

個別研究品質好壞有關係嗎？ 類比教學對於學生科學概念改變 成效後設分析*

謝進昌

國家教育研究院
測驗及評量研究中心

邱怡靜

國立臺灣師範大學
教育心理與輔導學系

隨著「以證據為基礎的教育」理念的興起，透過量化研究統合，以提出科學實徵證據，逐漸受到重視，進而使得如何判讀個別研究品質好壞，也成為重要議題。本研究採用系統性文獻回顧與後設分析，針對類比教學主題進行研究統合，目的除探討類比教學對於學生科學概念改變的統合成效外，焦點在於探討一般學者慣用指標（出版類型）與其它研究品質評估指標對於統合結果的影響，以尋求能兼顧效度理論與實務可行之替代性個別研究品質評估指標。本研究結果發現：1. 類比教學整體具有中等教學效果；2. 在未考量個別研究品質時，單一類比教學成效傾向低於非單一類比教學，如類比橋、多重類比教學成效，而針對國小學習階段的參與者使用非單一類比教學，具有最佳教學成效，然而，若同時考量不同個別研究品質時，其顯著調節效果會變成不顯著，顯示研究品質評估具有相當影響力；3. 以「出版類型」作為個別研究品質評估，容易同時涉及出版偏誤議題，而本文建議使用替代性研究品質評估指標，即同時符合前測表現基準可比較性、評量工具信效度、與基本參與者人數等三指標，似乎更能兼顧效度理論、實徵結果與實務可行性。文末，提出幾點結論與建議，供未來研究者參考。

關鍵詞：類比教學、概念改變、後設分析、量化研究統合、個別研究品質評估

* 1. 本文通訊作者：邱怡靜，通訊方式：80701010e@ntnu.edu.tw。

2. 感謝科技部經費支持，以利推動本研究，計畫編號：MOST 107-2410-H-656-007。

後設分析研究者進行量化研究統合 (quantitative research synthesis) 時, 為使文獻蒐集更具客觀與全面性, 其策略多強調窮竭式研究文獻搜尋 (exhaustive literature search), 針對不同出版類型的 research 文獻進行有系統性檢索, 使研究者不會遺漏任何類型的 research 文獻 (Cooper et al., 2009)。然而, 隨美國「不讓任何孩子落後」(No Child Left Behind Act) 法案的簽署, 強調「以科學研究為基礎」(scientifically based research), 提出研究者應以嚴格而客觀的科學方式, 進行系統性資料分析, 以驗證其研究假設與論證, 使得以證據為基礎的教育 (evidence based education) 思維興起, 遂愈來愈多學者、組織逐漸重視「使用良好個別研究品質, 以建立具效度之統合研究證據」, 例如, 美國的有效教育策略資料中心 (What Works Clearinghouse [WWC], 2020a) 或 Slavin (2008) 所提以最佳研究品質文獻進行統合以形成證據 (best evidence synthesis) 的概念。在此浪潮下, 有愈來愈多後設分析研究者, 例如, Sung 等人 (2016) 以「文獻出版類型」作為研究品質判斷初步指標, 僅納入發表在「同儕評閱期刊」的研究文獻作為統合標的, 此舉雖符合現行學者對於科學研究品質評估方式的認同, 但在量化研究統合脈絡下, 容易伴隨出版偏誤 (publication bias) 或結果呈現偏誤 (outcome report bias) (Pigott et al., 2013) 影響, 此外, 是否存在其它較適合、符合效度理論與實際可行之個別研究品質評估指標等, 這些議題現行較少為國內學者關注、探討。

有關主題聚焦方面, 作者在以數個關鍵字, 包含: 後設分析、整合分析、綜合分析、統合分析、系統性文獻評閱、系統性文獻回顧、彙總研究、meta analysis、systematic review、meta research、research synthesis、meta analytic studies 等, 針對國內現行完整收錄國內期刊、論文之資料庫進行搜尋時, 發現國內教學介入實驗量化統合研究主題, 多為資訊科技輔助、閱讀教學、創造或批判思考教學介入等主題, 較少涉及促進學生科學概念改變, 尤其科學教育頗受重視的類比教學, 故該主題的統合效果成為本研究重點。

類比 (analogy) 為人類認知的核心組成部分, 其涉及分類和學習, 亦提供了思考和解釋的工具, 其中, 類比是透過類比物相似的結構與具象化表徵, 連結學習者先備知識, 以建立新知識與抽象概念, 達到有效學習遷移, 因此類比成為學生是否產生概念改變中重要的軸心 (Duit, 1991), 也被視為促進概念改變有效的教學方式。然而, 過去研究發現並非所有類比教學研究皆呈現正向教學研究結果 (Thagard, 1992), 一如 Glynn (1991) 認為類比教學猶如雙面刃, 並非所有類比教學活動皆有益於學生學習遷移, 部分研究顯示類比物與類比標的相似性不足或限制, 可能使學習者產生迷思概念, 反而有礙新知識的建構 (Clement, 1987; Glynn et al., 1989)。整體而言, 類比教學實際效果未有定論外, 其影響是否會隨學生學習階段, 或不同個別研究品質判讀方式而具差異, 皆為本研究探究的重點。

本研究透過系統性文獻回顧與後設分析 (systematic review and meta analysis), 針對類比教學對於學生科學概念改變成效進行量化研究統合, 前者係以標準化程序進行有系統文獻蒐集與定位, 後者係針對收集的量化訊息, 進行量化統合分析, 而有關調節變項 (moderators), 本研究聚焦不同個別研究品質指標為首要, 包含文獻出版類型、前測表現基準是否均等、評量工具信效度、參與者人數等。茲依續條列本研究所擬回應之待答問題如下:

1. 經排除極端值, 探討類比教學對於促進學生科學學習概念改變之統合平均效果量為何?
2. 在問題一基礎下, 探討不同類比教學方式對於不同學習階段學生於科學學習改變之調節效果為何?
3. 在問題二基礎下, 探討不同個別研究品質評估指標 (出版類型、前測表現基準是否均等、評量工具信效度、參與者人數) 分別對於類比教學調節效果為何?

(一) 促進學生概念改變相關教學之統合研究

在回顧過去有關學生概念改變實驗教學之回顧性或統合研究文獻, 其主題多集中於概念構圖 (concept map) (Schroeder et al., 2018)、提供不同類型概念改變文本 (conceptual change texts) (Guzzetti et al., 1993), 前者在於提供鷹架, 以促進學生概念學習, 而後者在於提供學生對照反駁 (refutational) 的文本, 以突顯特定概念的迷思, 然而, 不論何者皆較鮮少涉及類比教學層面。

類比在於連結學生先備知識與教學標的間概念, 以達到學習遷移效果, 但由於類比是透過相似類比物的結構、特徵, 與目標物產生兩個不同概念體系連結 (Glynn, 1991), 若學習者對類比物不

熟悉，或所選取的類比物與標的物關係不明顯時，類比可能較難促進學習效果（Clement, 1987），甚至可能使學習者形成另有概念，阻礙新概念建立。因此，Spiro 等人（1989）建議當學習的新概念較複雜、結構較不完整時，可採用多重類比（multiple analogies）的策略，以新類比物補充、校正、變更等方式彌補原始類比物的不足，然而，提供多個類比卻也可能使學習者產生額外的認知負荷，以致影響教學成效。後續，相較於多重類比重視透過不同類比物間互補特性，Brown 與 Clement（1989）提出類比橋（bridging analogies）概念，利用學習者直觀思維為中介類比物，建立類比物和標的物關係，以避免過多類比物產生的干擾與負荷，進而達到學習新概念的目的，此外，電腦輔助類比動畫圖形教學是以電腦動畫形態來呈現類比物，藉由動態的視覺化影像，具體化概念關聯，以協助學生發現概念間變化關係（陳盈吉，2004）。

檢視過去個別實徵研究出現正反兩面的結果，類比如同雙面刃，部分學者發現適切的類比可幫助學習者建立抽象而複雜的新概念，並有助於概念改變（Clement, 1989; Duit, 1991），但不當使用類比非但無法協助學生建立正確的新概念，反而可能會增加學習者認知負荷，甚至迫使學習者產生迷思概念（Clement, 1987; Glynn et al., 1989; Spiro et al., 1989）。

（二）學生年級（齡）與類比教學效果

在檢視學生年級與類比教學成效關係時，Centner 與 Toupin（1986）認為學生瞭解目標物與其對應物間的系統性（systematicity）關係、目標物與其對應物相似程度的清晰度（transparency），對類比資訊傳遞的精準度存在重要意義，他們在針對 4-6 歲與 8-10 歲孩童進行遷移表現（transfer performance）研究中，發現兩組孩童皆會受類比清晰度影響，隨著對應的清晰度降低，遷移準確性會急速下降，然而，系統性僅會影響年齡較大的參與者。此外，Gentner（1989）在進行跨年齡研究中，亦發現當目標物與類比物間愈具結構、系統性，於類比連結時，學習者愈易覺察類比間對應，類比間連結也更趨完整，且隨學習者年齡增長，使用類比的能力會提高，更能促進學習者發現系統性的關係結構對應。

若就訊息處理與認知相關研究進行探討，Keane（1990）提出工作記憶限制（working memory constraints）、背景知識影響（influence of background knowledge）概念，認為工作記憶容量有限，當超出學習者工作記憶所能處理負荷量時，將造成學習者學習困難或類比錯誤，同時，背景知識會影響學習者類比對應的運作，亦會對學習產生負面效果，且工作記憶與表現會隨著年齡增加而逐漸衰退（Nyberg et al., 1997）。據此，類比教學重視目標物與其對應物間系統性與清晰度，而隨著學習者年齡增長，有助於其察覺兩者關係，然而，學習者工作記憶、背景知識與年齡產生共伴效果，使得年齡對於類比教學效果影響，更顯不確定性。

（三）量化研究統合脈絡下之個別研究品質評估

證據為基礎的教育理念推動，促使在研究統合脈絡下，「如何判讀個別研究品質」議題更顯重要，若回顧過去研究，可發現最常被研究者普遍使用的指標為文獻出版類型，即將經過同儕互評或專家認可的期刊視為具備品質之研究（Hall et al., 1988），在進行統合研究時，透過此指標作為品質判讀依據，其優點在於不僅具一定效度基礎，且簡單即可辨別研究好壞，卻也容易同時引發出版偏誤（publication bias）（Lipsey & Wilson, 1993）或結果呈現偏誤（outcome report bias）（Pigott et al., 2013），其偏誤來源皆指向統合研究者所蒐集的研究文獻，無法代表現行全部研究現況，而此被遺漏之研究屬非隨機遺漏（missing not at random）。因此，使用文獻出版類型作為個別研究品質判讀依據，雖然具有效度與簡便優勢，但於統合研究時，容易與其它偏誤議題相互共伴影響。

在統合研究中，過去學者發展出許多個別研究品質檢核指標，例如，Slavin（2008）、Valentine 與 Cooper（2008）、WWC（2020a）等，這些學者或組織對於指標發展，多以實驗教學效度理論為開端，配合如 Shadish 等人（2002）提出效度架構，但部分指標往往缺乏實際可行性考量，例如，WWC（2020a）非常強調參與者隨機分派，然而，在實際教育實驗研究設計中，卻鮮少真正以完全個體隨機分派方式進行。據此，本研究企圖平衡效度理論與實際現況，提出兩組別前測表現基準均等、評量工具信效度、參與者人數等替代性評估指標，不僅符合 Shadish 等人（2002）所重視內部、

外部、構念與統計結論效度，同時，這些指標的相關訊息也多存在於一般研究文獻，具實際分析可行性。

於內部與統計結論效度層面，研究者為確保實驗組與對照組參與者於前測表現基準均等（equivalent）或可比較（comparability），除了透過實驗研究設計外，例如，隨機分派參與者於實驗、對照組、使用迴歸－不連續設計（regression discontinuity design）等，多數研究會同時搭配使用統計分析（例如，使用單因子共變數分析），以校正可能存在的前測表現不均等，此時，相對應效果量（effect size）計算，就存在原始分數平均數效果量（mean-raw score）、共變數校正分數平均數效果量（mean-covariate adjusted score）的選擇（Morris & DeShon, 2002）議題。若教學研究使用適當研究設計，以確保前測表現基準可比較性，此時，不論使用哪種效果量計算途徑，皆能確保其符合此品質評估指標，然而，若研究並未透過研究設計，而是使用統計校正方式，此時，若該研究未使用對應之共變數校正效果量進行計算，則不能判讀該研究符合此指標，但此點卻容易被部分統合研究所忽略，即不論該個別實驗研究設計為何，皆一律使用原始分數平均效果量進行效果量計算，並視為具有相同研究品質。

除了「實驗研究設計－效果量計算途徑選擇」外，影響前測表現基準可比較性因素，尚存在「前測表現差異效果量是否過大」、「參與者整體與差異流失率」。就前者而言，Slavin 與 Lake（2008）認為若兩組別前測表現差異大於 0.5 個標準差，不論使用哪種統計校正，皆無法確保兩組表現基準具可比較性；就後者而言，學者認為若參與者於實驗中流失，會產生流失偏誤（withdrawal or dropout bias），以致影響實驗解釋與研究有效性，其中，Valentine 與 Cooper（2008）認為整體流失率應控制在 50% 範圍，實驗與對照組間差異流失率應控制在 10% 內。

在評估構念效度層面，研究者認為直接判讀個別研究所使用評量工具信、效度為較直接途徑，WWC（2020a）提出評量工具信度愈高，則其測驗誤差愈低，其中，使用內部一致性信度、評分者信度、重測信度為一般學者較慣用的方式，認為前兩者的分析結果至少得 0.5 以上，而重測信度至少為 0.4 以上，才符合基本準則。效度方面，研究者多慣以專家判讀方式建立評量工具內容品質或採用其它外在效標來進行效標關聯評估。至於統計結論效度方面，Slavin（2008）主張參與者人數必須具備基本門檻值的重要性，其論點涉及統計考驗力（power），他認為參與者人數會對於效果量產生重大影響，小型研究容易產生極端正或負效果量，以致錯誤高（低）估教學效果，因此，其後續應用研究，如 Slavin 與 Lake（2008）皆建議以實驗、對照組各具有至少 15 人，為基本達到研究品質之要求。

方法

本研究為量化統合研究，係使用系統性文獻回顧與後設分析技術等標準化程序進行研究。前者乃針對類比教學主題，有系統進行文獻蒐集與定位，以避免可能的文獻蒐集偏誤，並利用後者量化統計技術，經選擇適當效果量計算個別研究教學成效、調節變項分析等，以回應本研究待答問題。茲針對系統性文獻蒐集、研究納入與排除準則、指標建立、訊息建檔與文獻搜集過程維護、效果量計算與後設迴歸分析等，逐一說明如下：

（一）系統性文獻蒐集

本研究文獻檢索以期刊與博碩士論文為標的，並以過去曾發表相關回顧研究之參考文獻為複核依據，以避免遺漏重要著作。茲別文獻蒐集過程，說明如下：

1. 國內資料庫檢索

研究者以「類比、analogical reasoning、analogy、analogies」等詞彙於臺灣博碩士論文系統之論文名稱及關鍵字進行檢索，計回傳 1567 筆；再以前述詞彙於臺灣期刊論文索引系統之篇名與關鍵字進行檢索，共回傳 783 筆文獻；此外，再以前述詞彙於 Airiti Library 華藝線上圖書館之篇名、關鍵字檢索臺灣出版文獻，共回傳 599 筆期刊文章、1270 筆碩博士論文（截至 2011 年 11 月 5 日）。

2. 國外資料庫檢索

有關國外文獻資料庫檢索，研究者是以 analogical reasoning、analogy、analogies 等關鍵字，以指令就論文名稱和關鍵字進行搜尋。就論文名稱比對方面，在搜尋 EBSCOHOST 系統，包含 ERIC、PsycArticles、Academic Search Complete、Education Research Complete、Science Reference Center、Teacher Reference Center 等，初步回傳 3088 筆資料；於搜尋 Social Sciences Citation Index (Web of Science)、與 ProQuest Dissertations & Theses Database 時，分別回傳 1692、720 筆資料。在關鍵字比對方面，針對 ERIC、Academic Search Complete、Education Research Complete、PsycArticles 等資料庫檢索，分別回傳 360、412、105、131 筆資料，而就 ScienceDirect 進行檢索，回傳 1593 筆期刊（截至 2011 年 04 月 22 日止）。

3. 主題相關回顧研究之參考文獻複核

過去直接關乎類比教學的回顧或量化統合研究較缺乏，多屬回顧性研究，如：Dagher (1995)，此外，相關主題為有關科學概念改變者，包含 Guzzetti 等人 (1993)。研究者針對這類研究的參考文獻進行複核，以避免可能遺失的重要研究。

(二) 研究納入與排除準則

1. 刪除非類比教學研究文獻

本研究所論及的類比教學，包含採單一類比、類比橋、多重類比等教學研究文獻，但排除單純類比教學評論或介紹、不同階段樣本類比能力的發展、其它無關類比教學研究者，如：郭人仲 (1994) 一文為單純介紹類比教學模式等。

2. 刪除質性研究或未提供適當訊息以計算效果量文獻

該研究文獻若能滿足第一項準則，卻以分析質性資料為研究取向，亦為本研究刪除標的，例如：以觀察或訪談方式進行訊息收集，並無法轉換為效果量。

3. 僅納入探討科學概念學習表現之文獻

本研究所探討依變項以涉及科學概念學習表現為主軸，例如：自然、科學等著重迷思概念 (misconcept) 教學的學科，而未納入探討高層次能力、動作技能、語言之文獻，如：語言 (黃秀玉，2009) 等，其考量點在於語言類的類比多半涉及文字類比，其類比形態較不似科學範疇所使用。

4. 納入在學參與者及不同出版時間之文獻

本研究納入研究參與者年齡或年級是以在學學生為核心，包含學前、國中小、高中職至大學等階段，此外，根據本研究所採用的檢索系統，納入文獻的出版時間是介於民國 45 年 – 民國 100 年之博碩士論文及民國 59 年 – 民國 100 年的期刊 (檢索時間至 2011 年 11 月 5 日)。

(三) 指標建立、訊息建檔與文獻搜集過程維護

1. 指標建立、訊息建檔

本研究先針對蒐羅到的文獻，比對納入與排除準則進行篩選，於資料建檔前，研究者先與建檔者說明各指標內涵與分類，再由建檔者練習分類，並針對指標判斷有疑慮的部分進行討論，確認建檔者熟悉建檔方式後，由兩位建檔者正式施行訊息建檔。當兩位建檔者所建資料不一致時，研究者會回溯原文進行交叉驗證，以重新判定指標訊息，最後將資料進行簡化及轉碼。編碼內容包含文獻背景特徵、參與者、教學者、實驗教學與評量工具、計算效果量相關訊息等面向。其中，文獻背景記載作者、年代、出版類型；參與者訊息涵蓋參與者人數、學習階段、流失率；教學者標示教學者

人數、教學者是否為研究者；實驗教學與評量詳述類比教學法類型與教學取徑、實驗教學期程、實驗設計、評量工具信效度，而效果量計算會依照原始研究所提供前、後測資料，採用原始分數或共變項校正分數平均效果量進行估計。

在上述指標訊息建立中，屬於直觀、具體可分類指標者，不再贅述其內涵，部分可能存在主觀判讀、模糊空間的重要指標，以下將進行更具體說明：

(1) 類比教學取徑。本研究分為單一類比、多重類比、類比橋等，其中，單一類比教學是指實驗中對於單一概念之教學，僅採用單一類比進行教學；相對的，多重類比教學是對於單一概念之教學實驗中，採用多個類比物進行教學，且不同類比物間具互補性；至於類比橋教學則是採用學習者直觀思維為中介類比物，建立類比物和標的物關係，最後，電腦輔助類比動畫圖形教學是以電腦動態視覺化影像，以協助學生理解類比與概念間關係。

(2) 出版類型。本研究區分未出版研究與已出版研究兩大類，其中，未出版研究包含有博士、碩士論文兩部分，而已出版研究係包含期刊、與博碩士論文轉投期刊（此類不納入前者重複計算）兩類。

(3) 評量工具品質。本研究針對評量工具品質分為兩類，當該研究未提供評量工具信、效度具體說明時，視為未符合評量工具品質；當該研究有提供評量工具信、效度說明，同時，也符合 WWC (2020a) 建議準則（詳見文獻回顧）時，則將該筆研究歸類為符合評量工具品質。

(4) 前測表現基準均等。有關判讀研究內兩組別前測表現基準是否均等或可比較，其涉及到實驗研究設計與效果量選擇、兩組別前測表現差異效果量程度、參與者流失率等三個面向，其中，該研究於三個面向皆符合基本準則時，則視為前測表現基準均等，反之，則無。

就第一個實驗研究設計與效果量選擇面向而言，若該研究採隨機分派參與者於各組別、迴歸-不連續設計等，透過研究設計以確保前測表現基準均等者，其後不論採原始分數平均數效果量、共變數校正分數平均數效果量，皆視為符合本指標；若該研究僅使用統計校正（如，單因子共變數分析）以校正前測表現者，後續則以可計算校正分數平均數效果量者，才可視為符合本指標。

若該研究符合第一個面向，後續，仍需符合第二前測表現差異效果量、與第三參與者流失率面向的要求。就前者而言，依據 Slavin 與 Lake (2008) 建議，兩組別前測表現差異效果量不超過 0.5，視為符合本指標，反之，則無；就後者而言，就考量整體流失率應控制在 50% 範圍內、實驗與對照組間差異流失率應控制在 10% 範圍內，視為符合本指標，反之，則無。

(5) 參與者人數。依 Slavin (2008) 建議分兩類，實驗、控制組各皆低於 15 人者，歸為未具備基本人數，而實驗、控制組各皆高於 15 人，視為具備基本人數。

2. 文獻搜集過程維護

本研究在檢索國內外文獻資料庫後，初步以 EndNote 書目管理軟體，針對回傳文獻資料進行整理與重複文獻排除，後續，研究者經第一階段納入與排除準則篩選後，針對下表 1 之 149 篇文獻逐一檢視其內容，各刪除 53 篇期刊及 53 篇博碩士論文，共達 106 篇不符合準則的文獻，此外，就剩餘的 43 篇文獻中，有 8 篇文獻之研究參與者來源雷同，其原因為該文獻為博碩士論文轉投稿期刊，因此，研究者仍視為同 1 篇進行計算，最後，本文共計編碼 35 篇文獻。

表 1
文獻排除原因及其篇數

文獻排除原因	博碩士論文篇數	期刊篇數
非類比教學研究	35	43
以分析質性資料或未提供充足訊息以進行效果量計算	12	7
非探討學習表現	2	2
其它（如樣本同時接受兩種以上教學）	4	1

(四) 效果量計算與後設迴歸分析

為統合各研究間教學效果，本研究採標準化平均數差異效果量（standardized mean difference）為 Hedges 不偏 g 效果量（Hedges, 1982），其中，因不同研究所提供前、後測資料的差異，大致可將效果量分為原始分數平均數效果量、共變數校正分數平均數效果量（Morris & DeShon, 2002; WWC, 2020b）。

當原始研究文獻有提供前測表現校正相關資料時，包含如共變數分析統計量、校正平均數、標準差等，本研究使用共變項校正分數平均數效果量，其公式為：

$$g = \frac{\overline{Y}'^{G_1} - \overline{Y}'^{G_2}}{SD_{Y'}^{\text{pooled}}} \sqrt{1 - r_{XY}^2} \quad (1)$$

其中， \overline{Y}'^{G_1} 、 \overline{Y}'^{G_2} 分別為經共變項校正實驗組與控制組平均數， $SD_{Y'}^{\text{pooled}}$ 為共變項校正分數的整合標準差， r_{XY}^2 為共變項與後測相關係數平方值。此外，若原始研究文獻並未提供前述資料，而是各自提供前、後測平均數、標準差等資料時，依據 WWC（2020b）建議，當前後測評量工具一致時，可使用 difference-in-differences 進行計算，其公式為 $g = g_{\text{post}} - g_{\text{pre}}$ ，然為納入考量前後測相關性，本研究採用校正 difference-in-differences，其公式為：

$$g = g_{\text{post}} - r_{XY} * g_{\text{pre}} \quad (2)$$

g_{post} 為後測差異效果量， g_{pre} 為前測差異效果量， r_{XY} 為共變項與後測相關係數。

當原始研究文獻未提供任何前測資料，或僅提供後測相關統計量（如變異數分析統計量），研究者採用原始分數進行平均數效果量計算，而此公式也是大部分過去統合研究計算途徑，其公式為：

$$g = \frac{\overline{Y}^{G_1} - \overline{Y}^{G_2}}{SD_Y^{\text{pooled}}} \quad (3)$$

其中， \overline{Y}^{G_1} 、 \overline{Y}^{G_2} 分別為實驗組與控制組原始平均數， SD_Y^{pooled} 為整合的標準差。

在進行後設分析以統合量化資料結果與探討調節變項時，本研究預先使用 Viechtbauer 與 Cheung（2010）指標，針對影響與極端值進行評估與刪除，並以 Q 統計量與 Higgins 與 Thompson（2002） I^2 指標，來檢定研究間可能存在的異質性，隨後採用混合效果模式－後設迴歸分析（mixed effect model-meta regression）（Thompson & Higgins, 2002）進行待答問題探討，其一般估計模式表達為：

$$g_i = \beta_0 + \beta_1 M_{i1} + \dots + \beta_p M_{ip} + \mu_i + \varepsilon_i \quad (4)$$

g_i 為不同研究（1 至 i 個）效果量、 β_0 為截距、 β_1 至 β_p 為研究者所欲探討之調節變項 M （1 至 p 個）所估計對應迴歸係數值，顯示該調節變項間差異或影響程度、 μ_i 為不同研究間效果量之隨機變異，服從常態分配 $N(0, \tau^2)$ ，當 τ^2 愈大時，表研究間有待解釋調節變項、 ε_i 為不同研究內抽樣誤差。此外，在分析不同調節變項解釋力時，本研究使用 Raudenbush（2009）提出 $R^2 = (\tau_{RM}^2 - \tau_{ME}^2) / \tau_{RM}^2$ ，其表未投入某特定調節變項前，原始隨機效果模式所估計組間變異量（ τ_{RM}^2 ），減去投入該特定調節變項後，混合效果模式所估計組間殘差變異量（即 τ_{ME}^2 ）差異值，相對於原始變異量之比值，數值愈高，代表投入該調節變項，愈能解釋整體研究間異質性。

結果與討論

(一) 國內外類比教學對於學生科學概念學習實驗研究概況

研究者經對照前述納入與排除準則，針對剩餘 35 篇文獻進行資料鍵檔，然而，由於單一研究文獻，其實驗設計可能存在多個實驗組與對照組的情況，因此，總計產生 45 筆兩兩成對比較研究結果，其分佈概況如表 2 所示。

在文獻背景特徵方面，文獻出版類型分別為 3 筆博士論文、20 筆碩士論文、17 筆期刊及 5 筆論文改投期刊者；在這些文獻中，類比教學採用單一類比者，為 19 筆、類比橋 10 筆、多重類比 12 筆、電腦輔助類比動畫圖形教學 2 筆。在參與者學習階段方面，其分佈多以國中為主（25 筆）、其次是國小（9 筆），顯示類比實驗教學多施教於中等以上學生；至於研究評量工具方面，其來源來多為研究者自編評量工具（40 筆），只有 1 筆使用既有評量工具。

於檢視個別研究品質方面，在前測表現基準均等性中，未說明透過研究設計或統計校正，以確保前測表現基準者，有 4 筆，前測表現差異大於 0.5 標準差者 3 筆、參與者流失率過大者 1 筆，皆屬於不符合本指標；至於評量工具方面，未提供評量工具信、效度具體說明者有 8 筆，有提供且符合信效度數據者 37 筆；最後，有 4 筆研究屬於小型研究，實驗與控制組人數各低於 15 人，多數 41 筆研究皆能符合最低人數要求。整體而言，上述三指標皆符合標準者，大致兼具效度各層面者，計有 29 筆，有 16 筆研究不符合任一指標標準。

表 2
本文所納入研究文獻概況

研究背景特徵	調節變項虛擬建檔	細項	比較研究數
出版年代	--	1995 (含) 前	7
		1996-2000	12
		2001-2005	15
		2006-2010	11
出版類型	未出版文獻 (Ref.)	博士論文	3
		碩士論文	20
	已出版文獻	期刊	17
		論文轉投期刊	5
類比教學	單一類比 (Ref.)	單一類比	19
		類比橋	10
	非單一類比	多重類比	12
		電腦輔助類比動畫圖形教學	2
	未說明 (NA)	未說明	2
參與者學習階段	國小 (Ref.)	國小	9
		國中	25
	國中 (含) 後	高中職	4
		大學	4
		跨學習階段 (2 筆國高中、1 筆 18 歲至成年)	3
前測表現基準均等	前測表現基準不均等 (Ref.)	未說明透過研究設計或統計校正 (且選擇適當效果量) 以確保前測表現之可比較性	4
		兩組別前測表現差異效果量 > 0.5	3
		兩組別參與者流失率差異大於 10%	1
		前測表現基準均等	符合準則

(續下頁)

表 2
本文所納入研究文獻概況 (續)

研究背景特徵	調節變項虛擬建構	細項	比較研究數
評量工具來源 (NA)	自編	研究者自編評量工具	40
	引用	引用既有標準化評量工具	1
	未說明	未說明	4
評量工具信、效度	不符標準 (Ref.)	未提供評量工具信、效度具體說明	8
	符合標準	提供評量工具信效度相關說明、數據且符合基本水準	37
參與者人數	未具備基本人數 (Ref.)	實驗、控制組各皆低於 15 人	4
	具備基本人數	實驗、控制組各皆高於 15 人	41
三指標 (前測表現基 準、評量工具信效度、 總人數) 皆符合	任一指標不符合 (Ref.)		16
	皆符合		29

註：Ref. 指執行後設迴歸分析時，針對該類別變項進行虛擬變項建構轉換之參照水平。NA 係執行後設迴歸分析時，因比較研究數過少且不適合合併，視為缺失值，不計入調節變項分析。

(二) 類比教學對於學生科學學習表現影響統合結果

類比教學對於促進學生科學概念學習表現統合效果，如表 3 所示，在立即教學效果 45 筆資料中，其整體平均效果量為 0.65，若依照 Cohen (1992) 建議，為具有中強度教學效果，在評估研究間異質性，由 $Q(T)$ 值與 I^2 可發現不同研究間存在異質性，進一步檢核可能影響值，發現陳婉茹 (2003) 內兩筆比較研究，其效果量分別為 $g = 3.33$ 、 $g = 3.29$ ，經刪除該影響值後，整體平均效果量略降至 0.59，而各研究立即效果分佈森林圖，如圖 1 所示，多數研究效果量介於平均值兩側且效果量幾乎在 0 以上。此外， $Q(T)$ 值由 278.29 降至 218.11、 I^2 由 84.19 降至 80.74，結果顯示仍可能存在其它待解釋調節變項，有待後續探討。在延宕效果，14 筆資料整體平均效果量為 0.47，屬中等效果，經刪除 1 筆潛在影響值 (陳婉茹, 2003) ($g = 2.21$) 後，整體平均效果量降為 0.39 且 $Q(T)$ 值由 37.11 降至 18.43， I^2 由 64.97 降至 34.89，異質性檢定為不顯著，顯示現行類比教學對於延宕效果具有中等教學效果，而圖 2 延宕效果森林圖，可發現多數研究皆分佈平均值兩側且其效果量皆在 0 以上。

表 3
類比教學促進學生科學概念學習表現統合效果

類別	k	Hedges's 不偏 g	p_g	95%CI		$Q(T)$	$p_{Q(T)}$	I^2	τ^2
				LL	UL				
立即後測 (全部)	45	0.65	< .0001	0.52	0.79	278.29	< .0001	84.19	0.16
立即後測 - 刪除影響值	43	0.59	< .0001	0.47	0.71	218.11	< .0001	80.74	0.12
延宕後測 (全部)	14	0.47	< .0001	0.27	0.68	37.11	0.0004	64.97	0.09
延宕後測 - 刪除影響值	13	0.39	< .0001	0.24	0.54	18.43	0.1032	34.89	0.03

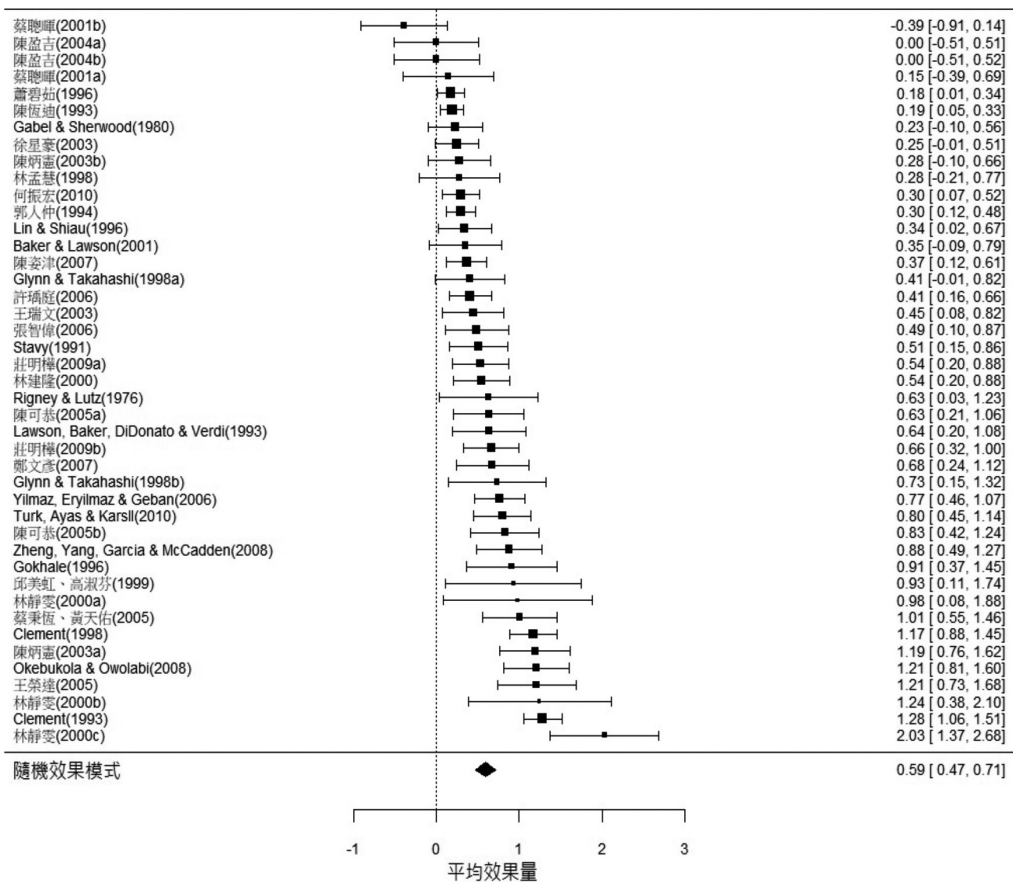
註： k = 總比較研究數； p_g = 整體平均效果量顯著考驗；CI = 整體平均效果量 95% 信賴區間；LL = 整體平均效果量最下限；UL = 整體平均效果量最上限； $Q(T)$ = 整體異質性評估 Q 值； p_Q = $Q(T)$ 值的顯著考驗； I^2 = Higgins 與 Thompson (2002) 異質性評估值。

(三) 影響類比教學效果之個別調節變項分析

就回應本研究待答問題二，其結果如表 4 所示，在探討不同出版類型對於類比教學效果之調節分析時，可發現已出版研究的迴歸係數 β 值為 0.31 ($p\beta = 0.009$) 達到顯著水準，未出版研究的統合平均效果量為 0.44 (即截距項)，已出版研究的統合平均效果量，依前述估計模式，可得 $g_i = \beta_0 + \beta_1 M_{i1} = 0.44 + 0.31 * 1 = 0.75$ ，顯示研究出版與否存在顯著差異的效果，本調節變項整體可解釋量為 8.20%。以研究品質觀點，研究者若僅篩選已出版期刊文獻作為統合標的，不僅可能高估教學效果，且依 Lipsey 與 Wilson (1993) 認為已出版研究效果量，顯著高於未出版研究效果時，可能受出版偏誤影響，使研究者若單純僅以「出版與否」作為個別研究品質評估，或作為研究文獻排除準則，容易讓統合效果摻雜其它影響干擾。

在探討類比教學取徑方面，經分析單一類比相較於其它如類比橋、多重類比、電腦輔助類比動畫等成效，其 β 值分別為 0.20 ($p_\beta = 0.21$)、0.22 ($p_\beta = 0.15$)、0.37 ($p_\beta = 0.25$)，雖然皆未達顯著差異，然而，後三類教學的整體平均教學效果量皆顯示比單一類比教學效果來得佳，在考量後三類教學可比較研究數量的多寡，及這三類教學都是在原類比外，透過其他輔助途徑來促進學生概念理解，具理論概念可合併性下，本研究將類比橋、多重類比、電腦輔助類比動畫等合併為「非單一類比」，以提升可比較研究數進行分析，結果顯示 β 值為 0.22 ($p_\beta = 0.08$)，雖然同樣未達顯著水準，然其 p 值似乎愈趨顯著或近臨界顯著。

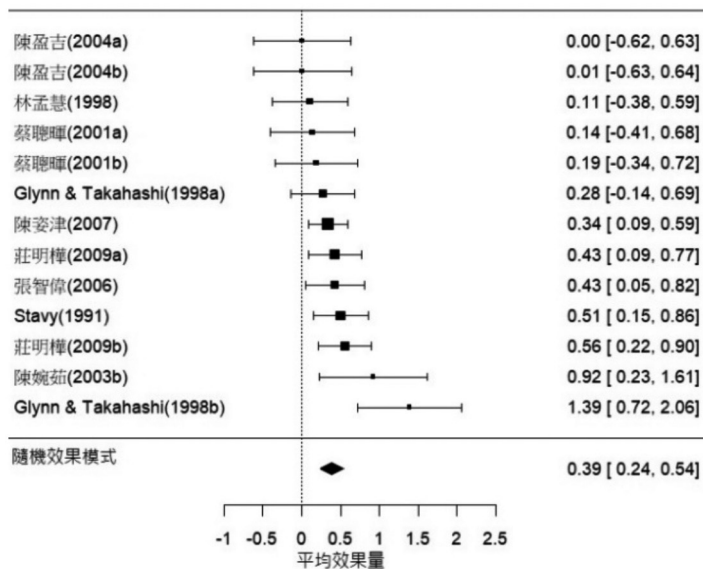
圖 1
立即效果森林圖



在參與者學習階段方面，經分析參與者為國中（含）以上學生，其 β 值為 -0.36 ($p_\beta = 0.02$)，達顯著水準，顯示類比教學對於國小階段學生的教學平均效果量為 0.88 ，而對於國中（含）以上學生的平均效果量，為 $0.88 - 0.36 = 0.52$ ，兩者是具顯著差異，而此調節變項整體可解釋量為 6.46% ，顯示類比教學對於不同學習階段學生，具有不同的教學效果，相對於國小階段，國中（含）以上學生待學習的科學概念較複雜，若類比不當，容易形成迷思，似乎可能是導致類比教學效果不佳的原因，間接突顯出類比教學存在雙面刃議題。

在尋求較適當以評估個別研究品質指標方面，本研究提出三項指標，分別為前測表現基準均等、評量工具信效度、參與者人數，以回應實驗研究品質評估效度的各面向。就前測表現基準均等指標而言，其 β 值為 0.10 ($p_\beta = 0.58$)，未達到顯著水準，顯示不論前測表現基準是否均等，該平均效果量並不具顯著差異。對此，若細部探討本研究對於效果量計算的選擇，可發現即使原始文獻研究參與者流失率過大（超過 10% 標準），或前測表現差異大於 0.5 標準差，被歸類為不符合指標標準時，只要該文獻有提供共變數分析相關統計量或前後測表現數據，本研究皆會使用共變數校正分數平均效果量，進行效果量估計與後續分析，此舉似乎已預先彌平部分兩者可能存在的差異，進而形成不顯著結果。

圖 2
延宕效果森林圖



在評估評量工具信效度指標方面，評量工具具備基本信效度者 β 值為 -0.57 ($p_\beta < .0001$)，達到顯著水準，此調節變項整體可解釋量為 44.19% ，顯示原始研究文獻若未提供或未說明評量工具品質，容易獲得較高的平均教學效果量 1.06 ，屬大效果，然而，若考量有提供且具備基本評量工具信效度者，其整體平均效果量僅為 $1.06 - 0.57 = 0.49$ ，屬中等效果。就參與者人數指標而言，具備基本參與者人數的 β 值為 -0.81 ($p_\beta = 0.003$)，同樣達到顯著水準，而此調節變項整體可解釋量為 10.71% ，顯示小型類比教學研究容易得到極大的平均效果 1.36 ，然而，一般類教學研究的整體平均效果為 $1.36 - 0.81 = 0.55$ ，屬中等效果，一如 Slavin (2008) 所表示，小型研究容易獲得極端正（或負）的效果。

同時符合三個指標方面，經檢核同時符合三項指標之研究，其 β 值為 -0.33 ($p_\beta = 0.007$)，達顯著水準，顯示具備兼顧效度各層面之個別研究品質者，其類比教學整體平均效果量為 $0.82 - 0.33 = 0.49$ ，具中等程度效果，整體解釋量為 23.87% 。整體而言，不論使用哪一個別研究品質評估指標，

本研究皆得到類似結果，不符合研究品質評估指標者之平均效果量，皆大於符合指標者，顯示出研究者若未考量個別研究品質，可能會高估類比教學效果。

表 4
個別調節變項逐一分析結果

類別	β	$SE \beta$	p_{β}	95%CI		$Q(E)/$ $p_{Q(E)}$	$Q(M)/$ $p_{Q(M)}$	I^2 (%)	R^2 (%)
				LL	UL				
出版類型									
截距 (未出版)	0.44	0.08	<.0001	0.27	0.60	194.51	6.90	78.92	8.20
已出版	0.31	0.12	0.009	0.08	0.55	<.0001	0.01		
教學取徑									
截距 (單一類比)	0.47	0.10	<.0001	0.27	0.66	192.39	3.27	80.77	---
類比橋	0.20	0.16	0.21	-0.12	0.52				
多重類比	0.22	0.15	0.15	-0.08	0.53				
電腦輔助類比動畫	0.37	0.33	0.25	-0.27	1.01				
教學取徑 – 合併									
截距 (單一類比)	0.47	0.10	<.0001	0.28	0.65	194.91	3.09	79.99	---
非單一類比	0.22	0.13	0.08	-0.03	0.47	<.0001	0.08		
參與者學習階段									
截距 (國小)	0.88	0.14	<.0001	0.61	1.16	202.18	5.20	79.72	6.46
國中 (含) 以上	-0.36	0.16	0.02	-0.66	-0.05	<.0001	0.02		
前測表現基準									
截距 (不均等)	0.51	0.17	0.003	0.17	0.84	217.73	0.30	81.17	---
前測表現均等	0.10	0.18	0.58	-0.26	0.46	<.0001	0.58		
評量工具信效度									
截距 (不符合)	1.06	0.13	<.0001	0.81	1.30	135.96	17.02	69.84	44.19
評量具備基本信效度	-0.57	0.14	<.0001	-0.84	-0.30	<.0001	<.0001		
參與者總人數									
截距 (不符合)	1.36	0.26	<.0001	0.84	1.87	197.43	9.10	79.23	10.71
具基本參與者人數	-0.81	0.27	0.003	-1.33	-0.28	<.0001	0.003		
符合三項指標									
截距 (不符合)	0.82	0.10	<.0001	0.62	1.02	170.36	7.20	75.93	23.87
符合三項研究品質評估指標	-0.33	0.12	0.007	-0.57	-0.09	<.0001	0.007		

註： β = 迴歸係數； $SE \beta$ = 迴歸係數標準誤； p_{β} = 迴歸係數顯著考驗；CI = 迴歸係數信賴區間；LL = 迴歸係數最下限；UL = 迴歸係數最上限； $Q(E)$ 、 $p_{Q(E)}$ = 組內異質性評估值及其顯著考驗； $Q(M)$ 、 $p_{Q(M)}$ = 組間異質性評估值及其顯著考驗； I^2 = Higgins 與 Thompson (2002) 異質性評估； R^2 = 整體解釋變異量。

(四) 不同研究品質下，不同類比教學與不同年齡學生科學學習成效

為進一步分析「不同個別研究品質指標對於已經存在且具影響力實務情境變項（如教學取徑、學習階段）之多元調節變項效果」，將部分變項合併，另考量調節變項間存在多元共線性，逐一納入該評估指標效果、模式變化等。說明如下：

1. 不同類比教學對於不同學習階段之調節效果

在前述表 4 個別調節變項分析中，可發現分別檢視類比教學取徑、參與者學習階段等實務變項，大致具備臨界顯著或顯著調節效果，因此，在此模式的建立，為同時投入此二變項，結果如表 5 所示，發現不同類比教學、參與者學習階段皆同時出現顯著效果，其 β 值分別是 0.27 ($p_\beta = 0.03$)、-0.41 ($p_\beta = 0.01$)，代表針對國小學習階段的學習者，給予單一類比教學成效為 $0.77 + 0*0.27 + 0*(-0.41) = 0.77$ ，具大程度教學效果，若是給予非單一類比教學的成效則為 $0.77 + 1*0.27 + 0*(-0.41) = 1.04$ ，具更大教學效果；此外，若針對國中（含）以上學習者，分別給予單一類比、非單一類比的教學效果，分別是 $0.77 + 0*0.27 + 1*(-0.41) = 0.36$ 、 $0.77 + 1*0.27 + 1*(-0.41) = 0.63$ ，大致屬中等程度上、下區間，而此二變項整體可解釋變異量達 13.90%。因此，不同類比教學對於不同學習階段學生，具有不同教學效果，且給予非單一類比教學具有較佳成效。然而，此結果是在未控制個別研究品質情況下，進行分析，接續，進一步探討納入品質評估之結果。

表 5
不同類比教學對於不同學習階段參與者於科學概念改變成效

類別	β	SE β	p_β	95%CI		Q(E)/ $p_{Q(E)}$	Q(M)/ $p_{Q(M)}$	I^2 (%)	R^2 (%)
				LL	UL				
截距	0.77	0.15	< .0001	0.48	1.06	174.85 < .0001	10.08 0.01	78.27	13.90
非單一類比	0.27	0.12	0.03	0.03	0.52				
國中（含）以上	-0.41	0.16	0.01	-0.71	-0.10				

註：符號意涵同表 4。

2. 不同個別研究品質指標對於不同類比教學、學習階段學生教學調節效果

在表 5 分析中，本研究發現不同類比教學模式與不同學生學習階段，同時具有顯著調節效果，以下進一步探討不同研究品質評估指標對於這兩調節變項可能影響。此外，由於本文比較研究數僅有 43 筆（刪除 2 筆極端值），若同時檢定三個調節變項，容易因某個水平細格研究數過少，影響統計檢定力，據此，本文分別以出版類型、前測表現基準均等、評量工具信效度、參與者人數、同時符合整體三指標等，各自針對類比教學取徑、學生學習階段分別進行分析，此舉雖然無法同時檢視各品質評估指標同時對於類比教學取徑與學生學習階段的影響，但仍能在相對穩定研究數量下，檢視各品質評估指標分別對於各實務變項情境影響。

在投入「出版類型」，分別針對類比教學取徑、與學生學習階段進行多元調節變項分析，結果如表 6、表 7，發現已出版 β 值為 0.31 ($p_\beta = 0.02$)、0.27 ($p_\beta = 0.025$)，皆達顯著，已出版研究傾向具有正向教學效果，整體解釋變異量分別為 8.13%、13.16%。此外，在同時檢視其它兩個調節變項時，可發現相對於表 4、表 5，類比教學取徑、學生學習階段變成不顯著或臨界顯著，代表有無同時控制出版類型，其對於類比教學類型、參與者學習階段，具有些許影響力。整體而言，本研究發現不同出版類型對於不同類比教學情境的教學研究結果，具有些許影響，然而，此是來自「真實研究品質」影響、抑或是「出版偏誤」影響，單純以出版類型作為評估指標，是無法有效釐清，尚待更具體指標的分析結果佐證。

表 6
不同出版類型下，不同類比教學取徑教學成效

類別	β	SE β	p_β	95%CI		Q(E)/ $p_{Q(E)}$	Q(M)/ $p_{Q(M)}$	I^2 (%)	R^2 (%)
				LL	UL				
截距	0.36	0.10	0.0005	0.16	0.57	180.29 < .0001	8.95 0.01	78.92	8.13
非單一類比	0.15	0.13	0.25	-0.10	0.40				
已出版	0.31	0.13	0.02	0.06	0.56				

註：符號意涵同表 4。

表 7
不同出版類型下，不同學生學習階段教學成效

類別	β	$SE \beta$	p_{β}	95%CI		$Q(E)/$ $p_{Q(E)}$	$Q(M)/$ $p_{Q(M)}$	F^2 (%)	R^2 (%)
				LL	UL				
截距	0.69	0.16	<.0001	0.37	1.01	181.89	10.40	78.01	13.16
國中(含)以上	-0.28	0.16	0.07	-0.59	0.02	<.0001	0.0055		
已出版	0.27	0.12	0.025	0.03	0.50				

註：符號意涵同表 4。

在投入「前測表現基準均等」指標，分別針對類比教學取徑、與學生學習階段進行多元調節變項分析，其結果分別如表 8、表 9 所示，發現前測表現基準具可比性之 β 值分別為 0.07 ($p_{\beta} = 0.72$)、0.03 ($p_{\beta} = 0.89$)，皆不具顯著水準，同時，檢視類比教學取徑、學生學習階段等調節變項分析結果，學生學習階段仍如同表 4 之結果，大致回應出本研究前面所述，在效果量選擇上，若研究者一開始即針對原始研究文獻的研究設計與所提供訊息，選擇適當的效果量計算量尺，能適當確保或調整各研究前測表現基準的可比較性。

表 8
是否具前測可比較性下，不同類比教學取徑教學成效

類別	β	$SE \beta$	p_{β}	95%CI		$Q(E)/$ $p_{Q(E)}$	$Q(M)/$ $p_{Q(M)}$	F^2 (%)	R^2 (%)
				LL	UL				
截距	0.40	0.21	0.06	-0.02	0.81				
非單一類比	0.23	0.13	0.08	-0.03	0.49	194.88	3.15	81.22	--
前測表現基準具可比性	0.07	0.20	0.72	-0.32	0.47	<.0001	0.21		

註：符號意涵同表 4。

表 9
是否具前測可比較性下，不同學生學習階段教學成效

類別	β	$SE \beta$	p_{β}	95%CI		$Q(E)/$ $p_{Q(E)}$	$Q(M)/$ $p_{Q(M)}$	F^2 (%)	R^2 (%)
				LL	UL				
截距	0.86	0.23	0.0002	0.41	1.31				
國中(含)以上	-0.35	0.16	0.03	-0.67	-0.04	202.16	5.13	80.21	3.05
前測表現基準具可比性	0.03	0.18	0.89	-0.33	0.38	<.0001	0.08		

註：符號意涵同表 4。

在投入「評量工具信效度」進行分析，結果分別如表 10、表 11，評量工具備基本信效度者 β 值分別為 -0.53 ($p_{\beta} = 0.0002$)、-0.51 ($p_{\beta} = 0.0008$)，皆達顯著水準，然而，與表 4、表 5 相比，類比教學取徑、學生學習階段等調節變項分析結果，由(臨界)顯著變成不顯著結果，代表有無同時控制評量工具信效度是否具基本水準，其對於採用何種類比教學，或針對哪一學習階段學生進行教學，具有相當影響，整體解釋變異量分別為 43.69%、43.10%。整體而言，評量工具信效度對於不同類比教學情境的教學研究結果具有一定程度影響，若研究者納入不具備或未說明評量工具基本水準者，可能會錯估其它調節變項的影響或高估整體教學效果。

表 10
評量是否具基本信效度下，不同類比教學取徑教學成效

類別	β	SE β	p_{β}	95%CI		Q(E)/ $p_{Q(E)}$	Q(M)/ $p_{Q(M)}$	I^2 (%)	R^2 (%)
				LL	UL				
截距	0.95	0.15	<.0001	0.65	1.25	126.59 <.0001	18.16 0.0001	69.98	43.69
非單一類比	0.14	0.11	0.20	-0.07	0.35				
評量具基本信效度	-0.53	0.14	0.0002	-0.81	-0.25				

註：符號意涵同表 4。

表 11
評量是否具基本信效度下，不同學生學習階段教學成效

類別	β	SE β	p_{β}	95%CI		Q(E)/ $p_{Q(E)}$	Q(M)/ $p_{Q(M)}$	I^2 (%)	R^2 (%)
				LL	UL				
截距	1.13	0.15	<.0001	0.84	1.41	134.55 <.0001	17.76 0.0001	70.27	43.10
國中(含)以上	-0.14	0.15	0.33	-0.43	0.14				
評量具基本信效度	-0.51	0.15	0.0008	-0.80	-0.21				

註：符號意涵同表 4。

在投入「參與者人數」指標，分析結果分別如表 12、表 13 所示，實驗研究參與者具備基本人數者 β 值分別為 -0.77 ($p_{\beta} = 0.004$)、-0.68 ($p_{\beta} = 0.02$)，達到顯著水準，相較於表 4、表 5，類比教學取徑、學生學習階段等調節效果，皆變成不顯著結果，代表有無同時控制參與者是否具備基本人數，其對於何種類比教學、或學習階段學生進行教學，具有相當影響，整體解釋變異量分別為 14.37%、12.99%。故可得知參與者基本人數對於不同類比教學情境的教學研究結果具有一定程度影響，若研究者納入小型研究，可能會錯估其它調節變項影響或高估整體教學效果，而此論點雷同 Slavin (2008) 之建議。

表 12
是否具基本參與者總人數下，不同類比教學取徑教學成效

類別	β	SE β	p_{β}	95%CI		Q(E)/ $p_{Q(E)}$	Q(M)/ $p_{Q(M)}$	I^2 (%)	R^2 (%)
				LL	UL				
截距	1.21	0.28	<.0001	0.67	1.75	176.90 <.0001	11.41 0.003	78.52	14.37
非單一類比	0.19	0.12	0.13	-0.05	0.43				
具基本參與者人數	-0.77	0.27	0.004	-1.29	-0.24				

註：符號意涵同表 4。

表 13
是否具基本參與者總人數下，不同學生學習階段教學成效

類別	β	SE β	p_{β}	95%CI		Q(E)/ $p_{Q(E)}$	Q(M)/ $p_{Q(M)}$	I^2 (%)	R^2 (%)
				LL	UL				
截距	1.41	0.26	<.0001	0.90	1.93	189.08 <.0001	11.03 0.004	78.84	12.99
國中(含)以上	-0.22	0.16	0.18	-0.54	0.10				
具基本參與者人數	-0.68	0.28	0.02	-1.23	-0.12				

註：符號意涵同表 4。

在前述研究品質評估指標分析中，研究者逐一探討各指標影響力，然而，部分指標內水平存在比較研究數量偏少情況，例如，如表 2，不具備參與者基本人數者僅 4 筆、不具備評量工具信效度基本水準者僅 8 筆，偏少的研究數量或多或少會影響多元調節變項分析結果穩定性，故本研究在考量能同時兼顧不同層面效度、與提供適當研究數量下，進行「是否同時符合三指標」之多元調節變項分析。

在投入「同時符合研究品質評估三指標」方面，結果分別如表 14、表 15 所示，同時符合個別研究品質評估三指標者， β 值分別為 -0.36 ($p_{\beta} = 0.005$)、 -0.28 ($p_{\beta} = 0.02$)，皆達顯著水準，經與表 4、表 5 的類比教學取徑、學生學習階段調節效果相比，由顯著變成不顯著或臨界顯著的結果，表示有無同時控制整體個別研究品質是否具備基本水準，其對類比教學、或學生學習階段具有相當影響力，整體解釋變異量為 31.74%、26.44%。整體而言，評估整體個別研究品質對於不同類比教學情境的教學研究結果具一定程度的影響，若研究者納入不具備基本整體研究品質者，可能會錯估其它調節變項影響或高估整體教學效果。

表 14
整體研究品質是否符合標準下，不同類比教學取徑教學成效

類別	β	$SE \beta$	p_{β}	95%CI		$Q(E)/$	$Q(M)/$	F^2 (%)	R^2 (%)
				LL	UL	$p_{Q(E)}$	$p_{Q(M)}$		
截距	0.77	0.14	<.0001	0.50	1.04				
非單一類比	0.14	0.12	0.21	-0.08	0.37	145.15	11.84	73.82	31.74
同時符合品質 評估三指標	-0.36	0.13	0.005	-0.61	-0.11	<.0001	0.003		

註：符號意涵同表 4。

表 15
整體研究品質是否符合標準下，不同學生學習階段教學成效

類別	β	$SE \beta$	p_{β}	95%CI		$Q(E)/$	$Q(M)/$	F^2 (%)	R^2 (%)
				LL	UL	$p_{Q(E)}$	$p_{Q(M)}$		
截距	1.02	0.15	<.0001	0.73	1.30				
國中(含)以上	-0.28	0.15	0.06	-0.57	0.01	162.28	10.98	75.35	26.44
同時符合品質 評估三指標	-0.28	0.12	0.02	-0.53	-0.04	<.0001	0.004		

註：符號意涵同表 4。

(五) 出版偏誤與個別研究品質評估

在出版偏誤影響方面，誠如前述表 4、表 6 與表 7 分析結果，皆具類似 Lipsey 與 Wilson (1993) 所發現已出版研究具備較高效果量現象，若以其它學者所建議統計分析技術進行檢核，預先以各研究效果量分佈的漏斗圖 (funnel plot) 進行檢視，繪製立即效果量漏斗圖分佈 (圖 3)、延宕效果量漏斗圖分佈 (圖 4)，以視覺檢核，立即效果左下角有些微不對稱性，經 Egger 等人 (1997) 線型迴歸考驗其不對稱性 (asymmetry)，在未置入任何調節變項時，立即效果不對稱性顯著考驗為 $z = 2.09$ ($p = 0.04$)，達到顯著結果，而延宕效果不對稱性檢核為 $z = 0.29$ ($p = 0.77 > 0.05$)，未達顯著結果，整體而言，立即效果存在輕微的不對稱，顯示部分「未達顯著效果」的研究可能未被發表或蒐集。

然而，一如 Vevea 等人 (2019) 認為以統計方法針對出版偏誤進行檢定，若出現顯著結果，往往也可能是伴隨其它調節變項所產生異質性影響，例如，本研究所關注個別研究品質差異，據此，

本研究進一步檢視在納入調節變項（自表 5 至表 15），再次以線型迴歸分析進行不對稱性檢驗，此時立即效果不對稱性顯著考驗依表的順序，分別為 $z = 1.14$ ($p = 0.26$)、 $z = 1.64$ ($p = 0.10$)、 $z = 1.35$ ($p = 0.18$)、 $z = 1.94$ ($p = 0.05$)、 $z = 1.47$ ($p = 0.14$)、 $z = 1.63$ ($p = 0.10$)、 $z = 1.53$ ($p = 0.13$)、 $z = 0.12$ ($p = 0.91$)、 $z = 0.02$ ($p = 0.98$)、 $z = 1.39$ ($p = 0.17$)、 $z = 1.12$ ($p = 0.26$)，各模式皆出現未達顯著結果。因此，本文認為此不對稱性乃源自於研究中其它調節變項異質性影響，若同時考量時，能有效避免其影響。

圖 3
立即效果量漏斗圖分佈

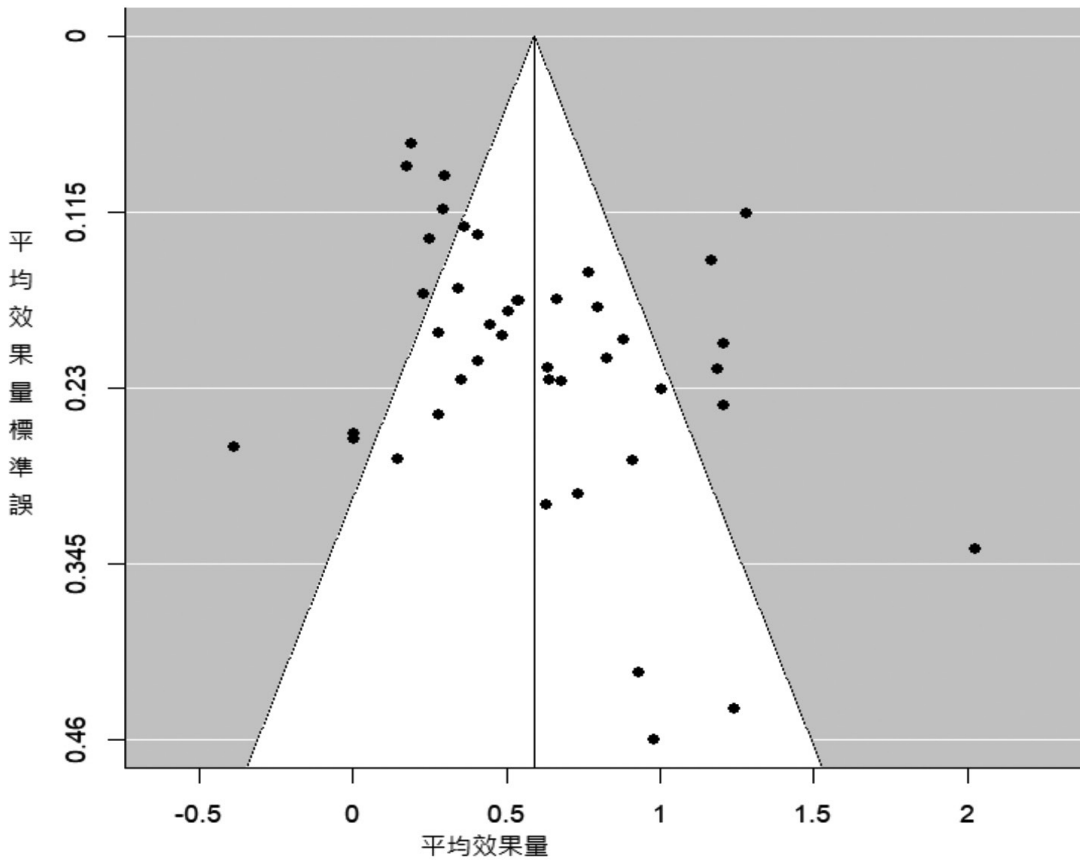
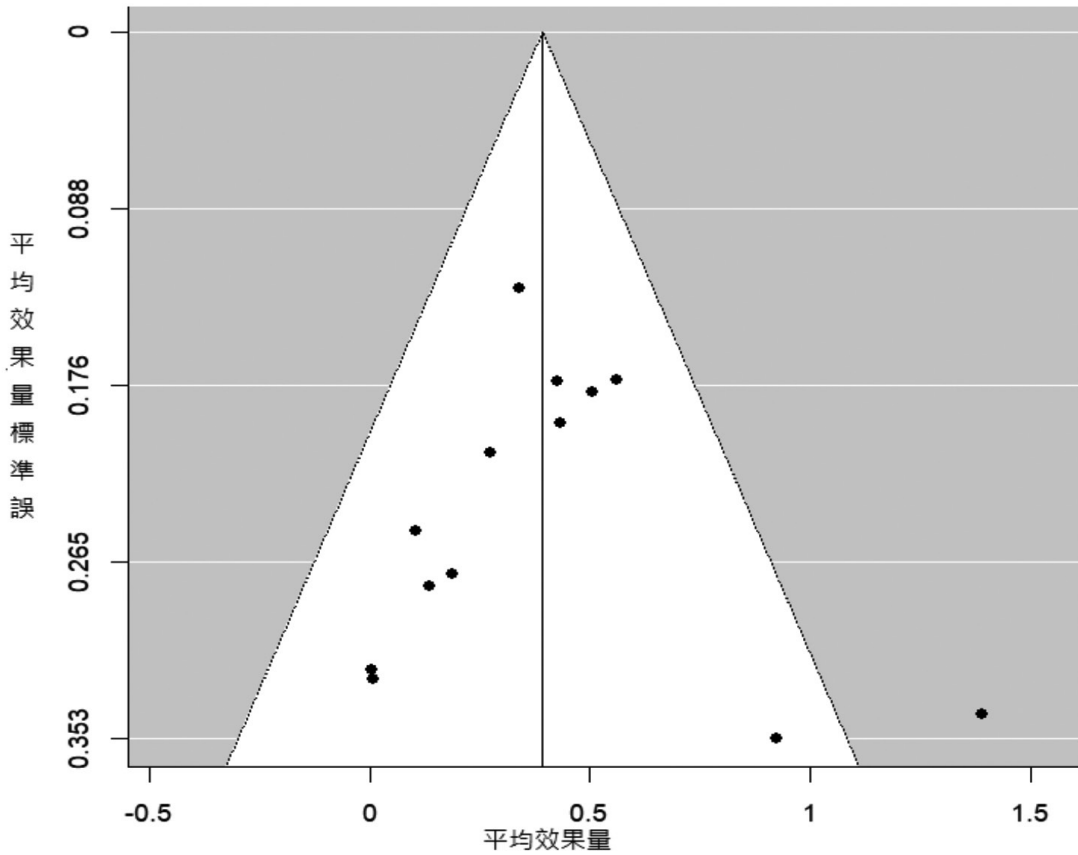


圖 4
延宕效果量漏斗圖分佈



結論與建議

(一) 類比教學普遍具有正向教學效果但隨調節變項不同而差異

在探討類比教學統合成效上，本研究發現整體具中等程度教學成效，然而，此點會受不同調節變項影響而具差異性，其中，單以教學取徑而言，進行非單一類比教學，包含多重類比、類比橋、電腦輔助類比動畫等，似乎傾向具有較大教學效果；單以參與者學習階段而言，類比教學對於國小學生具有較佳教學效果，而國中（含）以上階段，可能如同 Keane（1990）論述，國中以上科學概念較為複雜，易超出學習者工作記憶負荷，造成學習困難或類比錯誤，可能致使類比教學效果相對於國小階段來得小。然而，在進一步同時考量個別研究品質，本研究以評量工具是否具備基本信效度、前測表現基準是否均等、是否具備基本總人數、與同時符合三項個別研究品質指標等進行分析，雖然分析結果依舊發現整體類比教學是具有一定程度正向教學效果，然而，前述不同類比教學取徑、參與者學習階段的調節影響，多會由（臨界）顯著變成不顯著，顯示出是否控制個別研究品質，對於這些實務調節變項，具有相當影響力。整體而言，不論是從研究品質較佳的統合結果、抑或是研究品質較差的統合結果，整體類比教學皆具有一定正向教學成效，建議教師可多使用以協助學生概念連結，以達到學習遷移效果，然而，是否得根據不同學生學習階段、搭配不同類比教學取徑，才

得選擇出最佳的教學成效，由於容易受到個別研究品質優劣影響，還有賴未來納入更多研究進行分析。

（二）不同個別研究品質評估指標選擇會影響量化研究統合所形成證據有效性

在考量不同個別研究品質評估指標，以作為類比教學效果統合時，本研究發現在量化研究統合脈絡下，使用不同評估指標，如：出版類型、前測表現基準均等、評量工具信效度、參與者人數等，皆有其待回應或考量的因素，包含有效度理論層面、出版偏誤議題、實徵結果與實際原研究文獻資料可得性等。

研究者若單以出版類型作為品質評估指標，本研究分析結果發現其雖然具有調節效果，但容易同時與過去學者所提「出版偏誤」影響混淆，無法具體釐清該效果是受個別研究品質差異影響，抑或出版偏誤影響；此外，研究者若單以前測表現基準均等作為品質評估標的，就本文調節變項分析結果，似乎不具實徵影響力，但細究時，可發現研究者若一開始即針對原始研究文獻研究設計與所提供訊息，選擇適當的效果量計算量尺，即能確保或調整各研究前測表現基準的可比較性；若研究者單獨僅使用「評量工具信效度或參與者基本人數」作為研究品質評估指標，除符合效度理論外，其結果也顯示具有實徵影響力，但缺點在於原始文獻可獲得比較研究數量較少（如，多數研究會使用具信效度評量工具，較少獲得未使用或未說明之比較研究），致使分析穩定性較容易受到挑戰。整體而言，在同時考量能兼顧效度理論各層面、實徵影響力、資料可得性等條件下，本研究較建議研究者使用「同時前測表現基準均等、評量工具信效度與參與者基本人數」三個指標皆符合者，作為個別研究品質評估，不僅能適度避免出版偏誤影響，同時能全面回應品質評估議題，以增進統合結果所形成證據的有效性。

（三）不同主題領域，其替代性研究品質評估指標具差異性

本研究聚焦於探討類比教學對於學生科學概念學習成效，主題範疇屬科學教育，其中一項針對評量工具所使用品質檢核指標為評估其信效度，類似數位學習領域（李漢岳等人，2017）會使用以評估者，然而，就其它主題領域而言，例如，閱讀教學發現使用評量工具來源（研究者自編、或使用既有評量工具）作為檢核指標，會對統合結果產生影響，若使用研究者自編評量工具傾向獲得相對較高平均效果量（謝進昌，2019），此現象一如 Slavin 與 Madden（2011）所表示，研究者若未教授對照組學生相對應待評量的內容，容易產生內容教學不均衡的偏誤。然而，此點就本研究分析結果，可發現使用「既有評量工具」進行類比教學研究者，僅有 1 筆，即使該因素存在可能影響，受限於原始研究文獻資料可得性，無法提供相應實徵分析結果，代表在類比教學主題，以評量工具的來源可能不具備實際可作為品質評估的可行性。整體而言，在量化統合研究脈絡下，除效度理論、實徵結果外，研究者還得考量不同主題領域屬性所導致實際原始文獻資訊可得性，建議未來研究者可試著針對不同主題領域，探討各自適合的個別研究品質評估指標。

（四）可比較研究數量為研究限制

統合研究是一份涉及有系統性文獻回顧、資料編碼、驗證與分析的嚴謹過程，而在此程序中，往往得花上相當時間，準備與確認資料，若期間遇到不可抗力因素（如他方編碼者離職），就容易延宕研究推展，據此，受限於時間、人力與成員更迭因素，以致本研究僅能蒐集 1976–2010 年（35 年間）國內外類比教學實徵性文獻，獲得 45 筆比較研究數，而根據過往學者建議，在進行單一調節變項分析時，此數量仍可具備相當分析穩定性，但當隨著調節變項數量變多時，其穩定性則容易因各水平分佈數量不一，受到挑戰，致使本研究對於多元調節變項分析，也得降低同時納入分析之變項數，來回應研究待答問題。因此，本文後續預計進行編碼資料釋出，除透過資料透明化，以確保研究品質外，更為有助與建議後續研究者進行更新分析，以瞭解近年類比教學對於學生科學概念學習影響變化，並透過增加研究文獻數量，強化與增加同時投入多個調節變項分析結果穩定性。

參考文獻

- * 表示納入本文後設分析之研究文獻。
- * 王瑞文 (2003) : 《探究多重類比教學策略對國二學生的影響—以比熱概念為例》(未發表碩士論文), 國立高雄師範大學。[Wang, R.-W. (2003). *The influence of multiple analogies instructional strategies in specific heat conception for student the eighth grade* (Unpublished master's thesis). National Kaohsiung Normal University.]
- * 王榮達 (2005) : 《具體表徵在理化教學上的應用》(未發表碩士論文), 國立高雄師範大學。[Wang, J.-T. (2005). *The use of concrete instructional representation strategies in teaching particle concepts* (Unpublished master's thesis). National Kaohsiung Normal University.]
- * 何振宏 (2010) : 《結合 5E 探究及類比橋教學模組於八年級學生浮力概念改變成效之研究》(未發表碩士論文), 國立彰化師範大學。[He, Z.-H. (2010). *Research on effects of combining 5E inquiry learning cycle and analog bridge teaching modules on the conceptual changes in buoyancy for the eighth graders* (Unpublished master's thesis). National Changhua University.]
- 李漢岳、楊介銘、宋曜廷 (2017) : 〈數位學習實驗研究品質評估與現況分析：以行動學習為例〉。《教育科學研究期刊》, 62 (2) , 31-60。[Lee, H.-Y., Yang, J.-M., & Sung, Y.-T. (2017). Quality assessment and situational analysis of experimental elearning designs: A case study of mobile learning. *Journal of Research in Education Sciences*, 62(2), 31–60.] [https://doi.org/10.6209/JORIES.2017.62\(2\).02](https://doi.org/10.6209/JORIES.2017.62(2).02)
- * 林孟慧 (1998) : 《理化類比對國三學生地球科學概念學習之影響》(未發表碩士論文), 國立臺灣師範大學。[Lin, M.-H. (1998). *The influence of school-knowledge analogy on junior high school students' learning of earth science concepts* (Unpublished master's thesis). National Taiwan Normal University.]
- * 邱美虹、高淑芬 (1999) : 〈類比對應對學生建構 " 原子結構 " 心智表徵之影響〉。《師大學報》, 44 (1) , 31–59。[Chiu, M.-H., & Gau, S.-F. (1999). The effectiveness of analogies in learning of concepts of structure of atoms by eighth graders. *Journal of National Taiwan Normal University*, 44(1), 31–59.] <https://doi.org/10.6300/JNTNU.1999.44.03>
- * 林建隆 (2000) : 《合作學習的類比學習環對國中學生壓力概念學習成效之研究》(未發表博士論文), 國立彰化師範大學。[Lin, J.-L. (2000). *The study for analysis of cooperative - analogical learning cycle on the conceptual change of hydrostatic pressure at junior high level* (Unpublished doctoral dissertation). National Changhua University.]
- * 林靜雯 (2000) : 《由概念改變及心智模式初探多重類比對國小四年級學生電學概念學習之影響》(未發表碩士論文), 國立臺灣師範大學。[Lin, J.-W. (2000). *The study of the multiple analogies on the electricity learning through conceptual change and mental model for fourth grade students* (Unpublished master's thesis). National Taiwan Normal University.]
- * 徐星豪 (2004) : 《應用類比學習環探討國中學生大氣壓力概念學習成效的研究》(未發表碩士論文), 國立彰化師範大學。[Xu, X.-H. (2004). *The study of analogical learning cycle to explore*

- the effectiveness of concepts on atmospheric pressure at junior high school* (Unpublished master's thesis). National Changhua University.]
- * 郭人仲 (1994) : 《國中學生生物概念類比學習之研究》(未發表碩士論文), 國立彰化師範大學。 [Guo, R.-Z. (1994). *The study of analogical learning of concepts in junior high biology* (Unpublished master's thesis). National Changhua University.]
 - * 陳可恭 (2005) : 《從系統典範探討板塊構造學說多重類比教學 – 「凱利方格法」(RGT)之系統性應用》(未發表博士論文), 國立臺灣師範大學。 [Lin, M.-H. (2005). *A study on teaching plate tectonics with multiple analogies from the perspective of system paradigm: An application of repertory grid technique* (Unpublished doctoral dissertation). National Taiwan Normal University.]
 - * 莊明樺 (2009) : 《探討純粹比對類比與傳達屬性類比網路課程對國小學生科學概念建構與類比推理能力之影響》(未發表碩士論文), 國立交通大學。 [Chuang, M.-H. (2009). *Explore the impact of pure matching and carry over web-based learning content on 5th grade students' scientific concept construction and analogical reasoning ability* (Unpublished master's thesis). National Chiao Tung University.]
 - * 陳盈吉 (2004) : 《探究動態類比對於科學概念學習與概念改變歷程之研究 – 以國二學生學習氣體粒子概念為例》(未發表碩士論文), 國立臺灣師範大學。 [Chen, I.-J. (2004). *To inquiry the processes of students' learning and conceptual change when they use dynamic analogy for learning--An example of 8th students to learn the concepts about gas particles' movement* (Unpublished master's thesis). National Taiwan Normal University.]
 - * 陳恆迪 (1993) : 《國中學生物理概念類比學習之研究》(未發表碩士論文), 國立彰化師範大學。 [Chen, H.-D. (1993). *The study of analogical learning on physics conceptions for junior high students* (Unpublished master's thesis). National Changhua University.]
 - * 陳姿津 (2007) : 《『科學類比推理』網路互動學習研究 – 促進國中生電學概念之建構與推理能力》(未發表碩士論文), 國立交通大學。 [Chen, T.-C. (2007). *Research of "scientific analogical reasoning" within web-based interactive learning: Promoting middle school students' reasoning abilities and concept construction involving electricity* (Unpublished master's thesis). National Chiao Tung University.]
 - * 陳炳憲 (2003) : 《以類比學習環、ARCS 動機策略改進國二學生化學式學習之行動研究》(未發表碩士論文), 國立彰化師範大學。 [Chen, B.-S. (2003). *The action research of analogical learning cycle and ARCS motivational strategies to improve chemical formula learning on eighth grade students* (Unpublished master's thesis). National Changhua University.]
 - * 陳婉茹 (2003) : 《探討動態類比對於化學平衡概念學習之研究 – 八年級學生概念本體及心智模式之變化》(未發表碩士論文), 國立臺灣師範大學。 [Chen, J.-W. (2003). *The study of dynamic-analogy in chemical equilibrium concepts learning: The changes between mental model and conceptual ontology for eighth grade students* (Unpublished master's thesis). National Taiwan Normal University.]
 - * 張智偉 (2006) : 《結合 Edison 電腦模擬的類比學習環融入國中直流電路學習成效之研究》(未

- 發表碩士論文)，國立彰化師範大學。[Chang, C.-W. (2006). *The effect of direct circuit learning integrated cycle of analogical-learning with Edison computer simulation* (Unpublished master's thesis). National Changhua University.]
- * 許瑀庭 (2006)：《以類比學習環融入教學克服國中生電學迷思概念之研究》(未發表碩士論文)，國立彰化師範大學。[Xu, Y.-T. (2006). *The study of analogical learning cycle to improve the electricity misconceptions at junior high school* (Unpublished master's thesis). National Changhua University.]
- 黃秀玉 (2009)：《韻尾類比訓練對國小六年級學生英文讀字能力之成效研究》(未發表碩士論文)，國立政治大學。[Huang, S.-Y. (2009). *The effects of rime analogy training on word reading for 6th graders* (Unpublished master's thesis). National Chengchi University.]
- 謝進昌 (2019)：〈促進中文閱讀理解教學成效量化研究統合：調節變項影響與評估〉，《教育科學研究期刊》，64(4)，175–206。[Hsieh, J.-C. (2019). Synthesis of quantitative research on Chinese reading comprehension instruction: Analysis of moderating effects. *Journal of Research in Education Sciences*, 64(4), 175–206.] [https://doi.org/10.6209/JORIES.201912_64\(4\).0007](https://doi.org/10.6209/JORIES.201912_64(4).0007)
- * 鄭文彥 (2007)：《以類比橋融入浮力單元對國中生迷思概念改變之研究》(未發表碩士論文)，國立彰化師範大學。[Zheng, W.-Y. (2007). *The study of integrating analogical bridge into teaching buoyancy for overcoming students' misconceptions at junior high level* (Unpublished master's thesis). National Changhua University.]
- * 蔡秉恆、黃天佑 (2005)：〈促進國小學生科學直覺改變 – 運用網頁類比教材及視訊教學成效之研究〉。《資訊科學應用期刊》，1(1)，69–89。[Tsai, P.-H., & Huang, T.-Y. (2005). Facilitating changes of intuition in science in elementary school: Applying web analogy materials and video-conference. *Journal of Computer Science and Application*, 1(1), 69–89.]
- * 蔡聰暉 (2001)：《由心智模式探討學生導引之類比教學對國二學生溫度與熱概念學習之研究》(未發表碩士論文)，國立臺灣師範大學。[Tsai, T.-H. (2001). *Use mental model to explore student-guided analogical teaching how to affect the concept heat and temperature* (Unpublished master's thesis). National Taiwan Normal University.]
- * 蕭碧茹 (1996)：《圖形類比融入國中理化教學之研究》(未發表碩士論文)，國立高雄師範大學。[Xiao, B.-R. (1996). *The effectiveness of teaching science with pictorial analogies* (Unpublished master's thesis). National Kaohsiung Normal University.]
- * Baker, W., & Lawson, A. E. (2001). Complex instructional analogies and theoretical concept acquisition in college genetics. *Science Education*, 85(6), 665–683. <https://doi.org/10.1002/sce.1031>
- Brown, D. E., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: Abstract transfer versus explanatory model construction. *Instructional Science*, 18, 237–261.
- Centner, D., & Toupin, C. (1986). Systematicity and surface similarity in the development of analogy. *Cognitive Science*, 10, 277–300. https://doi.org/10.1207/s15516709cog1003_2
- Clement, J. (1987). Overcoming students' misconceptions in physics: The role of anchoring intuitions and analogical validity. In J. D. Novak (Ed.), *Proceedings of the international seminar*, Vol. 3.

- Misconceptions and educational strategies in science and mathematics* (pp. 84–97). Cornell University.
- Clement, J. (1989). The concept of variation and misconception in Cartesian graphing. *Focus on Learning Problems in Mathematics*, 11, 77–87.
- *Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching*, 30(10), 1241–1257. <https://doi.org/10.1002/tea.3660301007>
- *Clement, J. (1998). Expert novice similarities and instruction using analogies. *International Journal of Science Education*, 20(10), 1271–1286. <https://doi.org/10.1080/0950069980201007>
- Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage.
- Dagher, Z. R. (1995). Review of studies on the effectiveness of instructional analogies in science education. *Science Education*, 79(3), 295–312. <https://doi.org/10.1002/sce.3730790305>
- Duit, R. (1991). On the role of analogies and metaphors in learning science. *Science Education*, 75(6), 649–672.
- Egger, M., Davey, S. G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- *Gabel, D. L., & Sherwood, R. D. (1980). Effect of using analogies on chemistry achievement according to Piagetian level. *Science Education*, 64, 709–716. <https://doi.org/10.1002/sce.3730640516>
- Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). Cambridge University Press.
- Glynn, S. M. (1991). Explaining science concepts. In S. M. Glynn, R. H. Yeany, & B. K. Britton (Eds.), *The psychology of learning science* (pp. 219–240). Lawrence Erlbaum Associates. [https://doi.org/10.1002/\(SICI\)1098-2736\(199812\)35:10<1129::AID-TEA5>3.0.co;2-2](https://doi.org/10.1002/(SICI)1098-2736(199812)35:10<1129::AID-TEA5>3.0.co;2-2)
- Glynn, S. M., Britton, B. K., Semrud-Clikeman, M., & Muth, K. D. (1989). Analogical reasoning and problem solving in science textbooks. In J. Glover, R. Ronning, & C. Reynolds (Eds.), *Handbook of creativity: Assessment, theory, and research* (pp. 383–398). Plenum.
- *Glynn, S. M., & Takahashi, T. (1998). Learning from analogy-enhanced science text. *Journal of Research in Science Teaching*, 35(10), 1129–1149.
- *Gokhale, A. A. (1996). Using analogies to overcome misconceptions: A technology course example. *Journal of Technology Studies*, 22(1), 10–14.
- Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting conceptual change in science: A comparative meta-analysis of instructional interventions from reading education and science education. *Reading Research Quarterly*, 28(2), 116–159. <https://doi.org/10.2307/747886>
- Hall, B. W., Ward, A. W., & Comer, C. B. (1988). Published educational research: An empirical study of its

- quality. *The Journal of Educational Research*, 81(3), 182–189. <https://doi.org/10.1080/00220671.1988.10885820>
- Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, 92(2), 490–499. <https://doi.org/10.1037/0033-2909.92.2.490>
- Higgins, J. P. T., & Thomas, J. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. <https://doi.org/10.1002/sim.1186>
- Keane, M. T. (1990). Incremental analogizing: Theory and model. In K. J. Gilhooly, M. T. Keane, R. Logie, & G. Erdos (Eds.), *Lines of thinking: Reflections on the psychology of thought* (vol. 1, pp. 221–235). Wiley.
- *Lawson, A. E., Baker, W. P., Didonato, L., & Verdi, M. P. (1993). The role of hypothetico-deductive reasoning and physical analogues of molecular interactions in conceptual change. *Journal of Research in Science Teaching*, 30, 1073–1085. <https://doi.org/10.1002/tea.3660300906>
- *Lin, H.-S., Shiau, B.-R., & Lawrenz, F. (1996). The effectiveness of teaching science with pictorial analogies. *Research in Science Education*, 26(4), 495–511.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48(12), 1181–1209. <https://doi.org/10.1037/0003-066X.48.12.1181>
- Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-group designs. *Psychological Methods*, 7(1), 105–125. <https://doi.org/10.1037/1082-989X.7.1.105>
- Nyberg, L., Nilsson, L. G., Olofsson, U., & Backman, L. (1997). Effects of division of attention during encoding and retrieval on age differences in episodic memory. *Experiment Aging Research*, 23, 137–143. <https://doi.org/10.1080/03610739708254029>
- *Okebukola, F. O., & Owolabi, T. (2008). Teaching with analogies: The meeting point between science and language. *The International Journal of Learning*, 15(4), 147–150.
- Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher*, 42(8), 424–432. <https://doi.org/10.3102/0013189X13507104>
- Raudenbush, S. W. (2009). Analyzing effect sizes: Random effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 295–315). Russell Sage Foundation.
- *Rigney, J. W., & Lutz, K. A. (1976). Effect of graphic analogies of concepts in chemistry on learning and attitudes. *Journal of Educational Psychology*, 68(3), 305–311. <https://doi.org/10.1037/0022-0663.68.3.305>
- Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and constructing concept maps: A meta-analysis. *Educational Psychology Review*, 30, 431–455. <https://doi.org/10.1007/s10648-017-9403-9>

- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth, Cengage Learning.
- Slavin, R. E. (2008). Perspectives on evidence-based research in education: What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37(1), 5–14. <https://doi.org/10.3102/0013189X08314117>
- Slavin, R. E., & Lake, C. (2008). Effective programs for elementary mathematics: A best evidence synthesis. *Review of Educational Research*, 78(3), 427–515. <https://doi.org/10.3102/0034564308317473>
- Slavin, R. E., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, 4, 370–380. <https://doi.org/10.1080/19345747.2011.558986>
- Spiro, R. J., Feltovich, P. J., Coulson, R. L., & Anderson, D. K. (1989). Multiple analogies for complex concepts: Antidotes for analogy induced misconception in advanced knowledge acquisition. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 498–531). Cambridge University Press.
- *Stavy, R. (1991). Using analogy to overcome misconceptions about conservation of matter. *Journal of Research in Science Teaching*, 28, 305–313. <https://doi.org/10.1002/tea.3660280404>
- Sung Y.-T., Chang K.-E., & Liu T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, 94, 252–275. <https://doi.org/10.1016/j.compedu.2015.11.008>
- Thagard, P. (1992). Analogy, explanation, and education. *Journal of Research in Science Teaching*, 29(6), 537–544. <https://doi.org/10.1002/tea.3660290603>
- Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21(11), 1559–1573. <https://doi.org/10.1002/sim.1187>
- *Türk, F., Ayas, A., & Karslı, F. (2010). Effectiveness of analogy technique on students' achievement in general chemistry laboratory. *Procedia-Social and Behavioral Sciences*, 2(2), 2717–2721. <https://doi.org/10.1016/j.sbspro.2010.03.402>
- Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The study design and implementation assessment device (Study DIAD). *Psychological Methods*, 13, 130–149. <https://doi.org/10.1037/1082-989X.13.2.130>
- Vevea, J. L., Coburn, K., & Sutton, A. (2019). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 383–429). Russell Sage Foundation.
- Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1, 112–125. <https://doi.org/10.1002/jrsm.11>
- What Works Clearinghouse. (2020a). *WWC standards handbook* (Version 4.1). Institute of Education Sciences, Department of Education. <https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC->

[Standards-Handbook-v4-1-508.pdf](#)

What Works Clearinghouse. (2020b). *WWC Procedures handbook* (Version 4.1). Institute of Education Sciences, Department of Education. <https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Procedures-Handbook-v4-1-508.pdf>

*Yilmaz, S., Eryilmaz, A., & Geban, O. (2006). Assessing the impact of bridging analogies in mechanics. *School Science and Mathematics, 106*(6), 220–230. <https://doi.org/10.1111/j.1949-8594.2006.tb17911.x>

*Zheng, R. Z., Yang, W., Garcia, D., & McCadden, E. P. (2008). Effects of multimedia and schema induced analogical reasoning on science learning. *Journal of Computer Assisted Learning, 24*(6), 474–482. <https://doi.org/10.1111/j.1365-2729.2008.00282.x>

收稿日期：2020年08月19日

一稿修訂日期：2020年10月30日

二稿修訂日期：2020年11月25日

三稿修訂日期：2020年12月16日

接受刊登日期：2020年12月16日

Bulletin of Educational Psychology, 2021, 52(4), 935–962
National Taiwan Normal University, Taipei, Taiwan, R. O. C.

Does Assessing the Quality of Primary Research Matter? A Meta-Analysis of Analogical Instruction-Induced Changes in Students' Science Concepts

Jin-Chang Hsieh

Research Center for Testing and Assessment
National Academy for Educational Research

Yi-Ching Chiu

Department of Education Psychology
and Counseling
National Taiwan Normal University

As evidence-based education evolves, an increasing body of evidence has highlighted the issue of constructing high-quality indicators in the synthesis of quantitative research. However, what is considered optimal remains unclear and a topic of debate. Numerous researchers assess the single indicator of primary research such as the publication, pretest performance, reliability and validity, and sample size. This approach can result in bias. Therefore, this study proposed alternative quality indicators to replace the single indicator with suitable evidence synthesis. Research on analogy instruction was integrated through a systematic review and meta-analysis, and the effects of analogy instruction were explored through quantitative research synthesis.

With respect to literature reviews, document retrieval targets journal articles, master's theses, and doctoral dissertations on the basis of references to relevant retrospective studies. Regarding analogy instruction, systematic literature collection and positioning are performed to avoid missing relevant articles and causing sampling bias. Moreover, the participants included in the present study ranged from preschoolers to elementary, junior high school, vocational high school, and university students. The analogy instruction-related topics discussed included single analogies, analogy bridges, and multiple analogies. However, comments on or introductions of analogy instruction and scientific concept learning performance were excluded from analysis, as were articles on the development at different stages and other irrelevant topics. Because quantitative meta-analysis was conducted, qualitative studies were also disregarded, as were publications that failed to provide the information required for calculating effect sizes.

In meta-analyses, Hedges' g is used as a measure of the effect size of standardized mean differences (Hedges, 1982). Owing to pretest–posttest differences in data provided by different institutions, the effect sizes can be roughly classified as those of the mean raw score and the mean-covariate adjusted score (Morris & DeShon, 2002; WWC, 2020b). When the information on adjustments to pretest data, including statistics from analysis of covariance, adjusted means, and standard deviations, was provided, the effect size of the mean-covariate adjusted score was used. According to WWC recommendations (2020b), if this information was not provided but other data (e.g., pretest and posttest means and standard deviations) were and the pretest–posttest assessment tools were consistent, difference-in-differences was used. To account for pretest–posttest correlations, difference-in-differences adjustment was performed. When no pretest data were provided or only posttest-related statistics (e.g., from analysis of variance) were provided, the effect size was calculated from the mean-raw score.

Regarding the integration effect of analogy instruction in promoting students' scientific concept learning performance, a medium immediate effect size of 0.59 was obtained. The delayed effect size was reduced to 0.39 (medium).

Regarding the moderators of analogy instruction, single analogy regression failed to reach significance, but the p -value was 0.08, that appeared to become more significant or nearly significant. The effect size of analogy instruction for the elementary school students was 0.88, whereas that for students in junior high school and above was 0.52, constituting a significant difference.

With regard to seeking more appropriate indicators for assessing the quality of primary research, three indicators are proposed herein: baseline equivalence of the pretest, reliability and validity of assessment tools, and sample size. For baseline equivalence, the effect size was not significantly different regardless of whether the baseline equivalence of the pretest was even. Regarding the selection of effect calculations, even if the attrition rate was excessively high (more than 10%) or the pretest difference was greater than 0.5 standard deviations, they were classified as nonconforming indicators. By this standard, as long as statistics from covariate analysis or pretest and posttest data were provided, the effect size was estimated by using the mean score after covariate adjustment. This appeared to compensate for both types of possible differences, which in turn led to nonsignificant results.

Regarding the reliability and validity of assessment tools, the regression coefficient β was -0.57 ($p\beta < 0.0001$). The effect size of 1.06 was large. The result indicates that obtaining a large effect size is simple if the reliability and validity of assessment tools are not addressed. Regarding the sample size, the regression coefficient β for the basic number of participants was -0.81 ($p\beta = 0.003$), demonstrating that a favorable effect size of 1.36 could be easily obtained for small-scale studies on analogy instruction. As Slavin (2008) noted, small-scale studies are likely to obtain extremely positive or extremely negative result.

In terms of meeting the three indicators, primary research considering all levels of validity had a medium effect size of 0.49. Similar results were obtained regardless of the quality assessment method used. The effect sizes of the studies that did not meet the quality assessment standards exceeded those of the studies that did, indicating that the effects of analogy instruction in primary research may be overestimated if quality assessment is not considered.

Significant effects were observed for both analogy instruction approaches and learning stages. The regression coefficient β values were 0.27 ($p\beta = 0.03$) and -0.41 ($p\beta = 0.01$), respectively. For the elementary school learners, the effect size of 0.77 for single analogies was large. The effect size for approaches other than single analogies was larger (1.04). For single analogies and nonsingle analogies among students in junior high school and above, the effect sizes were 0.36 and 0.63, respectively—roughly in the upper and lower intervals of medium effect sizes. Therefore, different analogy instruction topics had different effect sizes for students at different learning stages, with better results for topics other than single analogies.

Regarding publication type, multiples analysis was performed on moderators of the analogy instruction topics and learning stages. A tendency toward positive effects was observed. When publication type is controlled and two moderators are examined at the same time, publication type had little influence on analogy instruction.

Regarding the baseline equivalence of the pretest, the comparable regression coefficient β values were 0.07 ($p\beta = 0.72$) and 0.03 ($p\beta = 0.89$) to the analogy instruction topics and learning stages, respectively. Neither were significant. According to the study design and the information provided, the initial selection of an appropriate calculation scale can facilitate the adjustment of the comparability of the baseline equivalence of the pretest.

When all three indicators (i.e., baseline equivalence, reliability and validity of assessment tools, and sample size) were met, the mentioned conditions exerted impacts on analogy instruction topics. When the conditions were not fulfilled, the influence of other moderators or the effects of analogy instruction could be misestimated or overestimated, respectively.

The conclusions are as follows: First, analogy instruction had an overall positive effect. Second, when the impacts of quality indicators were not controlled, single analogies were less efficient than other analogy instruction topics, such as multiple and bridging analogies. Furthermore, the elementary school students performed better, especially when they received instruction on topics other than single analogies. Finally, publication type is not a favorable indicator in the quality assessment of primary research because of mixed impacts with publication bias. However, alternatives can be well defined both theoretically and empirically through the examination of baseline equivalence, the reliability and validity of assessment tools, and the establishment of basic sample size requirements.

Keywords: analogy instruction, conceptual change, meta-analysis, quantitative research synthesis, quality assessment of primary research