的討論方面，本文不在於探討系統性評閱或後設分析策略本身的技術性議題，例如出版偏誤、語言偏誤、研究品質要如何評估等。研究者主要考量這方面的論述，目前已有許多國外學者撰書發表(如 Cooper, & Hedge, 1994; Rothstein, Sutton, & Borenstein, 2005)，有興趣的讀者，可逕自參考之；反而，本文是希望從本質性的議題著手，討論在推動系統性評閱與後設分析策略於建立以證據為本教育理念時，可能面臨的一些爭議(debate)與批評，某些是屬於研究本質的議題、某些則是屬於醫學與教育領域的差異，其中，某些議題可能至今仍處於辯論狀態，未有特定的答案或回應。而提出這些議題，關切的是一種自我的省思，期望在鼓吹藉由系統性評閱與後設分析策略，以推動國內建立以證據為本教育理念時，研究者能瞭解可能存在的疑慮，甚至在未來推動時，能化解或避免這些質疑可能產生的影響。

以證據為本位的醫學理念在英、美兩國的發展，確實促成醫療品質顯著的提昇，亦帶給了教育界許多的啟示，學者們思索著如何向醫學借鏡，以解決當前面臨的教育問題及癥結；Davies(1999)、Evans 與 Benefield(2001)就曾認為運用系統性評閱策略在醫學領域的成功，加以適當的引導至教育情境中，是可將評閱研究視為一種能呈現有關教育政策與實務累積效果的方法。而此舉所考量的，可能是系統性評閱與後設分析策略，確能呼應某些的需求，例如：其評閱流程強調的是外顯、可重覆驗證，若以此來獲得或證明有效的介入方案或教學策略，能清楚讓國會議員、民眾知道他們花的錢得到了哪些成效？但是否能完全轉移醫學的理念至教育研究上，其間不無疑義，而這些疑慮則多是醫學與教育領域、研究間的轉化、影響與隔閡，茲說明如下：

一、重視結果導向

醫學領域對於判斷某項治療的效果，也許有所謂的黃金標準(gold standard)，可能以治療樣本存活比率來代表某治療方案的成效，但在教育領域中，真的適合探討什麼才是有效(what works)的問題嗎？數學、科學分數的提昇，能當作是某教學策略有效的代名詞嗎？Hargreaves 教授在 1996 年的演講中認為，教育研究的主要任務在於改善教師的教學結果，但令人擔憂的是，我們會不會從強調以證據為本的教育理念轉化為強調以結果為本的教育(outcome-based education)作為(Elliot, 2001)。

Elliot(2001)認為將教學變得有效的想法，並沒有必要去承擔以結果為本的教育理念所產生的特定意識型態包袱，兩者不一定是在同一線上；Elliot 並不反對教學必須是有效的，但不贊成教師的心中都有一把相同的尺，認為每位學生都要達到相同的結果；教育並不是一種突破(breakthrough)，沒有單一的準則或標準，它是種一步一步改善(step by step improvement)的過程(Slavin, 2002)。

二、強調實證主義

Hammersley(1997, 2001)認為系統性評閱策略所強調的一種過程客觀性(procedural objectivity)，不外乎是想藉由統計上的推論，以達到客觀的結論，其本質上是屬於實證主義研究模式(positivist model of research)，而這項批評除了是對此策略在社會科學領域中，忽略主觀判斷的價值提出疑義外，更是針對 Hargreaves(1999)過度的強調在教學環境中，必須將隨機控制實驗結果視為研究證據所展現的成效，卻忽略了 Morrison(2001)認為過度依賴隨機實驗時，可能面臨的幾項問題，例如：對於證據的觀點過度限制在因果與可預測性的範圍、低估其它資訊來源的價值、將複雜的現實整體，過度不真實的簡化與模糊化、低估了以多面向觀點檢視什麼才是有效的重要性、無法掌握非線性現象的動態發展、忽視環境脈絡展現出的意涵等等。同時，Sanderson(2003)也有類似的想法，認為過度強調隨機控制實驗，會造成工具性主義(instrumental rationality)興起，如此去脈絡化(de-contextualize)的觀點，是限制了研究者觀察可能潛在的影響效果，以致無法詳細瞭解某教學策略是「如何」產生這樣的效果；Sanderson 認為應該將觀點做個轉換，將什麼才是有效的觀念，轉變為什麼才是適當的(what is appropriate)，以此來因應複雜與矛盾的各項問題。

前述對於實證主義之評述，早先的哲學家，就已提出類似的觀點，Husserl 認為實證主義是企圖將意識以及所有觀念與規範自然化，他認為每個意識狀態都有其指向性(intentionality)，即意識之流指向世界的知覺、想像、分析、回憶等主體性，強調必須將現象置於所處的脈絡背景上，才能使觀察者客觀的瞭解生活世界的圖像，所謂存而不論(phenomenological reduction)的過程，更著重回歸事物本質，主體自我反省、觀照事物本質的過程(引自德穆·莫倫，2000)；此外，Schutz 則是反對單純思考研究對於觀察者的意義，卻忽略對於行動者的意義，因此，他強調「理解」的方法，尤其是對於「被觀察的行動者有何意義？」(舒茲，1967)。而在教育實證研究過程中，研究者對於發現教育現象背後的因果關係，是以科學語句如以描述、驗證後的科學法則，這種方式往往是將原因當作手段，而將結果當成想達到的目的，是明顯忽略教育的社會過程，以 Gadamar 來說，就是忽略了社會過程的情境考量(楊深坑，2002)。

對於上述的批評，Gough(2004)則反駁認為，將系統性評閱策略單純認為是實證主義，這種假設多半是將實證主義視為具有負向的特性，而且窄化了系統性評閱的定義；系統性評閱策略是可以納入質性研究，雖然在評閱流程中，對於質性準則的使用與判斷偏誤的觀點，兩者可能互有不同，但很多主要原則卻是相通的；此外，Davies(1999)早就主張，證據為本的教育理念是整合了個人教學與學習的專業、經驗，並搭配來自系統性評閱中所得之最佳、可行之外部證據，所調合、擴增的結果，因此，融入教師、學生自身主觀的判斷原本就是其中重要元素。

三、證據的來源與研究本質

若由過程客觀性觀點來檢視系統性評閱策略時，Hammersley(2001)認為會產生幾項問題：首先，為什麼研究者在執行研究整合時，會偏好某些與主題充分相關的證據，但卻忽略或排除了某些具有價值的證據，這意謂著證據被特權化嗎(evidence being privileged)?我們需要多少證據、哪類證據，是依據知識的本質需求(nature of knowledge claim made)而定，這些來源可能是建立在各式研究(描述性、解釋性或理論性)上，並搭配證據可信、有效的程度進行主觀判斷的；其次，在使用固定、標準的準則來評估研究設計時，此舉已經是忽略了變動的來源，例如，若某個評閱過程是無法具體過程化時，若強迫其如此，只會加深其研究結果的偏差；最後，不同研究間是可加性(additive)的嗎?Hammersley 認為研究間其實是互補的(complementary)，一連串的研究是透過不間斷的修訂，以增加其解釋力，不同的研究發現是無法單純加在一起，每一個研究都是對該理論做出了真實的貢獻。

Gough(2004)認為上述對於證據引用的批評，是過度推論系統性評閱扮演的角色，我們強調的是在決策上，研究證據所扮演的理性角色，是需要被提昇，而不斷投入研究的努力，將會使得研究者變得更瞭解其它研究的發展，因而，能更有效的使用這些合作努力的結果，而系統性評閱策略是其中的方法之一，並未主張應該利用此策略去支配所有的決策；此外，系統性評閱策略對於研究的搜尋、組織與分析過程，是強調必須在主題的理論架構下進行，即使是編碼的變項，都必須陳述其合理性(rationale)(Harwell & Maeda, 2008)，因而，並不是所有研究都可以納入研究整合中。

四、醫學與教育領域之異質性

Hargreaves(1996)在對教師訓練中心進行年度演講時，認為教育研究應該對於教師的專業實務發展，有更多的相關與影響力，但是，Hammersley(1997)質疑這項主張，認為醫學與教育領域兩者間是存在許多的異質性，例如教學比較不像醫療，其本質是比較接近實務(需主觀判斷)，而非技術面(單純依循準則(rule-following))、醫學研究的本質重視實證主義，是更勝於對脈絡的感知，但教育學門則非如此、即使是醫學中所獲得的研究知識，其與實務之間，亦具備相當的隔閡，更何況是教學領域呢？；但 Hargreaves(1997)則是從醫師與教師扮演的角色、醫學與教育專業研究的本質、醫學與教育領域中研究與實務的關係等三個面向，逐一進行反駁，認為不論是教師或醫生在決策時，都必然牽涉到複雜的主觀判斷、其實務性都很強烈；此外，教師在某層面上亦是技術導向，會依循特定準則執行例行性的教室決策；而對於實證主義的宣言，Hargreaves 認為是過度簡化醫學實務的知識基礎，他認為知識的本質至少可區分為三種層次，即使是醫學領域中，亦有研究是針對治療病人的臨床決策效果進行評估的，如此，則需相當依賴醫生對於病人的脈絡感知，最後，他承認相對於醫師，教師確實擁有較少的科學知識，來協助他們進行實務決策，當然，現行存在的有效證據確是非常的稀少，造成教師無法將這些證據視為是一種資源，以引領教學實務的推動。

五、研究、實務與政策

在以系統性評閱與後設分析策略，推動證據為本理念時，可能引發的疑慮之外，更重要的議題在於研究、實務與政策間，要如何有效的整合呢？這是個難解、複雜的問題，就如同 Davies, Nutley 與 Smith(1999)認為我們現在面臨了一個難以解決的困境，多生產證據並不保證我們能把這些我們認為有效的東西，轉化為實務，其背後交織影響的因素，可能是複雜的政治背景、行動者本身專業智慧與訊息、媒體的傳播等(De Vibe, Hovland, & Young, 2002)，尤其當彼此的觀點，相互衝突時，我們要如何因應？這個問題，研究者只能留待未來思考，但要重申的是 Davies(1999)、Gough(20004)、Hargreaves(1996)等人都曾強調過的想法，系統性評閱策略與其推動的證據為本理念，並不是解決當前教育問題的特效藥，但它可以是作為教育人員在面對各式政策或實務問題時，作出專業判斷的原則與方針。

陸、國內系統性評閱與後設分析研究之建議

資訊的發達、績效責任意識的提昇，促成國際間對於證據本位的訴求，其執行核心的系統性評閱與後設分析策略，亦隨之受到重視，引起許多學者熱烈的回響與推廣(馬信行，2002；Davies, 1999；Hargreaves, 1996, 1997)；反觀國內教育學門，對其發展仍處於相對弱勢，當然有其存在的限制與面臨困境，但從另一個角度視之，這不外乎是一種轉機，借鏡國際組織的發展經驗，可做為國內推展的架構，反思可能存在的問題，不斷的精進並往前展望。以下茲提供幾項建議，供國內實務與研究工作者之參考。

一、系統性評閱與後設分析策略在國內未來的發展

後設分析能運用的範疇十分廣泛，諸如醫學、經濟、教育等，都可見其足跡，但若觀察國內教育學門的發展，其應用多單純集中在幾個領域(如人文、醫學教育)，反觀其它如科學教育，每年雖然有為數不少的量化研究或文獻被發表，但後設分析研究的產出仍非常的稀少。在運用的層面外，對於技術性的議題，國內亦多集中在後設分析方法的介紹與進階的統計技術議題討論(如黃寶園、林世華(2007)對於後設分析與結構方程式模式的結合；多層次的後設分析(洪來發、王文中，1999)、馬信行教授發展對單一受試者研究之整合方法(馬信行，2007a))，對於系統性評閱延伸的議題，則很少觸及，例如語言偏誤、出版偏誤等。舉例來說，對於比較相同主題但不同語言之研究整合結果，在國際間早已行之有年(Grégoire, Derderian, & LeLorier, 1995)，但對於具有中文閱讀優勢的台灣研究者而言，卻是目前仍未開發的方向；此外，國內研究對於出版偏誤的觀點，仍多集中引用 Rosenthal(1979)fail-safe N ，但卻忽略許多已發展或待開發的領域(如隨機試驗的註冊、其它更精確的統計檢定方法)(Rothstein, Sutton, & Borenstein, 2005)。如此等等，都是國內對於系統性評閱與後設分析策略的發展，未來可思索的方向。

二、對國內系統性評閱組織的催生

早自以證據為本位的理念開始推動時，Coalition for Evidence-Based Policy(2002)、Davies(1999)、Petrosino 等人(2001)就不斷的鼓吹建立證據資料庫的重要性，而各國對此亦有所回應，美國教育部於2002年成立了 What Works Clearinghouse(<http://ies.ed.gov/ncee/wwc>)、國際非營利機構移植了 Cochrane 模式，成立 Campbell 合作組織 (<http://www.campbellcollaboration.org/>)，而在英國則成立 EPPI-Centre(<http://eppi.ioe.ac.uk/cms/>)等。專責機構的成立，相繼建立起系統性評閱資料庫與試驗資料庫，為期望提供決策者即時、更新與嚴謹的證據而努力；反觀國內，就研究者觀察教育領域，目前尚未有任何正式專職系統性評閱組織成立，在借鏡國外組織的發展下，建議政府能考慮指定相關機構成立專責單位，以建立起國家級的證據資料庫。

在國家級的系統性評閱組織外，國際學者們為建立證據為本的知識努力而成立的小型合作聯盟，亦相繼興起，彼此透過網站平台分享資料、相互交流，例如來自瑞典、荷蘭與美國的學者—Cuijpers, Straten, Warmerdam 與 Anderson(2008)—針對憂鬱症相關的心理治療介入方案建立隨機實驗資料庫，並執行一系列的系統性評閱研究，其成果可透過網站：Psychotherapy: Randomized controlled and comparative trials (<http://www.psychotherapyrccts.org/index.php?id=25>)自由下載；流行病學領域系統性評閱與後設分析研究的 Berkeley Systematic Reviews Group (<http://www.medepi.net/meta/index.html>)、搜集社會方案(social program)研究的 Social Programs That Works (<http://www.evidencebasedprograms.org/>)等都屬之。在有限的資源下，研究者透過小型系統性評閱組織的成立，為建立證據本位的理念而努力，亦是國內教育學者們可參考的方向。

柒、結語

以證據為本的教育理念，是逐漸受到國內外許多學者的推廣，而建立此理念的方法之一，就是援用系統性文獻評閱策略，其產生的證據來源，是具有整合性、嚴謹性，同時，若能輔以成立專責的系統性評閱組織以推動相關事務，不僅能為國內教育提供長期性的證據，更能有效因應現行後設分析所面臨的資料搜集費時、費力、連續性研究資料累積的困難與出版偏誤、語言偏誤等難題，這概念就如同現行國內外大型教育評比資料庫的建置一般，它們都是由專責機構針對某個主題，負責執行長期性的調查、並提供可靠的資料為研究者所使用，不僅節省人力的重複浪費、更能提高研究成果的可信度，目的都在於深入瞭解某特定教育主題的概況或現象，但其中的差別在於系統性評閱策略所面對的是過去的各式研究結果與未來可能產生的文獻資料。

研究者在推行這項策略時，不可避免的會面臨許多難題與批評，諸如：以結果為導向、過於實證主義、如何面對醫學與教育領域的異質性等議題，這些都是系統性評閱策略者所需積極克服與面對的，雖然，某些議題仍處於辯論狀態或未有特定回應，但我們必須強調這個策略只是所有可行方法的其中之一，在現行國內普遍缺乏證據資源下，是有其倡導的價值性，但未主張應該利用此策略去支配所有的決策，它只是所有證據來源的可行選項之一，並非證據的全部，更重要的是決策者如何能在研究、實務與政策間，有效的整合以推動，這才是證據能有效發揮的最大智慧。

參考文獻

- 吳宜諄(2005)。外籍配偶子女之家庭學習環境對其國語文能力發展的影響。國立台北教育大學特殊教育學系碩士論文，未出版，臺北市。
- 林邦傑(1987)。整合分析的理論及其在國內的應用。**教育與心理研究**，**10**，1-38。
- 林璣萍(2003)。台灣新興的弱勢學生：外籍新娘子女學校適應現況之研究。國立台東大學教育研究所碩士論文，未出版，台東縣。
- 柯淑慧(2005)。外籍母親與本籍母親之子女學業成就之比較研究？以基隆市國小一年級學生為例。國立台北師範學院幼兒教育學系碩士班碩士論文，未出版，臺北市。
- 洪來發、王文中(1999)。效應量集中與分散程度之多層次整合性分析。**測驗學刊**，**46**(2)，61-72。
- 孫旻儀、蔡明學(2007)。社經地位和學生學業成就關係之後設分析。**國教學報**，**19**，199-222。
- 秦夢群、吳勁甫、濮世緯(2007)。校長轉型、交易領導與教師工作滿意關係之後設分析。**教育與心理研究**，**30**(4)，27-55。
- 馬信行(2002)。教育科學導論。臺北：元照。
- 馬信行(2007a)。後設分析之方法論問題之探討。**量化研究學刊**，**1**(1)，175-189。
- 馬信行(2007b)。導致學生中輟的因素：一個後設分析。**教育與心理研究**，**30**(1)，123-140。
- 馬信行(2008)。影響學生學業成就之因素：對美國國家長期教育追蹤調查研究結果之綜合分析。**教育政策論壇**，**11**(4)，1-31。
- 張高賓、戴嘉南(2005)。國內兒童團體諮商與心理治療整合分析研究。**中華輔導學報**，**18**，77-119。
- 章勝傑、林烘煜(2009)。二分依變項相關性研究的後設分析：以留級與超齡對學童中輟的影響為例。**台東大學教育學報**，**20**(1)，199-230。
- 許崇憲(2002)。家庭背景因素與子女學業成就之關係：臺灣樣本的後設分析。**中正教育研究**，**1**(2)，25-62。
- 陳湘淇(2004)。國小一年級外籍配偶子女在智力、語文能力及學業成就表現之研究。台南師範學院教師在職進修幼教碩士學位班碩士論文，未出版，臺南市。
- 郭鐘隆、黃久美、蘇鳳足、邱靜如、劉貴雲、劉潔心、張婷婷(2003)。衛生教育與行為科學介入研究成效之統合分析。**衛生教育學報**，**20**，71-88。
- 彭瓊慧(2002)。我國資優教育研究之回顧與後設分析研究。國立台灣師範大學特殊教育研究所碩士論文，未出版，臺北市。
- 舒茲（1991）。社會世界的現象學。（盧嵐蘭譯）。臺北：久大。（原著出版年：1967年）
- 黃淑滿、周麗端、葉明芬(2008)。依附與其相關因素之後設分析：台灣近二十年文獻的研究。**教育心理學報**，**40**(1)，39-62。
- 黃瑛琪、戴嘉南、張高賓、連廷嘉(2005)。台灣青少年生涯團體諮商效果之整合分析研究。**高師諮商輔導學報**，**12**，70-99。
- 黃毅志、陳怡靖(2005)。台灣的升學問題：教育社會學理論與研究之檢討。**臺灣教育社**

- 會學研究，5(1)，77-118。
- 黃寶園、林世華(2002)。合作學習對學習效果影響之研究：統合分析。教育心理學報，34(1)，21-41。
- 黃寶園、林世華(2007)。人格特質與社會支持對壓力反應歷程影響之研究：結合統合分析與結構方程模式二計量方法。教育心理學報，39(2)，263-294。
- 楊深坑(2002)。科學理論與教育學發展。台北：心理。
- 葉連祺(1998)。國內有關學校效能研究之後設分析。國民教育研究學報，4，265-296。
- 葉連祺、張盈霏(2001)。證據本位管理之論析。學校行政雙月刊，12，59-69。
- 鄒浮安(1994)。家庭社經地位與學業成就之關係：後設分析。教育研究資訊，2(3)，38-47。
- 廖遠光、黃淑敏(2003)。電腦網路學習與傳統教學對學生學習成效影響之後設分析。管理與教育研究學報，1，161-182。
- 德穆·莫倫(2000)。現象學導論。蔡錚雲(譯)。臺北：桂冠。
- 賴志峰(2004)。證據本位教育政策之理念與啓示。初等教育學刊，17，83-104。
- 謝慶皇(2004)。外籍配偶子女學業成就及其相關因素探討。台南師範學院教師在職進修特殊碩士學位班碩士論文，未出版，臺南市。
- 鍾鳳嬌、王國川、陳永朗(2006)。屏東地區外籍與本國籍配偶子女在語文、心智能力發展與學習行為之比較研究：探析家庭背景的影響。教育心理學報，37(4)，411-429。
- Antman, E. M., Lau, J., Kupelnick, B., Mosteller, F., & Chalmers, T. C. (1992). A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: Treatments for myocardial infarction. *The Journal of American Medical Association*, 268(2), 240-248.
- Berlin, J. A., & Ghersi, D. (2005). Preventing publication bias: Registries and prospective meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 35-48). NJ: John Wiley & Son.
- Bero, L., & Rennie, D. (1995). The Cochrane Collaboration: Preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Journal of the American Medical Association*, 274, 1935-1938.
- Boruch, R., Petrosino, A., & Chalmers, I. (1999). The Campbell Collaboration: A proposal for systematic, multinational, and continuous reviews of evidence. In P. Davies, A. Petrosino, & I. Chalmers (Eds.), *The effects of social and educational interventions: Developing an infrastructure for international collaboration to prepare, maintain and promote the accessibility of systematic reviews of relevant research* (pp. 1-22). London: University College London School of Public Policy.
- Campbell Collaboration Web Site. (2008). *The Campbell Collaboration: What helps? What harms? Based on what evidence?* Retrieved September 10, 2008 from the World Wide Web: <http://www.campbellcollaboration.org/>
- Caldwell, B. J. (1994). *Beyond positivism: Economic methodology in the twentieth century*.

- London: Routledge.
- Coalition for Evidence-Based Policy (2002). *Bringing Evidence-Driven Progress: A Recommended Strategy for the U.S. Department of Education*. Retrieved February 28, 2009 from the world wide web: <http://www.excelgov.org/usermedia/images/uploads/PDFs/coalitionFinRpt.pdf>
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10, 101-129.
- Cochrane Collaboration Web Site. (2008). *The Cochrane Collaboration*. Retrieved September 10, 2008 from the world wide web: <http://www.cochrane.org/>
- Cooper, H. M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research*, 52, 291-302.
- Cooper, H., & Hedge, L. V. (1994). *The handbook of research synthesis* (Eds.). New York: Russell Sage Foundation.
- Cuijpers, P., van Straten, A., Warmerdam, L., & Andersson, G. (2008). Psychological treatment of depression: A meta-analytic database of randomized studies. *BMC Psychiatry*, 8(36), Retrieved September 24 from the world wide web: <http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=2408566&blobtype=pdf>
- Davies, H. T. O., Nutley, S. M., & Smith, P. C. (1999). Editorial: What works? The role of evidence in public sector policy and practice. *Public Money & Management*, 19(1), 3-5.
- Davies, H. T. O., Nutley, S. M., & Smith, P. C. (2000). *What works? Evidence-based policy and practice in public services* (Eds.). Bristol: The Policy Press.
- Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, 47(2), 108-121.
- De Vibe, M. D., Hovland, I., & Young, J. (2002). *Bridging research and policy: An annotated bibliography (Working paper No. 174)*. London: Overseas Development Institute. Retrieved February 28, 2009 from the world wide web: <http://www.odi.org.uk/rapid/Publications/Documents/wp174.pdf>
- Dickersin, K., Manheimer, E., Wieland, S., Robinson, K. A., Lefebvre, C., & McDonald, S. (2002). Development of the Cochrane Collaboration's CENTRAL register of controlled clinical trials. *Evaluation & the Health Professions*, 25, 38-64.
- Dixon-Woods, M., Bonas, S., Booth, A., Jones, D. R., Miller, T., Sutton, A. J., Shaw, R. L., Smith, J. A., & Young, B. (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research*, 6, 27 - 44.
- Egger, M., Smith, G. D., & Altman, D. (2001). *Systematic reviews in health care: Meta analysis in context (2nd ed.)*. London: BMJ Books.
- Elliott, J. (2001). Making evidence-based practice educational. *British Educational Research Journal*, 27(5), 555-574.
- EPPI-Centre Web Site. (2008). *Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre)*. Retrieved September 10, 2008 from the World

Wide Web: <http://eppi.ioe.ac.uk/cms/>

- Evans, J., & Benefield, P. (2001). Systematic reviews of educational research: Does the medical model fit? *British Educational Research Journal*, 27(5), 527-541.
- Feldman, K. A. (1971). Using the work of others: Some observations on reviewing and integrating. *Sociology of Education*, 4, 86-102.
- Glass, G. V. (1976). Primary, secondary and meta-analysis of research, *Educational researcher*, 5, 3-8.
- Glass, G. V. (1982). Meta-analysis: An approach to the synthesis of research results. *Journal of research in science teaching*, 19(2), 93-112.
- Gough, D. (2004). Systematic research synthesis. In G. Thomas & R. Pring (Eds.), *Evidence-based practice in education* (pp.44-62). New York: The Open University Press.
- Grégoire, G., Derderian, F., & LeLorier, J. (1995). Selecting the language of the publications included in a meta-analysis: is there a Tower of Babel bias? *Journal of Clinical Epidemiological*, 48, 159-163.
- Hammersley, M. (1997). Educational research and teaching: A response to David Hrgreaves' TTA lecture. *British Educational Research Journal*, 23(2), 141-161.
- Hammersley, M. (2001). On 'systematic' reviews of research literatures: A 'narrative' response to Evans & Benefield. *British Educational Research Journal*, 27(5), 543-554.
- Hargreaves, D. H. (1996). Teaching as a research-based profession: Possibility and prospects. In B. Moon, J. Butcher, & E. Bird (Eds.), *Leading Professional Development in Education* (pp. 200-210). London: The Open University.
- Hargreaves, D. H. (1997). In defense of research for evidence-based teaching: A rejoinder to Martyn Hammersley. *British Educational Research Journal*, 23(4), 405-419.
- Hargreaves, D. H. (1999). The Knowledge-creating School. *British Journal of Educational Studies*, 47(2), pp. 122-144.
- Harwell, M., & Maeda, Y. (2008). Deficiencies of reporting in meta-analyses and some remedies. *The Journal of Experimental Education*, 76(4), 403-428.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Higgins, J. P. T., & Greens, S. (Eds.). (2006). *Cochrane handbook for systematic reviews of interventions 4.2.6* (updated September 2006). Retrieved September 10, 2008 from the world wide web: <http://www.cochrane.org/resources/handbook/hbook.htm>
- Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(1), 539-558.
- Hoffert, S. P. (1997). Meta-analysis is gaining status in science and policymaking. *The Scientist*, 11(18), 1-6.
- Hunter, J. E., & Schmidt, F. L. (1996). Cumulative research knowledge and social policy formulation: The critical role of meta-analysis. *Psychology, Public Policy, and Law*, 2,

- 324-247.
- Jackson, G. B. (1980). Methods for integrative reviews. *Review of Educational Research*, 50, 438-460.
- Jadad, A. R., Cook, D. J., Jones, A., Klassen, T. P., Tugwell, P., Moher, M., & Moher, D. (1998). Methodology and reports of systematic reviews and meta-analyses: A comparison of Cochrane reviews with articles published in paper-based journals. *Journal of American Medical Association*, 280(3), 278-280.
- Lin, N. (1999). Social networks and status attainment. *Annual Review of Sociology*, 25, 467-487.
- Lipsey, M. W. (1994). Identifying potentially interesting variables and analysis opportunities. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 111-123). New York: Russell Sage Foundation.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, London: Sage Publications.
- Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, T., & Walsh, S. (1995). Assessing the quality of randomized controlled trials: An annotated bibliography of scales and checklists. *Controlled Clinical Trials*, 16, 62-73.
- Moher, D., Tsertsvadze, A., Tricco, A. C., Eccles, M., Grimshaw, J., Sampson, M., & Barrowman, N. (2007). A systematic review identified few methods and strategies describing when and how to update systematic reviews. *Journal of Clinical Epidemiology*, 60, 1095-1104.
- Morrison, K. (2001). Randomized controlled trials for evidence-based education: Some problems in judging 'what works'. *Evaluation and Research in Education*, 15(2), 69-83.
- Mulrow, C. D. (1994). Rationale for systematic reviews. *British Medical Journal*, 309, 597-599.
- Oakley, A. (2001). Making evidence-based practice education: A rejoinder to John Elliott. *British Educational Research Journal*, 27(5), 575-576.
- Petrosino, A. J., Boruch, R. F., Rounding, C., McDonald, S., & Chalmers, I. (2000). The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) to Facilitate the Preparation and Maintenance of Systematic Reviews of Social and Educational Interventions. *Evaluation and Research in Education*, 14, 206-219.
- Petrosino, A., Boruch, R. F., Soydan, H., Duggan, L., & Sanchez-Meca, J. (2001). Meeting the challenges of evidence-based policy: The Campbell Collaboration. *The ANNALS of the American Academy of Political and Social Science*, 578(1), 14-34.
- Prout, H. T., & DeMartino, R. A. (1986). A meta-analysis of school based studies of psychotherapy. *Journal of School Psychology*, 24, 285-292.
- Prout, S. M., & Prout, H. T. (1998). A meta-analysis of school-based studies of counseling and psychotherapy: An update. *Journal of School Psychology*, 36(2), 121-136.

- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments* (Eds.). NJ: John Wiley & Son.
- Sanderson, I. (2003). Is it 'what works' that matters? Evaluation and evidence-based policy-making. *Research Papers in Education*, 18(4), 331-345.
- Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15-21.
- Soll, R. F. (2008). Updating reviews: The experience of the Cochrane Neonatal review group. *Paediatric and Perinatal Epidemiology*, 22, 29-32.
- Sutton, A. J., & Higgins, J. P. T. (2008). Recent developments in meta-analysis. *Statistics in Medicine*, 27, 625-650.
- Taveggia, T. C. (1974). Resolving research controversy through empirical cumulation: Toward reliable sociological knowledge. *Sociological Methods and Research*, 2, 395-407.
- U.S. Department of Education. (2002a). *No Child Left Behind: A desktop reference*. Washington, DC: Author.
- U.S. Department of Education. (2002b). *Strategic Plan, 2002-2007*. Washington, DC: Author.
- What Works Clearinghouse. (2008). *What works clearinghouse*. Retrieved September 10, 2008 from the World Wide Web: <http://ies.ed.gov/ncee/wwc>
- White, V. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91(3), 461-481.
- Wortman, P. M. (1994). Judging research quality. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 97-109). New York: Russell Sage Foundation.

An Introduction and Suggestion on Systematic Review and meta-analysis in Domestic Education

Hsieh Jin-Chang²

Assistant Research Fellow,
National Academy for Educational Research PreParatory Office

Abstract

The belief of evidence-based education has been becoming widely accepted and promoted by many domestic and international scholars. And one of the best strategies to accomplish and disseminate this belief is systematic review, which is rarely discussed in domestic education field. Therefore, the purpose of this paper is to introduce the concept of systematic review and to describe the development and possible dilemma domestic scholars may face. Furthermore, some suggestions were proposed in dealing with the dilemma by referring to establishment of systematic review groups in international education and medicine domain. At the end, some debates and prospects in disseminating the belief of evidence-based education by systematic review are also proposed and discussed. We hope these would be referred by domestic practitioners and researchers in the future.

Key word : Evidence-based education, Meta-analysis, Systematic review,
Systematic review group

². The authors would like to thank the National Science Council of the Republic of China, Taiwan for financially/partially supporting this research under Contract No. NSC-98-2511-S-656-001-MY3