

以統合研究建立相互教學有效性的在地證據：研究方法特徵影響評估

謝進昌* 陳敏瑜**

摘要

檢視過去統合研究，發現學者多關注文獻出版類型可能導致偏誤，鮮少針對研究方法特徵於統合結果影響進行評估與討論。據此，本文目的在探究研究方法特徵（實驗研究設計、評量工具與教學者來源）對於相互教學在促進不同學習階段學生，中文閱讀理解統合成效的影響。研究結果顯示，相互教學對於學生中文閱讀理解平均效果量為近中等程度教學成效，此外，進一步評估研究方法特徵影響時，發現當納入實驗研究設計因子，相互教學對參與者為國小高年級（ $\beta=-0.2736$, $p_{\beta}=0.0048$ ）、或國中以上學生（ $\beta=-0.2427$, $p_{\beta}=0.0481$ ）之教學效果皆顯著低於參與者為國小中年級學生。然而，評量工具來源、教學者是否為研究者則未產生顯著影響。最後，本文針對分析結果，提出未來實務與研究建議。

關鍵字：後設分析、研究方法特徵、研究品質評估、量化研究統合

* 第一作者：國家教育研究院副研究員

** 通訊作者：臺北市立大學兼任助理教授

Email：d970103@gmail.com

誌謝：本文作者感謝科技部經費補助，以使本文順利完成。研究計畫編號：MOST 106-2410-H-656

-003

投稿日期：109年3月9日；修改日期：109年4月24日；採用日期：109年6月19日

壹、研究背景與目的

學者近幾年提出以實徵證據作為導引決策的理念(evidence-based education)(鈕文英, 2010; Slavin, 2017), 證據來源之一即量化研究統合(quantitative research synthesis)結果, 其策略強調系統性進行窮竭式文獻回顧, 並藉由後設分析, 整合各研究量化數據, 形成可靠結論, 再轉化為政策推動基礎(Cooper, 2016)。隨著此理念擴散, 學者開始關注「研究品質」於形成證據重要性, 策略之一是透過負面表列可能偏誤進行影響評估, 而過去最常被學者探究議題即出版偏誤(publication bias)(謝進昌、陳敏瑜, 2011; Rosenthal, 1979), 其成因源自於研究者搜集文獻結果, 無法代表現行已執行全部研究成果, 因而可能產生已出版類型(有搜集)研究平均效果顯著高於未出版類型(未搜集)研究之平均效果量(Lipsey & Wilson, 1993), 致使研究者可能提出高估的統合結果。然而, 作者觀察近幾年不同教育介入主題的統合研究概況時, 發現某些替代性研究品質評估指標, 是源自研究方法特徵(methodological features), 似乎也具有類似影響力與具備效度理論, 卻鮮少為學者進行探究。

針對統合研究的個別研究品質評估, 教育學者提出許多正面表列評估指標(Gersten, Fuchs, Compton, Coyne, Greenwood, & Innocenti, 2005; Valentine & Cooper, 2008), 美國教育部教育科學院(Institution of Educational Studies, IES)成立之「有效教育介入資料中心」(What Works Clearinghouse, WWC)(WWC, 2017a)則建立出系列性評估機制與具體篩選標準。這些評估指標多以實驗研究效度理論基礎, 雖然具嚴謹性卻不見得適合一般統合研究的推動, 例如: Valentine 與 McHugh (2007)發現參與者流失率與效果量, 並不具顯著關聯性, 然而, 此指標卻在 WWC 評估機制中, 扮演重要角色。作者進一步發現過去統合實徵研究結果源自於「實驗研究設計、評量涉及內容教學的不均衡、研究者教學期望」等成因, 可能具有較直接的潛在影響力進行探究, 其中, 實驗研究設計係涉及 WWC 評估機制確保實驗、對照組表現基準之可比較性, 而評量涉及內容教學的不均衡、研究者教學期望則曾為過去學者所關注(Allen, 2015; Slavin & Madden, 2011)。

本文主題聚焦於促進學生中文閱讀理解, 以 Palincsar 與 Brown (1984) 早年提出的相互教學較受到學者重視, 發展出許多延伸或修訂教學方法, 教學效果也多獲得量化研究統合的支持(連啟舜, 2002; 謝進昌, 2015; Davis, 2010; Galloway, 2003; Rosenshine & Meister, 1994), 然而, 在統合研究脈絡下, 發現調節變項(moderator)

(或稱脈絡變項 (contextual variables)) 分析仍存在待發展空間。本文作者依知識探索面向，將調節變項分為兩類，一是教學實務焦點變項，其功能為回應現行策略教學效益，是否存在其他實務教學可操弄與解釋的知識性探索，包含教學方法、理解策略類型、學生特徵等，皆是實務教學現場所關注焦點，教師可就研究統合結果，選擇有效教學路徑，而本文聚焦於學生年級，原因在於相互教學屬於多元理解策略教學 (multiple comprehension strategy instruction) (Davis, 2013; National Institute of Child Health and Human Development(NICHHD), 2000)，其成效涉及學習者自我調整與後設認知的發展，使學生年級成為重要探索因子，然而，此變項較鮮少為相互教學統合研究者進行分析。另一是源自實驗研究方法特徵之變項，源自研究方法特徵導致偏誤，其功能為回應現行策略教學效益，是否存在研究方法特徵所導致統合結果偏誤，其角色作為執行後設迴歸分析時，進行對於教學效果量影響控制的用途，以降低可能干擾真實教學效果統合估計之因子，本文聚焦於實驗研究設計、評量工具來源、研究者是否為教學者等。

就上述研究背景下，本文關注研究方法特徵對於教學成效統合之調節，除系統性文獻回顧過去臺灣相互教學促進不同學習階段學生閱讀理解之成效外，同時評估實驗研究設計、評量工具與研究者來源可能造成的影響，以期研究結果能提供實驗教學研究者了解，可能影響教學效果之研究方法特徵，並於未來設計時適時避免，同時提供實務教學者了解相互教學對於不同學習階段學生之教學效果。本研究待答問題如下：

問題一：相互教學對學生中文閱讀理解之統合平均效果量為何？

問題二：相互教學對促進不同學習階段學生中文閱讀理解之統合平均效果量為何？

問題三：研究方法特徵（實驗研究設計、評量工具來源、研究者是否為教學者）對相互教學促進不同學習階段學生中文閱讀理解統合成效影響為何？

貳、文獻探討

茲就本文相關議題進行文獻回顧如下：

一、名詞定義

茲針對本文重要名詞進行定義。首先，有關統合研究係指針對某特定主題，透

過有系統的程序進行文獻檢索、定位、蒐集與訊息建檔、品質評估後，透過後設分析針對量化數據進行效果量計算與結果整合，以提出綜整性結論；相互教學定義係源自 Palincsar 與 Brown 於 1984 年所提出融合師生對話討論與閱讀策略教學活動，其核心為透過師生互動以作為鷹架協助（scaffolding），再藉由教導預測、提問、摘要與澄清等策略，以協助學生逐步能獨立發展與建構自身閱讀理解；在地證據係源自於以實徵證據作為導引決策的理念，該理念強調決策的基礎建立在有品質之研究統合實徵證據，進而，本文加以援引，並強調在地（以臺灣參與者為標的）實驗研究的統合結果，最能反映臺灣（在地）教學效果，故，稱為在地證據；最後，有關研究方法特徵影響評估，係指在進行統合研究時，對於該研究品質的評估，由於相關指標十分多元，因此，本文根據效度理論與實際可得性，針對該研究方法特徵，包含有實驗研究設計、評量工具來源、研究者是否為教學者等，作為品質評估指標。

二、研究品質與證據形成

在這股證據為基礎進行有效教學潮流下，學者們提出了以最佳品質研究證據進行統合（best evidence synthesis, BES）（Slavin, 1986, 2008）概念，此理念是廣為美國 What Works Clearinghouse（WWC, n.d.）、或非營利組織 The Campbell Collaboration（C2, n.d.）、Best Evidence Encyclopedia（BEE, n.d.）等採用，其推動概念在執行系統性文獻回顧時，採用「事前排除」策略，即預先針對「研究品質」設定最低納入準則，僅就符合準則之最佳品質研究進行後設分析並提出統合結果作為教學有效性證據，然而，利用此途徑進行研究統合，最大議題在於訂立研究品質篩選指標與機制。單就教學介入實驗研究而言，Slavin（1986, 2008）早先曾提出以有無設立控制組、受試者是否隨機分派、實驗與控制組前測表現基準是否均等、受試者流失率、成果評量適切性與教學時間等數項指標進行研究品質篩選，後續，Gersten 等人（2005）針對實驗與準實驗教學研究提出 18 項研究評估問題，Valentine 與 Cooper（2008）則區分出四大類，共計 35 項評估問題等，由此發現，過去學者所提出的研究品質篩選或評估指標（quality indicator）相當多，而相關組織也曾利用這些指標的組合建立符合自身目的之研究篩選機制，然而，這些準則大多是從實驗研究效度理論（Shadish, Cook, & Campbell, 2002）框架，由上而下（top to bottom）建構而成，未考量研究統合實務現況與可能問題、影響。舉例來說，WWC（2017a）研究品質篩選機制強調該研究受試者是否隨機分派至不同實驗組別，但受限於教育實務現場條件，研究者往往較少依

照此準則進行研究設計，如此，容易導致 WWC 統合研究數量只有個位數、甚至完全沒有符合準則的研究，例如，WWC (2010) 在系統性文獻回顧相互教學對於一般學生閱讀理解成效時，經檢核 164 篇研究，僅納入 5 個符合品質標準及 1 個保守符合品質標準的研究，排除 158 篇研究。

透過最佳品質研究進行統合以形成證據策略，其強調統合研究結果具有較佳(內部)效度，然而，此策略卻容易相對降低統合研究原有功能-對於結論概化 (generalization) 程度、及可能產生研究選擇偏誤等，據此，有別於前述事前排除策略，本文企圖從研究方法特徵著手，進行事後影響評估，此舉不僅可確保研究數量，亦即在控制或評估不同研究品質影響下，探究相互教學有效性。

三、相互教學與量化統合研究

閱讀教學方法十分多元，本文以促進一般學生中文閱讀理解策略的教學為核心，再聚焦於 Palincsar 與 Brown (1984) 相互教學 (或譯為交互教學) (reciprocal teaching) 及其延伸相關研究，並透過系統性文獻回顧與後設分析 (或譯為統合分析) (systematic review and meta-analysis) 策略進行量化研究統合 (quantitative research synthesis) (Cooper, Hedges, & Valentine, 2009)。

美國閱讀小組 (Reading Study Group, RAND) 認為閱讀歷程係為讀者與書面語言互動中所進行一連串訊息提取與意義建構過程，不僅是一項學習活動，也是一項學習其它知識與獲取其它能力的途徑，而影響閱讀歷程因素，大致牽涉到讀者、文本、閱讀情境脈絡與讀者、文本間交流過程等 (RAND, 2002)。其中，透過以策略為本教學 (strategy-based instruction) 與以討論為本教學 (discussion-based instruction) 則有助於學生了解文本結構、類型、及促進自身與文本交流中意義的建構 (Goldman, 2012)。

Palincsar 與 Brown 於 1984 年提出融合對話討論、與策略教學活動，現行稱為相互教學，其理念著重師生間對話互動提供鷹架協助，藉由系列性課室閱讀策略教學，協助學生建構與監控閱讀理解，再逐步拆鷹架、釋放責任於學生，獨立發展自身閱讀能力。在後續研究，Rosenshine 與 Meister (1994) 將相互教學區分為單純相互教學與相互教學前進行明示策略教學，差別在於前者對於策略指導是融入師生討論文本意義對話，後者則在建構文本意義前，先進行策略明示教學，但不論為何，策略大致包含預測、提問、摘要與澄清等。預測策略係連結學生先備知識與文本線索，透過

假設形成與驗證循環過程，以推測文本可能發展；提問策略係以問題為核心導引學生聚焦文本重點，而最早提問問題類型大致以 6W（Who, What, Where, When, Why, How）為主軸（Palincsar, 1987）；摘要係藉由指導學生刪除不重要與重複訊息，找出文本主要概念，並透過歸納及合併，經濃縮以建構文本意義；澄清係引導學生確認文本困惑部分，並提供適當澄清措施。

相互教學對於促進學生閱讀理解有效性，早已經獲得許多量化研究統合支持（連啟舜, 2002; 謝進昌, 2015; Davis, 2010; Galloway, 2003; Rosenshine & Meister, 1994），然而，較鮮少有系統分析研究方法特徵對於統合效果影響。據此，在納入這類關係研究品質調節因子下，其統合結果是否有別於過去研究，為本研究探究焦點。

四、學生年級與閱讀教學統合成效之關聯

Chall 於 1967 年提出閱讀發展階段理論，指出四年級學生是學習閱讀及透過閱讀進行學習間歷程轉換關鍵期，突顯出學生於不同學習階段閱讀發展重要性。後續，Myers 與 Paris（1978）、Paris、Wasik 與 Turner（1991）等學者也提出閱讀後設認知知識、策略發展研究，認為學生年級與其擁有文本結構、多元閱讀目標與策略知識，具有關聯性，顯示學生年級所扮演關鍵角色。直至近年，閱讀教學效益與學生年級調節關聯，也逐漸受到統合學者關注（連啟舜, 2002; Davis, 2010; Herbert, Bohaty, Nelson, & Brown, 2016; Suggate, 2010），然而，除了早先 Galloway（2003）、Rosenshine 與 Meister（1994）統合研究外，鮮少研究曾針對特定相互教學取徑進行相關議題探索。

回顧過去閱讀教學統合結果，連啟舜（2002）發現中文閱讀教學對於國小學習階段學生教學效果量為 0.4429，相較於國中學習階段學生教學效果之 1.0423 相對偏低，而就結果方向性是呼應 Galloway（2003）統合發現，然而，Galloway 進行統計顯著考驗時，並未發現年級間具有統計顯著差異。此外，Herbert 等人（2016）在透過後設迴歸分析進行文本結構教學效果，也發現國中學習階段學生效果高於國小學習階段學生，但其迴歸係數值並未達統計顯著；另一方面，Suggate（2010）統合多種閱讀教學時，發現五至七年級學生具有較高教學效果量（0.58），但未達顯著水準，其次則是二年級學生所獲得顯著 0.52 與三至四年級學生顯著 0.46；最後，Davis（2010）針對促進閱讀理解策略教學進行學生年級調節分析，發現四、七年級學生具有較高教學成效，但似乎會隨著研究者所使用評量工具來源差異而不同，顯示年級因素外，似

乎存在其他研究方法特徵干擾影響，而此結論也雷同 Rosenshine 與 Meister (1994)，認為當使用研究者發展的評量工具時，教學顯著成效會產生於四年級學生至成年人。

五、研究方法特徵與閱讀教學統合成效關聯

研究統合脈絡下，個別研究方法特徵對統合成效影響，某程度代表研究統合可能存在來自研究偏誤干擾，而檢視過去閱讀教學較受關注變項，除了出版偏誤（謝進昌、陳敏瑜，2011）外，大致可歸納為實驗研究設計、評量工具所涉內容教學不均衡、研究者教學期望等。

實驗研究設計類型多涉及內部效度，研究者透過實驗設計、或搭配統計分析，以排除、控制無關干擾因素影響，例如，隨機分派參與者、透過事前配對或分層進行參與者分派、抑或利用事後共變項校正統計分析技術等（Shadish et al., 2002），目的皆在確保或控制實驗、對照組表現基準是在一個可比較基礎，據此，實驗研究設計往往是統合研究者作為重要控制變項考量之一（Slavin, Lake, Chambers, Cheung, & Davis, 2009; WWC, 2017a）。就過去閱讀教學統合結果，Davis (2010) 發現使用標準化評量工具，針對前測不均等進行統計校正之研究會獲得相對顯著較低的教學效果，後設迴歸係數值 β 為 -0.32，而 Suggate (2010) 研究結果顯示準實驗研究（quasi-experimental design）所獲得統合效果量（0.64），大於透過參與者隨機分派研究設計（randomized-control design）（0.41），此外，若該統合研究已預先將研究品質作為文獻納入準則，實驗研究設計差異對於統合效果量影響，其影響會顯得較輕微（謝進昌，2015；Herbert et al., 2016; Slavin et al., 2009）。

就評量工具所涉內容教學的不均衡，概念如同 Slavin 與 Madden (2011) 在教育介入實驗研究發現，認為以教學方法（instruction based）為實驗教學主軸研究中（例如：本文的相互教學），評量內容是否均衡教導於實驗、對照組學生，會對教學效果產生顯著差異影響，若研究者未教導對照組學生相應之評量內容（稱為 treatment inherent measures）就容易產生內容教學不均衡的偏誤，而過去統合研究對此偏誤評估方式，多是以評量工具來源（即研究者自編/標準化評量工具）作為檢核指標，其背後理論假設研究者自編評量工具內容傾向與實驗教學內容重疊，進而偏好實驗組學生以干擾教學效果。若檢視過閱讀教學、或相互教學統合結果，此議題常受到研究者關注，會發現採用研究者自編評量工具會獲得相對較高平均教學效果（連啟舜，2002；謝進昌，2015；Davis, 2010; Galloway, 2003; Rosenshine & Meister, 1994），然而，礙

於研究者分析途徑差異，分析結果不必然產生統計顯著差異。

研究者的教學期望效果 (experimenter expectancy effect) 曾受到實驗教學研究者關注 (Allen, 2015)，認為研究者可能會因自身特性，例如，個人期望、自我價值或性別等因素，以致影響實驗研究結果。在統合研究脈絡下，Gersten 等人 (2005) 建議研究者可從實驗教學者特徵來檢視，其中，以檢視「研究者是否同時兼任教學者」為評估指標，較常受到重視，而就過去閱讀教學統合結果，連啟舜 (2002)、Davis (2010)、Galloway (2003)、Rosenshine 與 Meister (1994) 等研究皆發現不論教學者是否為研究者、或其它身分別，皆未產生顯著差異教學效果。

參、研究方法

本研究採用系統性文獻回顧與後設分析，前者係以標準化程序進行文獻蒐集，再依據納入與排除準則進行文獻定位，續透過資料建檔指標與品質維護後，建立本研究資料，後者則透過量化統計技術，經效果量計算、後設與調節變項分析，以回應本研究待答問題，茲分述如下：

一、標準化程序進行文獻蒐集

研究者透過臺灣期刊與博碩士論文資料庫進行文獻資料檢索，以主題相關關鍵字，包含相互教學、交互教學、互惠式教學、閱讀理解、reciprocal teaching、reciprocal instruction、reading comprehension 等中英文詞彙進行搜尋，截至 2019 年 1 月 10 日，採聯集方式於臺灣博碩士論文知識加值系統的論文名稱、關鍵字進行檢索時，總計回傳 1,995 筆資料，於臺灣期刊論文索引系統的篇名與關鍵詞進行檢索時，總計回傳 442 筆資料。在文獻資料庫檢索外，研究者覆核臺灣曾發表相關閱讀教學統合研究 (吳佩勳, 2015; 連啟舜, 2002; 謝進昌, 2015) 參考文獻，經交叉比對，以避免缺漏重要文獻。此外，對於可能發表於國外期刊、論文之相關文獻，本研究針對 EBSCOhost、ProQuest 系統，透過上述英語關鍵字與 Taiwan 的組合，於標題、摘要進行檢索，截至 2020 年 4 月 23 日前，分別回傳 282、208 筆，此外，同步檢核發表於英文期刊之交互教學相關統合或回顧性文獻，以避免可能缺漏。

研究者採兩階段篩選步驟針對每筆回傳訊息進行檢核，首先，就發表於國內檢索資料庫者，先利用下述文獻納入與排除準則，就主題名進行初步篩選，獲得 248

筆資料，其次，再就摘要、全文進行第二階段篩選。經檢核，排除非利用相互教學進行中文閱讀理解之個別實驗教學研究（132 篇）、針對特殊生或以華語為第二外語者（25 篇）、無提供對照組或採用單一受試者研究設計（40 篇）、與 2 篇作者不公開論文，此外，就發表於國外檢索資料庫者，經檢核，全部文獻係屬非臺灣參與者、或探究臺灣外語學習者之英語文理解為主，故排除。最後，本文共計獲得 49 篇文獻，覆經檢核使用同一研究參與者，排除重覆之 6 篇文獻係碩士論文轉投稿期刊者，最後，本文正式以 43 篇文獻進行統合。

二、納入與排除準則進行文獻定位

文獻納入與排除準則依下列五面向：

- (一) 必須涉入透過相互(交互)教學方法教導學生閱讀理解策略，其策略包含預測、提問、摘要與澄清等四類或其延伸，排除單一閱讀理解策略教學。
- (二) 參與者以繁體中文為母語之各學習階段一般(或中低學習成就)臺灣學生，排除特殊生，如閱讀障礙、學習障礙者，或以華語(英語)為第二外語者。
- (三) 教學成效必須涉及中文閱讀理解之評量，如，字面文意理解、推論理解等。
- (四) 必須涉及個別實驗教學研究，以佐證相互教學成效，設計包含參與者隨機分派研究、或其它提供對照組別之實驗設計，排除單一受試者、或未提供對照組之相關研究設計。
- (五) 結果必須提供適當閱讀理解表現量化數據，以計算效果量。

三、資料建檔指標與品質維護

(一) 資料建檔指標

本文資料建檔指標大致包含文獻背景特徵、相互教學、教學者、中文閱讀理解評量、研究參與者、實驗研究設計、效果量計算訊息等面向。除一般具體從原文內容，可直接記載與分類的指標，包含文獻背景(作者、出版年代、出版類型)、參與者年級、實驗教學時間、延宕施測時間、效果量計算數據等資料外，某些涉及專業判斷的指標，會由研究者依據下述建檔品質維護程序進行，包含相互教學、教學者、中文閱讀理解評量、實驗研究設計等，茲分述如下：

- 1、相互教學：本文係指 Palincsar 與 Brown (1984) 所發展教學方法，閱讀理解策略著重預測、提問、摘要與澄清等、或相關後設認知閱讀理解策略之調整教學，

而教學內涵重視師生討論與合作、閱讀策略教學、與責任轉移。

- 2、教學者：該文獻實際執行實驗教學者來源，區分為研究者、與非研究者，前者係指該研究作者就是實驗教學者，後者係指該研究作者非實驗教學者。
- 3、中文閱讀理解評量：該文獻所使用於評量參與者中文閱讀理解評量工具來源，區分為研究者自編、與引用既有標準化評量工具。
- 4、實驗研究設計：資料建檔者依據該文獻所使用之實驗設計類型，進行歸類者，區分為一透過研究設計可確保或合理假設實驗、對照組參與者前測表現基準均等之相關設計，包含：參與者個體隨機分派、參與者群體隨機分派、參與者隨機分派之迴歸-不連續設計（regression discontinuity design），前兩者係指透過隨機分派機制將參與者以個體或群體為單位分派至各實驗組別，而後者係指研究者預先依據某關鍵變項（例如，智力表現）結果，進行各組別參與者分派；另一為透過統計分析技術以校正實驗、對照組參與者前測表現的不均等，例如：採用共變數分析、WWC（2017b）採納 difference in differences 校正。最後，其他無法依照上述原則進行判斷者，則歸類為其他或未說明。

（二）資料建檔品質維護

本文就建檔者訓練、訊息建檔與覆核等面向著手。研究者預先針對建檔指標進行內涵說明，再交由建檔者實作練習，復經疑義討論與回饋修訂等，以便建檔者熟悉各指標。此外，為降低潛在人為建檔誤差，各建檔指標皆要求依照原文內容進行記載，再進行判讀歸類，以降低建檔者當下認知負荷，同時，每篇文獻訊息皆經由兩位建檔者依指標進行建檔與交叉覆核，若訊息存在不一致，會再次比對原文與討論，以確保訊息正確性，使各指標訊息建檔達到完全一致。最後，本文兩位建檔者皆長期涉入本領域者，皆有數年以上實驗教學研究建檔經驗，具備一定執行專業度。

四、效果量、後設與調節變項分析

本文對於實驗教學效果量計算乃採用標準化平均數差異效果量（standardized mean difference），公式則引用 Hedges（1981）針對參與者人數偏少進行校正之不偏 g 效果量，此外，考量不同文獻所提供數據具相當歧異，本文優先以能進行共變數（或前測）校正分數之效果量進行計算（謝進昌，2012；Borenstein, 2009；Morris & DeShon, 2002），其原始與校正之 g 效果量公式可分別表達為（1）、（2）：

$$g = \frac{\bar{Y}^{G1} - \bar{Y}^{G2}}{SD_{Y'}^{\text{pooled}}} \sqrt{1 - r_{XY}^2} \quad (1)$$

$$g^u = \left(1 - \frac{3}{4df-1}\right)g \quad (2)$$

自由度df為 $n^{G1} + n^{G2} - 3$ （假設有 1 共變項）， n^{G1} 、 n^{G2} 分別為實驗、對照組參與者人數，而 \bar{Y}^{G1} 為共變項校正之實驗組平均數、 \bar{Y}^{G2} 為共變項校正之對照組平均數、與共變項校正分數的整合標準差 $SD_{Y'}^{\text{pooled}}$ ，同時 r_{XY}^2 為參與者於共變項、依變項表現相關係數平方值。其次，本文則是以提供原始後測平均表現分數效果量進行計算，其公式與前述差別僅在除了是採用原始（實驗、對照組）後測分數平均數及其整合標準差外， r_{XY}^2 假設為 0，詳細計算公式可參考 Lipsey 與 Wilson（2001）。另一方面，由於計算公式涉及共變項、後測表現相關係數值，本文是以該篇研究有提供相關係數者進行計算、抑或是統一就其他有提供數據者，以固定效果模式進行後設分析，進行補插，其值為 0.7241。最後，考量該研究若未提供整體閱讀理解表現數據時，本文會視情況整合多個子評量、或子群體數據，進行該研究效果量計算，以避免可能產生的計算偏誤（Borenstein et al., 2009, p.215）。

在執行後設與調節變項分析前，本文採納 Marín-Martínez 與 Sánchez-Meca（2010）以變異數倒數（inverse variance）作為各個別研究加權值建議，並參考 Viechtbauer 與 Cheung（2010）對於影響值（influential case）檢測結果，以評估適合後續納入後設分析之個別研究。此外，本文認為除了部份組間效果量變異可由調節變項解釋之系統性影響來源外，各研究仍存在少許無法被解釋隨機變異，據此，採用混合效果模式之後設迴歸分析（mixed effect model-meta regression）（Borenstein et al., 2009; Thompson & Higgins, 2002），其模式可表達為：

$$g_i^u = \beta_0 + \beta_1 M_{i1} + \dots + \beta_p M_{ip} + \mu_i + \varepsilon_i \quad (3)$$

g_i^u 代表各研究(i)效果量， β_0 為模式截距項，若未投入任何調節變項時，它代表整體教學平均效果量，而 β_1 至 β_p 各調節變項(M_1 至 M_p)迴歸係數值，代表其對於整體平均效果量的影響權重，其中， μ_i 代表解釋研究間效果量之隨機變異效果，是服從常態分配 $N(0, \tau^2)$ ， ε_i 為各研究內抽樣誤差，服從常態分配 $(0, \nu_i)$ 。有關於各參數估計，本文採用限制最大概似估計法（restricted maximum likelihood）（Thompson & Sharp, 1999）並透過 Z 檢定進行各調節變項迴歸係數值之顯著考驗。有關研究間效果量異

質性評估，本文依照 Q 統計量來檢定異質性顯著性、及 Higgins 與 Thompson (2002) 建議 I^2 指標，顯示在隨機造成的機運外，整體異質程度有多少百分比是來自於實質異質性，其中，Higgins 等人 (2019) 提供了粗略建議，認為 I^2 界於 0-40% 為不重要、30-60% 為中等異質、50-90% 為大量異質，而 75-100% 為相當異質。此外，本文透過 Raudenbush (2009) 提出 $R^2 = (\tau_{RM}^2 - \tau_{ME}^2) / \tau_{RM}^2$ ，係為原始隨機效果模式（未投入調節變項前）所估計組間變異量（即 τ_{RM}^2 ）相對於投入調節變項後之混合效果模式所估計組間殘差變異量（即 τ_{ME}^2 ）差異值，來詮釋投入該調節變項後，整體異質性降低比率概況。

有關出版偏誤可能的影響評估，本文利用漏斗圖 (funnel plot) (Light & Pillemer, 1984) 及以漏斗圖為基礎所發展的線型迴歸檢測圖形對稱性 (Egger, Davey, Schneider, & Minder, 1997)、與 trim 與 fill 方法 (Duval & Tweedie, 2000) 評估可能因出版偏誤所導致缺失研究及其影響。最後，本文使用 Viechtbauer (2010, 2020) 開發 metafor 套件，在自由軟體 R 平台進行相關後設與調節變項分析程序。

肆、結果與討論

茲就結果與討論，分述如下：

一、臺灣相互教學實驗研究現況

本文計獲得 43 篇文獻，特徵摘要如表 1 所示。各文獻出版年代界於 1990 至 2017 年，發表最高峰是在 2000 年至 2015 年，顯示臺灣近二十年來才興起對相互教學的應用，就參與者學習階段，文獻多分佈於國小高、與中年級，分別有 19 篇、10 篇研究，顯示相互教學多具一定程度認知發展學生為實驗對象，而不同研究總實驗教學時間，介於 180 至 1920 分（約 4 節至 48 節課）。就參與者總人數，實驗組累計為 1,493 名、對照組累計為 1,437 名，總計 2,930 名，其中，參與者整體樣本流失率多為 0%（計 38 篇），其餘 5 篇文獻整體樣本流失率皆是小於 15%，加諸，兩群體參與者流失差異比率也均小於 10%，皆符合 WWC (2017a) 所界定的低參與者流失率，整體而言，參與者流失屬輕微，未對本文統合結果造成過多影響。

就實驗研究設計而言，12 個研究宣稱以參與者群體（例如：班級）為單位進行隨機分派，2 個研究為迴歸-不連續設計，此外，28 個研究僅以統計分析技術進行前

測表現不均等之校正，最後，有 1 筆研究未採用上述策略確保前測表現的均等性。就評量工具來源而言，21 個使用研究者自編評量工具，22 個直接引用既有標準化評量工具，進行閱讀理解測量，最後，有關教學者來源，僅有 6 個研究之教學者為非研究者，其餘皆為研究者，顯示教學者多由研究者（即可能是實務現場之教師）擔任的現況。整體而言，這些變項是否可能是造成教學偏差之影響因子，仍待後續顯著考驗分析。

表 1

本文所納入研究文獻概況

研究特徵	調節變項虛擬建檔	細項	比較研究數
出版年代	--	1990-1995	4
		1996-2000	3
		2001-2005	10
		2006-2010	12
		2011-2015	13
		2016-2017	1
出版類型	未出版文獻 (Ref.)	博士論文	1
		碩士論文	34
	已出版文獻	期刊	2
		論文轉投期刊	6
參與者學習階段	國小低年級與 學前	學前幼兒	1
		國小低年級	7
	國小中年級 (Ref.)	國小中年級	10
		國小高年級	19
	國中以上	國中	6
	高中職以上	3	
教學時間	--	180 至 1920 分鐘	43
實驗研究設計	以設計確保前 測均等(Ref.)	參與者群體隨機分派	12
		參與者隨機分派之迴歸- 不連續設計	2
	以統計方法校	依前測表現進行統計校	22

正前測表現		正(共變數分析)	
		WWC(2017b)difference	
		in differences 校正	6
評量工具來源	其他(NA)	其他	1
	自編	研究者自編評量工具	21
	引用(Ref.)	引用既有標準化評量工具	22
實驗教學者是 否為研究者	非研究者	教學者非研究者	6
	研究者(Ref.)	教學者為研究者	36
參與者總人數	--	60 人以內	24
	--	界於 61 至 90	16
	--	大於 91 人	3
參與者整體樣 本流失率	--	0%	38
	--	小於 15%	5
參與者於各組 別樣本流失差 異比率	--	0%	38
	--	小於 10%	5

註：1. Ref.係指執行後設迴歸分析時，針對該類別變項進行虛擬變項建檔轉換之參照水平。2. NA 係指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、相互教學於學生中文閱讀理解統合成效

回應研究問題一，本文以隨機效果模式進行後設分析，結果如表 2 所示，顯示 43 筆研究立即效果為 0.5593[95%CI=0.4397, 0.6788]、13 筆研究延宕效果為 0.3767 [95%CI=0.2781,0.4754]，分別屬於 Cohen (1992) 建議中等、與介於小至中等效果量，然而，以 Viechtbauer 與 Cheung (2010) 策略進行影響值評估，發現立即效果有 2 筆潛在影響值，分別是林容妃(2005)、吳潔蓉(2010)，加諸刪除劉永慈、陳明珠(2017) 特別極端值(平均效果量為 6.1518)，而延宕效果則無發現任何影響值。經排除這 3 筆研究後，本文再次後設分析 40 筆研究之立即效果，統合結果略降為 0.4756[95%CI=0.3995, 0.5516]，其中，整體異質性評估指標 $Q(T)$ 值變為 83.59($p_{Q(T)} < .0001$)，而 I^2 變為 52.58%，依據 Cochrane 建議準則，已經由大量異質 (substantial) 降至中等異質 (moderate) (Higgins, et. al., 2019)。整體而言，相互

教學具有促進學生閱讀理解的正向效益，而此結果也雷同於過去統合研究發現（連啟舜，2002；謝進昌，2015；Davis, 2010; Galloway, 2003; Rosenshine & Meister, 1994）。

檢核出版偏誤可能影響時，由於不同方法皆有其缺點與限制情境，故，本研究採用多重方法來評估出版偏誤可能影響，首先，本文以漏斗圖進行圖形對稱性檢核，結果分別如圖 1（立即效果）、2（延宕效果）所示，以主觀檢視，延宕效果似乎在右下角出現不對稱現象，進一步以線性迴歸分析、與 trim 與 fill 方法進行統計考驗。不論是立即（ $z = 1.1473, p = 0.2512$ ）或延宕（ $z = -0.8944, p = 0.3711$ ）效果，線性迴歸分析皆未出現顯著結果，然而，trim 與 fill 方法檢定出延宕效果疑似存在 2 個潛在缺失值，經補插，如表 2 所示，其統合結果為 0.3969[95%CI=0.3014, 0.4925]。此外，參考 Glass、McGaw 與 Smith（1981）對於出版偏誤評估建議，本文經分析已出版/未出版文獻的實驗教學研究效果，結果並未顯示出版類型具有調節效果，其迴歸係數值分別是 0.0714($p_{\beta}=.5089$)（立即效果）、-0.1106($p_{\beta}=.3502$)（延宕效果），結論大致雷同於前述發現，可大致忽略出版偏誤可能產生之影響。

表 2

相互教學促進學生閱讀理解之立即、延宕統合效果

類別	k	Hedges's		95% CI		Q(T)	$p_{Q(T)}$	I^2	τ^2
		unbiased g	p_g	LL	UL				
立即後測 (全部)	43	0.5593	<.0001	0.4397	0.6788	179.11	<.0001	81.65%	0.1216
立即後測- 刪除影響 值	40	0.4756	<.0001	0.3995	0.5516	83.59	<.0001	52.58%	0.0288
延宕後測 (全部)	13	0.3767	<.0001	0.2781	0.4754	6.83	0.8684	0%	0
延宕後測 -trim & fill 補插	15	0.3969	<.0001	0.3014	0.4925	9.51	0.7974	0%	0

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 異質性評估值。

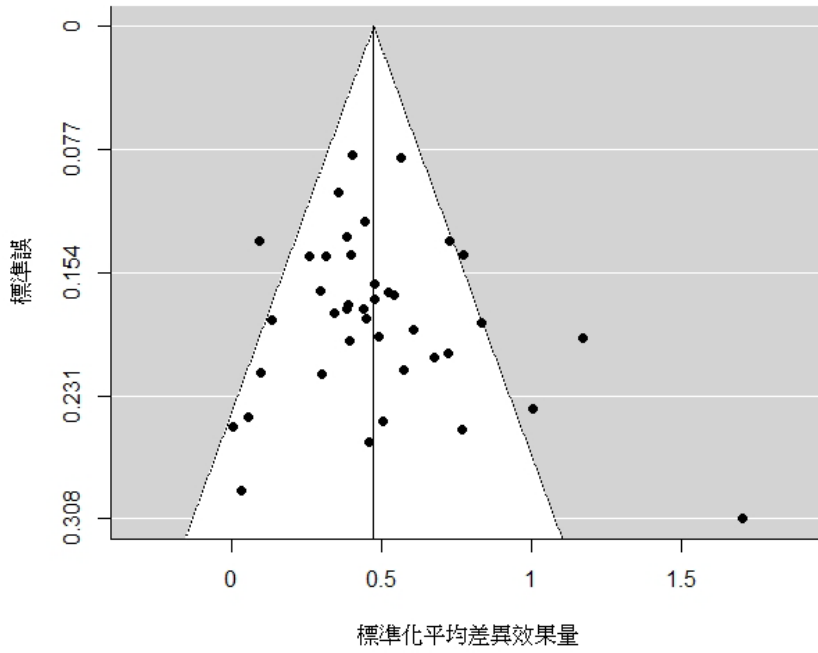


圖 1 立即效果量漏斗圖分佈

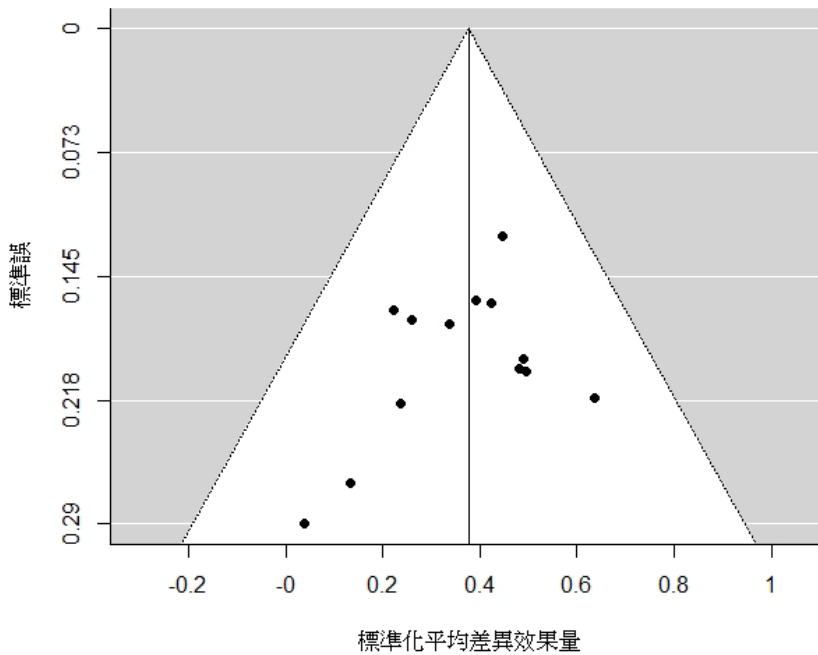


圖 2 延宕效果量漏斗圖分佈

三、影響相互教學成效之年級、研究方法特徵分析結果

就回應研究問題二，在檢視學生年級為焦點變項時，探討相互教學對於不同學習階段學生之統合成效，經考量研究數量多寡、年級合併適當性，本文合併學生年級，區分為國小低、中、高、與國中以上等學習階段，並以參與者屬於第二學習階段（即國小三、四年級）為對照組，進行後設迴歸分析，結果如表 3 所示，顯示相互教學對於國小低年級、高年級、與國中以上學生的教學效果皆低於國小中年級學生，迴歸係數值皆為負值，然而，在統計顯著水準 p 值標準設定為 .05 時，僅有國小高年級 VS. 國小中年級的教學效果達到統計顯著，其迴歸係數值 $\beta = -0.2227$ ($p_\beta = 0.0273$)，而國中以上 VS. 國小中年級則是屬於接近顯著水準 ($p_\beta = 0.0913$)，此外，投入學習階段變項時，組間效果量整體異質性 I^2 是自 52.58% (見表 2) 降為 49.47%，而可解釋變異量 R^2 為 9.56%。整體而言，本文顯示學生學習階段對於相互教學成效的調節效果，略不同於過去以非特定(交互教學)取徑之閱讀教學統合結果，並未發現國中學習階段學生（相對國小學習階段）具有較佳教學成效（連啟舜，2002；Galloway, 2003；Herbert, et. al., 2016），可能原因在於不同閱讀教學取徑具有各自適用學生學習階段，而這些研究文獻皆非針對特定相互教學取徑進行分析；然而，本結果符合 Davis（2010）發現四年級學生具有較高教學成效，不過，其間是否會隨著其它研究方法特徵影響而差異，仍待後續探討。

表 3

相互教學對於不同學習階段學生教學效果

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	0.6358	0.0851	<.0001	0.4690	0.8026				
國小 低年 級	-0.1159	0.1279	0.3649	-0.3665	0.1348	74.58 0.0002	5.43 0.1430	49.47%	9.56%
國小 高年 級	-0.2227	0.1009	0.0273	-0.4205	-0.0250				
國中 以上	-0.2154	0.1276	0.0913	-0.4655	0.0346				

註： β =迴歸係數； SE_{β} =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 異質性評估指標； R^2 =整體解釋變異量。

研究待答問題三，是探討當控制不同研究方法特徵時，統合成效之調節效果與變化，以了解來自「實驗研究設計、評量工具來源、研究者是否為教學者」等可能影響。本文預先投入學生學習階段變項為基礎，再逐一投入實驗研究設計、評量工具來源、教學者是否為研究者等變項，以多元後設迴歸分析各調節變項直接效果、交互作用效果，結果如表 4 所示。以實驗研究設計而言，預先檢視模式 2 分析結果，可發現其與學生學習階段的交互作用效果並未達顯著水準，迴歸係數 β 分別為 0.0603 ($p_{\beta}=0.8672$)、0.0028 ($p_{\beta}=0.9914$)、0.0180 ($p_{\beta}=0.9558$)，代表可直接檢視模式 1 的直接效果，結果顯示實驗研究設計為使用統計方法進行前測校正時，相互教學效果顯著低於透過研究設計（如以參與者群體為單位進行隨機分派）確保前測均等者 ($\beta=-0.1559$, $p_{\beta}=0.0485$)，此外，納入實驗研究設計因子，相互教學對不同學習階段學生的教學效益也相應產生變化，顯示參與者為國小高年級 ($\beta=-0.2736$, $p_{\beta}=0.0048$)、或國中以上學生 ($\beta=-0.2427$, $p_{\beta}=0.0481$) 之教學效果皆顯著低於參與者為國小中年級學生，而組間效果量整體異質性 I^2 降為 41.16%，可解釋變異量 R^2 提

高為 38.02%，是具有相當程度解釋力，此結果大致與 Davis（2010）發現針對前測不均等進行統計校正者，會獲得相對顯著較低的教學效果相符合。

在分別檢視評量工具來源、教學者是否為研究者時，就模式 4、6 分析結果，可發現各自交互作用效果皆不顯著，進而，就模式 3、5 進行分析時，顯示採用評量工具為研究者自編者，教學效果並未顯著不同於評量工具來源為引用既有標準化評量工具（ $\beta=-0.0023$, $p_{\beta}=0.9767$ ），教學者非研究者之教學效果也未顯著不同於教學者為研究者（ $\beta=0.0434$, $p_{\beta}=0.6876$ ），其中，前者結果不同於過去統合研究發現，後者則大略符合（連啟舜，2002；謝進昌，2015；Davis, 2010; Galloway, 2003; Rosenshine & Meister, 1994）。整體而言，本文在比較控制不同研究方法特徵時，發現實驗研究設計對於教學成效具有相對顯著影響力，而當控制其影響時，相互教學對於國小中年級學生具有最佳的教學效益。

表 4

在控制不同研究方法特徵下，相互教學於不同學習階段學生之教學成效

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				
模式 1									
截距	0.7634	0.1028	<.0001	0.5618	0.9650				
國小 低年 級	-0.1178	0.1186	0.3208	-0.3502	0.1147				
國小 高年 級	-0.2736	0.0971	0.0048	-0.4638	-0.0833	65.63 0.0009	10.08 0.0391	41.16%	38.02%
國中 以上	-0.2427	0.1228	0.0481	-0.4834	-0.0020				
研究 設計- 統計 校正	-0.1559	0.0790	0.0485	-0.3107	-0.0011				
模式 2									

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	0.7734	0.2176	0.0004	0.3470	1.1998				
國小 低年 級	-0.1654	0.3322	0.6186	-0.8166	0.4858				
國小 高年 級	-0.2824	0.2306	0.2208	-0.7344	0.1696				
國中 以上	-0.2609	0.2868	0.3630	-0.8231	0.3013				
研究 設計- 統計 校正	-0.1625	0.2368	0.4926	-0.6267	0.3017				
國小 低年 級*研 究設 計-統 計校 正	0.0603	0.3605	0.8672	-0.6464	0.7670	65.60 0.0003	8.60 0.2828	49.96%	10.59%
國小 高年 級*研 究設 計-統 計校 正	0.0028	0.2609	0.9914	-0.5085	0.5142				
國中 以上* 研究 設計-	0.0180	0.3248	0.9558	-0.6187	0.6547				

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				
統計 校正									
模式 3									
截距	0.6376	0.0905	<.0001	0.4601	0.8150				
國小 低年 級	-0.1144	0.1316	0.3850	-0.3724	0.1437				
國小 高年 級	-0.2236	0.1034	0.0307	-0.4263	-0.0208	74.41 0.0001	5.25 0.2623	51.43%	1.59%
國中 以上	-0.2156	0.1321	0.1025	-0.4744	0.0432				
評量 工具- 自編	-0.0023	0.0794	0.9767	-0.1580	0.1534				
模式 4									
截距	0.6636	0.1122	<.0001	0.4436	0.8835				
國小 低年 級	-0.0938	0.1850	0.6123	-0.4564	0.2688				
國小 高年 級	-0.2731	0.1360	0.0446	-0.5395	-0.0066	74.00 <.0001	5.27 0.6268	57.02%	<0.01%
國中 以上	-0.2375	0.1947	0.2225	-0.6192	0.1441				
評量 工具- 自編	-0.0773	0.2018	0.7019	-0.4728	0.3183				
國小 低年 級*評	-0.0055	0.2888	0.9848	-0.5716	0.5606				

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				
量工 具-自 編									
國小 高年 級*評 量工 具-自 編	0.1326	0.2350	0.5727	-0.3281	0.5933				
國中 以上* 評量 工具- 自編	0.0689	0.2898	0.8120	-0.4990	0.6368				
模式 5									
截距	0.6018	0.0954	<.0001	0.4147	0.7888				
國小 低年 級	-0.0932	0.1359	0.4928	-0.3595	0.1731				
國小 高年 級	-0.1961	0.1086	0.0710	-0.4091	0.0168	71.95 0.0002	3.99 0.4068	51.14%	<0.01%
國中 以上 教學 者-非 研究 者	-0.1811	0.1362	0.1835	-0.4481	0.0858				
模式 6									
截距	0.5870	0.1124	<.0001	0.3667	0.8072	71.36	5.58	56.76%	<0.01%

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				
國小 低年 級	-0.1418	0.1689	0.4013	-0.4728	0.1893	< .0001	0.4717		
國小 高年 級	-0.1562	0.1303	0.2306	-0.4117	0.0992				
國中 以上	-0.1657	0.1531	0.2791	-0.4657	0.1344				
教學 者-非 研究 者	0.1406	0.2620	0.5915	-0.3729	0.6541				
國小 低年 級*教 學者- 非研 究者	0.1327	0.3469	0.7019	-0.5471	0.8126				
國小 高年 級*教 學者- 非研 究者	-0.2463	0.3075	0.4231	-0.8490	0.3563				
國中 以上* 教學 者-非 研究 者	---	---	---	---	---				

註：符號意涵同表 3。

伍、研究結論與建議

統合研究除能提供特定教學主題暫時性的整合結果，以利研究者評估其效益外，其優點在於透過調節變項評估，得以探討個別實驗教學較未能觸及的因子，而本文目的在聚焦評估研究方法特徵的影響，假設來自實驗研究設計、評量工具所涉內容教學的不均衡、研究者的教學期望效果等因素，可能干擾統合結果，若能同時納入考量與評估，可提供較適當的分析結果與結論。

本文經刪除影響值、評估出版偏誤可能影響，再經後設分析，結果顯示相互教學對於學生中文閱讀理解教學平均效果量，分別具有立即正向 0.4756 平均效果、與延宕正向 0.3969 平均效果，若依據 Cohen (1992) 建議，皆為中等程度教學成效，大致雷同於過去統合研究發現。此外，在評估焦點變項時，結果顯示學生學習階段會調節相互教學效益，加諸，逐一評估與控制實驗研究設計、評量工具來源、與教學者是否為研究者等研究方法特徵時，發現在納入實驗研究設計因子，相互教學對於參與者為國小高年級、或國中以上學生之教學統合效果，皆是顯著低於參與者為國小中年級學生，然而，本文在檢視評量工具來源、教學者是否為研究者之影響時，結果則未發現類似顯著影響。整體而言，在同時評估研究方法特徵可能影響下，相互教學對於促進國小中年級學生的中文閱讀理解，是具有最佳教學效果，值得教學現場推廣。

統合研究有效性逐漸受到學者關注，而此點可從 WWC (2017a) 提出研究品質評估準則、或 Slavin (2008) 重視以最佳品質研究證據進行統合 (best evidence synthesis) 理念一窺端倪，然而，前述作法多從 (事前) 正向表列可能研究特徵，進行品質判斷，而本文則是從 (事後) 負向評估可能偏誤影響著手，前者優點在於具有充分效度理論基礎，然，其缺點就是逐一檢核可能研究特徵，所費不貲，同時，並非所有指標皆具有鑑別力，後者優點能聚焦特定因子，節省資料建檔成本，加諸透過統計方法得以同時評估各因子影響力，以獲得較穩定統合結果。若依據本文分析結果發現實驗研究設計具有相當影響力，然而，評量工具為研究者自編、或引用既有標準化評量工具則未如過去統合研究發現，具有顯著影響力，這代表臺灣相互教學研究文獻中，多數教學研究者在設計評量內容時，能顧及其與教學內容的差異性，避免產生偏好實驗組學生情況，此外，教學者是否為研究者的分析結果則如同過去研究發現，並未產生任何顯著影響力，顯示本因子對於現行閱讀教學研究效果干擾並不高。整體而言，在出版偏誤可能影響評估外，建議未來統合研究者可多考量來自實驗研究設計、評量工具來源

等因子的干擾。

在研究限制方面，本文雖然透過窮竭式搜尋策略進行本主題研究文獻擷取，避免可能研究缺漏影響，然而，截至 2018 年，本文僅能納入 43 筆文獻，若執行多元調節變項分析時，仍不可避免會因各類別研究數目多寡而影響統計考驗力，而此點可從「迴歸係數值標準誤偏高」一窺端倪，據此，本文建議未來研究者可考量逐步更新與累積研究文獻，以進行深入分析。

參考文獻

中文部分

- 吳佩勳(2015)。**國內閱讀交互教學研究成效之統合分析研究**(未出版碩士論文)。
國立中正大學，嘉義縣。
- 吳潔蓉(2010)。**運用交互教學法促進國小學童科學閱讀之成效**(未出版碩士論文)。
國立臺北教育大學，臺北市。
- 林容妃(2005)。**兒童科普讀物的閱讀理解教學對國小學童自然科學習之相關研究**
(未出版碩士論文)。國立臺北師範學院，臺北市。
- 連啟舜(2002)。**國內閱讀理解教學研究成效之統合分析研究**(未出版碩士論文)。
國立臺灣師範大學，臺北市。
- 鈕文英(2010)。**特殊教育證據本位實務之建立、鑑識與運用**。**南屏特殊教育**，1，
1-24。
- 劉永慈、陳明珠(2017)。**以圖畫故事書為教材運用交互教學法提升幼兒的閱讀理解能力**。**清華教育學報**，34(2)，1-44。
- 謝進昌(2012)。**不同實驗設計之標準化平均數差異效果量整合探討**。**屏東教育大學學報**，38，57-92。
- 謝進昌(2015)。**有效的中文閱讀理解策略：國內最佳教學實徵研究整合**。**教育科學研究期刊**，60(2)，33-77。
- 謝進昌、陳敏瑜(2011)。**國內教育、心理後設分析研究出版偏誤檢定之實徵分析**。**測驗學刊**，58(2)，391-422。

外文部分

- Allen, M. (2015). The experimenter expectancy effect: An inevitable component of school science? *Research in Education*, 94(1), 13-29.
- Best Evidence Encyclopedia[BEE] (n.d.). *Best Evidence Encyclopedia : Empowering educators with evidences*. Retrieved 23April, 2020 from the world wide web: <http://www.bestevidence.org/>
- Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta analysis* (pp. 221-235). New York: Russell Sage Foundation.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons.
- Chall, J. S. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.
- Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin*, 112(1), 155-159.
- Cooper, H. (2016). *Research synthesis and meta-analysis* (5th ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Cooper, H., Hedges, L.V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*. Russel Sage Foundation, New York.
- Davis, D. S. (2010). *A met-analysis of comprehension strategy instruction for upper elementary and middle school students*. Unpublished Doctoral dissertation, Graduate School of Vanderbilt University, Nashville, Tennessee.
- Davis, D.S. (2013). Multiple comprehension strategies instruction in the intermediate grades: Three remarks about content and pedagogy in the intervention literature. *Review of Education*, 1(2), 194-224.
- Duval, S. J., & Tweedie, R. L. (2000). A nonparametric trim and fill method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95, 89-98.
- 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.

- Galloway, A. M. (2003). *Improving reading comprehension through metacognitive strategy instruction: Evaluating the evidence for the effectiveness of the reciprocal teaching procedure*. Unpublished Doctoral dissertation, The University of Nebraska–Lincoln, Lincoln, Nebraska.
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*, 149-164.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Goldman, S. R. (2012). Adolescent literacy: Learning and understanding content. *The Future of Children, 22*(2), 89-116.
- Herbert, M., Bohaty, J. J., Nelson, J. R., & Brown, J. (2016). The effects of text structure instruction on expository reading comprehension: A meta-analysis. *Journal of Educational Psychology, 108*(5), 609-629.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128.
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539-1558.
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., Welch, V. A. (Eds.)(2019). *Cochrane handbook for systematic reviews of interventions*(2nd Ed.). Chichester (UK): John Wiley & Sons.
- Light, R., & Pillemer, D. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181-1209.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, London: Sage Publications.
- Marín-Martínez, F., & Sánchez-Meca, J. (2010). Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educational and Psychological*

Measurement, 70, 56-73.

- 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.
- Myers, M., & Paris, S. G. (1978). Children's metacognitive knowledge about reading. *Journal of Educational Psychology, 70, 680-690.*
- National Institute of Child Health and Human Development [NICHD]. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.
- Palincsar, A. N. (1987). *Collaborating for collaborative learning of text comprehension.* Paper presented at the annual conference of the American Educational Research Association, Washington, D.C.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension fostering and comprehension-monitoring activities. *Cognition and Instruction, 2, 117-175.*
- Paris, S. G., Wasik, B. A., & Turner, J. C. (1991). The development of strategic readers. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp. 609-639). Mahwah, NJ: Erlbaum.
- RAND Reading Study Group (2002). *Reading for understanding: Toward an R & D program in reading comprehension.* Santa Monica, CA: RAND Corporation.
- 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). New York: Russell Sage Foundation.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*(4), 479-530.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86, 638-641.*
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Belmont: Wadsworth, Cengage Learning.

- Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytical and traditional reviews. *Educational Researcher*, 9(15), 5-11.
- 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.
- Slavin, R. E. (2017). Evidence-based reform in education. *Journal of Education for Students Placed at Risk*, 22(3), 178-184.
- Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational Research*, 79(4), 1391-1466.
- Slavin, R., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, 4, 370-380.
- Suggate, S. P. (2010). Why “what” we teach depends on “when”: Grade and reading intervention modality moderate effect size. *Developmental Psychology*, 46, 1556-1579.
- The Campbell Collaboration[C2](n.d.). *The Campbell Collaboration website*. Retrieved April 23, 2020 from the world wide web: <http://www.campbellcollaboration.org/>
- Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21(11), 1559-1573.
- Thompson, S. G., Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, 18, 2693-2708.
- 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(2), 130-149.
- Valentine, J. C., & McHugh, C. (2007). The effects of attrition on baseline comparability in randomized experiments in education: A meta-analysis. *Psychological Methods*, 12(3), 268-282.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package.

Journal of Statistical Software, 36(3), 1-48.

Viechtbauer, W. (2020). *metafor: Meta-analysis package for R*. Retrieved April 23, 2020 from <https://cran.r-project.org/web/packages/metafor/metafor.pdf>

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1, 112-125.

What Works Clearinghouse[WWC] (2010). *WWC intervention report: Reciprocal teaching(Adolescent literacy)*. Princeton, NJ: US Department of Education, Institute of Education Sciences. Retrieved April 23, 2020 from the world wide web: http://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_rec_teach_091410.pdf

What Works Clearinghouse[WWC] (n.d.). *What Works Clearinghouse website*. Retrieved April 23, 2020 from the world wide web: <http://ies.ed.gov/ncee/wwc/>

What Works Clearinghouse[WWC] (2017a). *What Works Clearinghouse: Standards handbook(version 4.0)*. Retrieved April 23, 2020 from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

What Works Clearinghouse[WWC] (2017b). *What Works Clearinghouse: Procedures handbook (Version 4)*. Washington, DC: Institute of Education Sciences, Department of Education. Retrieved April 23, 2020 from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

Local Evidence-based Research Synthesis: How Methodological Features Moderate the Effects of Reciprocal Instruction

Jin-Chang Hsieh* Min-Yu Chen**

Abstract

Systematic review and meta-analysis, one of the effective and widespread techniques in forming research-based and holistic evidences, was widely accepted by researchers. With the evolution of this technique, the approach in assessing quality of primary research has become an important topic in this area. However, except publication bias, research has seldom been conducted to investigate biases deriving from methodological features under the context of research synthesis. The purpose of this paper is to evaluate the moderating effects concerning experimental design, measurement tools, and instructor-as-researcher, while investigation the effects of reciprocal instruction on Chinese reading comprehension across students' grades. Systematic literature review processes were implemented to search, identify, and collect relevant studies. After few keywords were made in searching Taiwan literature database, a total of 2,437 articles were first retrieved, and followed a two-step literature screening process according to inclusion and exclusion rules, 43 articles were formally adopted. Random effect model was adopted while meta-analyzing findings of the included studies, an overall average effect size of +0.4756 and +0.3969 were obtained separately for immediate and follow-up effect. As to the analysis of moderator effects, the findings showed a significant lower average effect size separately for students in fifth-to-sixth grade, and above middle school ages in comparison to students in third-to-fourth grade while using multiple meta-regression with controlling the impact of experimental design. However, no significant moderating effects were found in measurement tools, and instructor-as-researcher. Finally, suggestions and implications were raised for further research.

以統合研究建立相互教學有效性的在地證據：研究方法特徵影響評估

**Key words: meta-analysis, methodological features, research quality
assessment, research synthesis**

* 1stAuthor: Nation Academy for Educational Research, Associate Research Fellow

** corresponding author: University of Taipei, Assistant Professor

Email : d970103@gmail.com