

壹、前言

因應科技的快速發展，數位學習的方式也歷經各種階段。包括早期的電腦輔助教學（Computer Assisted Instruction, CAI）、網際網路盛行後的遠距教學（distance learning）、加入行動載具輔助學習的行動學習（mobile learning）、到結合各種感測裝置的情境感知學習（context awareness learning）。然而，在各個數位學習的階段中，探討新工具對於教學效果的影響一直都是熱門議題（張菀珍、葉榮木，2014；Liu, Lin, & Paas, 2014; Sung, Chang, & Liu, 2016; Sung, Chang, & Yang, 2015）。

以行動學習為例，Wu 等（2012）回顧的 164 篇文獻中，即以「評估行動裝置輔助學習效果」為研究目的的文獻占最大宗（58%）。這些研究有些強調行動載具輔助學習者在教室中的創新教學（Kerawalla et al., 2007; Ketamo, 2003; Liu, Chou, Liu, & Yang, 2006; Zurita, Nussbaum, & Salinas, 2005），有些強調應用行動載具輔助學生探索博物館的展品（Hsi, 2003; Sung, Chang, Lee, & Yu, 2008），有些強調應用行動載具於戶外教學活動或實地觀察（Chen, Kao, & Sheu, 2005; Chu, Hwang, Huang, & Wu, 2008; Tan, Liu, & Chang, 2007），有些利用行動載具協助學習者在教室、實驗室以及戶外活動間進行無縫學習（seamless learning）（Liu & Hong, 2007）。由此可知，目前學界對於行動載具所能帶來的學習效果深感好奇與重視。

評估一項新的教學工具或教學方法的成效，最強而有力的方法即是實驗研究法。實驗研究可藉由操弄自變項，同時使其他變項保持恆定的情況下，來觀察依變項的變化結果。也因此，若期待實驗研究應用在教育上也可以證實預期的因果關係，就有賴對實驗研究的正確認識與實驗嚴謹度的精確講求。歷年來外界對很多教育研究的品質有不少批評，包括研究是否嚴謹、研究方法是否合宜以及研究結果的應用性等。Slavin（2002）就曾提醒教育研究在實驗法的運用與品質現況，已經落後醫療、農業、科學等領域一個世紀以上。此外，美國自 2002 年起一系列以實證本位的改革，也都在在強調實驗研究嚴謹度的重要性（Slavin, 2003; U.S. Department of Education, 2002）。而數位學習是個科技整合領域，許多研究者所受的實驗法訓練並不一致，且對實驗研究嚴謹度的要求亦不相同。因此，檢視行動學習實驗研究的品質，對於改進後續研究品質，並提供研究者參考有很大的價值。

Cheung 與 Slavin（2013）在回顧過去 30 年教育科技應用在數學學習的實驗研究時，即提醒讀者多數研究均面臨嚴重的方法學問題。主要的研究缺失包含實驗設計缺乏控制組、忽略實驗組與控制組起點能力不均等的問題、教學時間過於短暫、成效評量工具的品質不佳、採用的研究樣本數過少等。上述問題都可能對實驗結果的有效性形成嚴重威脅。Burston（2015）回顧過去 20 年行動裝置輔助語言學習的文獻也指出，儘管目前已累積上百篇的實驗研究成果，但半數以上的研究對於學習成果的測量並不可靠，研究者很少採取客觀與可信賴的測驗

工具，反而經常以教師主觀評估或學生的自我評估作為學習成效的工具。此外，許多研究也出現嚴重的實驗設計缺失，包含採用的樣本數過少、混淆變項控制不佳、缺乏控制組或對於控制組的描述相當模糊、未包含前測證據以說明實驗組與控制組在教學前的恆等性、不適當的統計分析程序等。

上述回顧性研究的結果，對於加深對實驗研究品質的重視具有很好的提醒作用：第一，實驗品質是累積正確證據的關鍵。數位學習與行動學習目前正處於快速發展的時期，我們是否更該在此階段檢視當前實驗研究方法的嚴謹度，以為未來此領域累積更多有意義的證據？Wu 等（2012）對於 2003-2010 年的行動學習研究回顧發現，86%的研究均具有正向的結果（positive outcome）。然而，缺乏研究方法嚴謹度的檢驗，我們如何確定行動學習的效果真有如此樂觀？這些研究中是否也包含一定程度有缺陷的實驗設計而誇大了實驗效果？第二，現有文獻缺乏對行動學習研究進行品質評估的回顧。過去有關行動學習的回顧性文獻大致包括探討筆記型電腦在學校學習的使用情形（Bebell & O'Dwyer, 2010; Fleischer, 2012; Penuel, 2006）、分析行動學習研究的發表趨勢（Frohberg, Goth, & Schwabe, 2009; Hung & Zhang, 2012; Hwang & Tsai, 2011; Wu et al., 2012）、分析行動載具對教學與學習的影響，例如對合作學習教學法的影響（Hsu & Ching, 2013）、分析行動載具對無縫學習的影響（Wong & Looi, 2011）。然而，目前尚缺乏以評估實驗研究品質為主題所進行的回顧性研究。

因此，本研究目的就是先介紹評估實驗嚴謹度的各種評估指標，以瞭解評閱或進行一份實驗研究時需要注意的重點。接著，透過搜尋與整理 ERIC 和 SSCI 資料庫中自 2003 年至 2013 年間，行動裝置輔助學習的實驗研究，以發現 10 年間數位學習實驗研究品質的概況與趨勢。

貳、實驗研究品質評估指標

本節統整適用於教育研究的兩份品質評估量表（WWC 與 DIAD）、相關學者的提醒，以及實驗方法教科書的建議作為介紹。其中，WWC 是美國教育部教育科學院（Institution of Educational Studies, IES）成立之「有效教育策略資料中心」（What Works Clearinghouse, WWC）所發布的教育研究評估標準（WWC, 2014）。該中心成立於 2002 年，主要任務為回顧各重要教育主題的實徵研究，以建立教育決策中可靠的證據來源。其發展的教育研究評估標準，用以辨識出哪些教育介入研究存有清楚的因果關係，得以提供該介入對學生學習是真正有效的證據（what works）。DIAD 則是由 Valentine 與 Cooper 於 2008 年所提出的研究設計與實施評估裝置（Study Design and Implementation Assessment Device, Study DIAD）。DIAD 由四個層面（內在效度、外在效度、統計結果效度、構念效度）與 32 個細節問題所組成，以系統性和全面性地評估個別實驗研究為其特色。有關 DIAD 詳細評估指標內容請參見 Valentine 與 Cooper（2008, p. 144）的研究。此外，近年來許多教育相關領域的研究皆使用 WWC 或 DIAD 的標