科學教育學刊
2011, 第十九卷第二期, 169-184

Chinese Journal of Science Education
2011, 19(2), 169-184

網路化學習歷程檔案學習者自評之信效度

張基成[1、*]　吳炳宏[2]

[1]國立臺灣師範大學　科技應用與人力資源發展系
[2]臺北市立松山家商　資料處理科

(投稿日期：民國99年8月7日，修訂日期：民國100年3月10日，接受日期：民國100年4月13日)

摘要：本研究探討網路化檔案學習者自評的信、效度。研究樣本爲某高職修習「計算機概論」課程的72位學生。學生藉由網路化檔案評量系統進行個人檔案的製作、觀摩與自評。結果顯示：(1)學習者兩次檔案自評結果具有高度的一致性；(2)學習者自評與教學者評量結果具高度一致性且無顯著差異；(3)學習者檔案自評分數與測驗成績結果具高度一致性，顯示網路化檔案自評可以反應學習成就。簡言之，網路化檔案自評兼具信、效度。

關鍵詞：自評、學習歷程檔案、網路化學習歷程檔案、網路化檔案評量

壹、研究背景

學習歷程檔案(learning portfolio)乃學習者經過一段時間，有目的、有組織地蒐集其學習過程、進步紀錄、心得感想、反省資料、學習成果等資料，可供檢視學習者學習成就。Russell與Butcher (1999)、Springfiled (2001)指出，傳統紙本式學習歷程檔案會因爲長時間進行資料蒐集以致數量增加，造成展示與觀摩上的不便。王等元(2000)、岳修平與王郁青(2000)不約而同的表示學習歷程檔案(learning portfolio)的網路化將是未來的應用趨勢；Yancey (2001)在從事網路化學習歷程檔案的研究之後亦提出這樣的看法。Chang (2001)、Chang與Tseng (2009a)。指出藉由網路上學習檔案的分享，不但能幫助教師瞭解學生的學習狀況，更有助於師生彼此間互動的增進；同時對學生而言，亦能瞭解同儕間的學習成果及過程。傳統紙本式學習歷程檔案在儲存空間及管理上的問題是可以藉由網路技術加以克服。

學習歷程檔案評量(portfolio assessment，以下簡稱檔案評量)乃依據特定目的有系統地將被評量者在學習過程與結果等相關紀

*通訊作者：張基成

錄集結成冊，藉以評定學習者的努力、成長、進步與成就情形的一種評量方式(Chang, 2008)。簡言之，檔案評量即學習歷程檔案用於評量學習者學習成就。網路化檔案評量(web-based portfolio assessment)指的是透過網路來輔助檔案評量的進行，評量者可於線上進行學習者自評、同儕互評、教學者評量或其他相關評量等活動(Chang & Tseng, 2009b)。Gadbury-Amyot等(2003)認爲，在課程當中讓學生進行自評活動可以提升大學生的學業能力，而這些自評活動可以透過檔案評量來進行。通常在網路化檔案評量過程中，常會要求學生針對自己的檔案在網路上做自評。更進一步來說，應是以自己的檔案內所呈現的學習過程與結果資料，對自己的學習在線上做出評論與反省，包括學習成果是否有達到預期目標、是否有進步、有哪些地方需要改善等。

目前，許多研究較常探討同儕互評議題，較少單獨討論學習者自評的議題。尤其是檔案評量過程中自評的信、效度問題，更少討論。Oskay, Schallies與Morgil (2008)綜合幾個檔案評量信、效度的研究結果，結論認爲檔案評量是一個恰當且可靠的評量方法，但可惜的是並未說明是否也包括學生自評。Knowles, Holton與Swanson (2005)宣稱，自評通常是有信度，但效度通常不足。Sadler與Good (2006)、Sung, Chang, Chiou與Hou (2005)研究顯示，學生自評與教師評量結果具一致性。由上述研究顯示，各研究有不同的結果，可能因研究對象的教育層級與課程屬性不同所致。

Bouzidi與Jaillet (2009)指出，電腦化評量的信、效度較一般紙本式評量的信、效度更難檢測，也更難(或更容易)達到，因爲有更多因素會影響(提高或減低)其信、效度的檢測與達成，譬如科技的支援、人在科技環境下的心智等。

由於前面所述研究大多是針對紙本式的檔案評量，因此在網路化學習歷程檔案環境下，自評的信度如何？自評的效度如何？如何進行網路化學習歷程檔案的自評？這些都是值得探討的議題。根據上述研究背景，本研究之目的爲探討網路化學習歷程檔案學習者自評的信、效度。相對應的研究問題與統計方法如表1。

貳、文獻探討

一、網路化檔案評量

岳修平與王郁青(2000)在其研究中表示，採用學習歷程檔案(learning portfolio)使學生對自己所學與評量具所有權感，並對自己的作品產生責任感，透過其內容的真實性讓學生檢驗他們的知識與技能。張麗麗(2002)指出透過檔案可以評量學習過程與結果，並提供認知與情意並重及尊重學生個別差異的檢測。學習歷程檔案是一種可以真實評估出學習者能力的評量方式，透過學習者自行組織檔案內容可以真正檢視

表1：研究目的、研究問題與統計方法對照表

研究目的	研究問題	進行方式	統計方法
學習者自評的信度	學習者兩次自評結果是否一致？	重測一致性	皮爾森(Pearson)積差相關
學習者自評的效度	學習者自評與教學者評量結果是否一致？是否有顯著差異？	外在效標關聯效度 (效標爲授課教師所評分數)	Pearson積差相關 *t* 檢定
	學習者自評分數是否與測驗成績一致？	外在效標關聯效度 (效標爲測驗成績)	Pearson積差相關

到學習者的學習成效。Hult (2001)表示學習歷程檔案可以讓學生自己明確發現及確認自己的問題。就後設認知理論中的自我監控(self-monitored)及自我調整(self-regulated)策略而言，具有自我監控能力的人會不斷的注意自己學習進展的情形，分辨較佳與較差之處，在學習中不斷地發現及澄清問題(王等元，2000)。Luca與McMahon (2006)將網路化學習檔案做爲線上學習的自我監控與評量工具。Chang與Tseng (2009b)的研究證實網路化檔案評量的使用對學生的自評能力有提升作用。綜上所述，在網路化檔案評量環境下，學習者線上自我檢視與評量扮演重要角色。

二、自評信效度

(一)自評信度

陳英豪與吳裕益(1990)指出信度是評量結果的一致性，即評量結果可不可靠。余民寧(2002)認爲信度是多次評量結果間的一致性或穩定性。陳嘉鴻(2002)指信度是相同的人在不同的時間，以相同的評量所得結果的一致性；如果兩次評量結果一致，表示結果具有穩定性、可靠性或可預測。王家玲(2002)指信度是評量結果具有一致性、穩定性的程度。對於信度的檢驗，一般多以內部一致性及重測信度來表示評量信度的高低。評量分數要具有意義，它必須具有相當程度的一致性與可靠性(張麗麗，2002)。

針對自評信度，研究者提出如下定義：是對於同一份評量採用複本再測，或者在不同時間與情境下施測結果的一致、穩定及可靠程度。Knowles等(2005)宣稱，自評通常是有信度的；因爲對任何人而言，不同時間自評結果的一致性並不難達成。對於信度的檢驗方法上，可以採用Pearson積差相關找出兩次評量結果的相關係數(余民寧，2002；張麗麗，2002；劉旨峰、林珊如、袁賢銘，2002；Derham & DiPerna, 2007; Gadbury-Amyot et al., 2003; Lenze, 2003; Lin, Liu, & Yuan, 2001; Liu, Lin, Chiu, & Yuan, 2001)。評量表的內部一致性可用Cronbach's α 係數來檢測(余民寧；Derham & DiPerna)。

(二)自評效度

陳英豪與吳裕益(1990)指出效度是一個評量能夠測量到它所欲測量的特質或功能的程度，即評量結果正不正確。余民寧(2002)表示效度是評量分數的有效程度，亦即評量能夠提供適切資料以作決策的程度。簡言之，效度是指評量分數的正確性，即一個評量能夠測量到他所想要測量的特質的程度。針對自評效度，研究者提出如下定義：是指一個評量能夠測量到他所欲測量的特質或功能的程度，亦可稱爲某一評量能提供適切資料以作成決策參考的程度。

Knowles等(2005)宣稱，自評結果通常不讓人信任，亦即其效度令人懷疑；因爲對任何人而言，自評容易因自我膨脹與誇大使其結果的正確性減低。通常若欲檢驗學習者自評的效度，可用授課教師的評分或教學助理的評分作爲校標以茲參照(Bouzidi & Jaillet, 2009; Cho, Schunn, & Wilson, 2006; Falchikov & Goldfinch, 2000; Sluijsmans, Dochy, & Moerkerke, 1999)。余民寧(2002)表示，若以外在效度來解釋評量的效度，需要一個外在效標以供參考，而授課教師的評分可以視爲一個外在效度指標。Sung等(2005)以大學生爲研究對象，結果顯示學生自評與教師評量結果具一致性。Sadler與Good (2006)以高中四個不同課程的學生爲研究對象，結果顯示學生自評與教師評量結果有高度一致性。由上述

顯示，自評效度的研究有不同的結果。

Gadbury-Amyot等(2003)的研究是以學生測驗分數當做外在效標，檢驗其與檔案評量分數的一致性。兩者一致性越高，表示檔案評量的效度越高。研究者認爲，學習者自評的效度可以運用外在效標的方式予以檢驗，教學者評分與學生測驗分數都可以是外在效標的來源。張麗麗(2002)表示，若教學者數量多一些或接受過評分訓練，外在效標會具有更高的效度。研究者認爲如果可以的話，再加上對做爲外在效標的教學者評分結果之間一致性的檢驗，外在效標的參考程度會更高。

參、研究方法與步驟

一、研究對象

參與實驗者爲臺灣北部某高級職業學校會計事務科二年級修習「計算機概論」課程的一個班級學生，共79人。刪除7位檔案內容不完整者(譬如學習目標未設定、作品未繳交、反思撰寫不全等)，剩72人之檔案作爲統計分析用；其中男生34名，女生38名。該門課學分數爲3學分，每週上課3小時。教學實驗爲期12週，教學內容爲「Word文書處理：版面設定及長文件編輯」(兩單元)。授課教師要求完成的作業，是此次實驗中學習者學習歷程檔案內蒐集的主要內容。實驗期間，學習者使用本研究自行開發的網路化檔案評量系統進行個人檔案的製作、觀摩與自評，教師則透過系統檢視學習者的檔案並評量其學習表現。

二、檔案評量標準發展

根據文獻歸納出初步的網路化檔案評量標準，經與授課教師討論及修正後製成量表。之後委請三位學者專家協助內容效度的建立，針對所給的意見進行修正。最後，再對修課學生說明並經討論與修正，以提升學生對檔案評量的動機與參與程度，這也是許多研究建議的做法(Bloxham & West, 2004; Price & O'Donovan, 2006; Prins, Sluijsmans, Kirschner, & Strijbos, 2005; Tsai, Lin, & Yuan, 2002)。評量標準分爲檔案製作、學習目標、作品、反思、態度、及其他等六個面向，共27項指標。量尺以李克特五等第爲理念，每項指標的滿分分數爲5分。但考慮實際評分上的方便性，每個指標依據被評量者所達到的績效水準給予1、1.5、2、2.5、3、3.5、4、4.5與5的點數。每項指標的點數間皆定義不同的績效表現水準，以利學生評分時能正確地給分。

在檔案製作的衡量上，包括內容完整性、內容恰當程度、內容豐富程度、內容的組織與呈現。在學習目標的衡量上，關注學習者個人學習目標的達成程度。作品面向細分爲正確性、恰當性、豐富性及完整性、難度、創意與創新等題項，同時也納入作品歷程紀錄。在反思面向中，考核的重點包含學習者對學習目標、作品、學習成就、學習態度等的反省思考。倘若學習者在這些項目中的反思明確指出自己的缺點及可改進的地方，則可獲較高評價；但如果反思內容呈現的是該位學習者學習上的負面缺失，如「過程中自己抱持偷懶的態度進行學習，下一次的學習要改正此缺點，態度要轉爲積極」。針對上述情況，本研究顧及學習者因擔心扣分而怯於寫出缺點，因而在評分時並不予扣分。在學習態度的衡量上，著重在網路上的情意表現。「其他」這個面向是檢視其他的事蹟或證據。

相較於傳統紙面式檔案評量，網路可增添傳統檔案所欠缺的互動性(Yancey, 2001)，因而此評量標準融入了一些網路的特性。在

檔案製作面向上，內容豐富程度指標考量了學習者上傳掃描圖檔、Word檔、PowerPoint檔、影音檔等狀況；在反思面向上，增加了對觀摩同儕表現的反思，針對他人回饋的反思的指標；在態度面向上，增加了線上的表現，譬如觀摩、欣賞、互評與回饋的表現、網路資源的分享、線上溝通討論、知識分享、意見交換、問題解決、線上互動的數量及品質等評量項目。

三、檔案評量標準項目分析

首先針對預試量表進行項目分析(item analysis)。在鑑別度(discrimination)部分，研究者將每一指標平均數分爲高(27.00%)、低分(27.00%)兩組後，再進行兩組平均數比較(*t* 檢定)。結果顯示，每一題 *t* 檢定的結果(*t* 值，即決斷值)皆達顯著水準，表示每一題的鑑別力足夠，不必刪除。在一致性部分，研究者將每一指標平均數與所有指標平均數進行Pearson積差相關檢定。結果顯示，每一題檢定結果(Pearson相關係數)皆達顯著水準，顯示每一指標之間的一致性足夠，不必刪除。

四、檔案評量標準效度檢驗

抽取共同因素面向方式是採用主成份分析法(Principal Component Analysis, PCA)，萃取特徵值大於1的因素面向。量表預試各面向的抽樣適切性量數(Kaiser-Meyer-Olkin, KMO)值皆大於0.50(表2)，可進行因素分析並建構效度。因考量各因素面向之間具一定程度之相關性，採斜交轉軸法(oblique rotation)；而Bartlett球形檢定之近似卡方分配皆達顯著，表示各指標之間有共同因素存在，適合進行因素分析。第一次因素分析結果顯示，態度面向內有一指標的因素負荷量小於0.30，將此指標刪除。第二次因素分析結果顯示，因素負荷量皆大於0.30，故保留所有指標。

最後共確定五個特徵值大於1的因素面向。各面向的解釋變異量皆大於45.00%，顯示各面向皆具一定之效度。作品面向的解釋變異量最大，態度面向最小。

五、檔案評量標準信度檢驗

如表3所示，預試與正式評量各面向的Cronbach's α 值皆大於 .70，屬於高信度。顯示各項指標之間具有高度的內部一致性，因此預試後保留各項指標。

六、網路化學習歷程檔案評量系統

此網路化檔案評量系統提供檔案內容項目按鈕及輸入資料的表單視窗，使用者只要點選檔案內容項目按鈕及根據表單上的說明輸入資料，即可完成個人檔案的製作。每人的檔案內容項目名稱與檔案內容格式相同，

表2：量表預試因素分析

面向	KMO值	解釋變異量(%)
檔案製作	0.67	56.26
學習目標	0.84	58.10
作品	0.89	66.71
反思	0.80	50.64
態度	0.69	45.41
其他	無	無
整體	0.86	72.09

註：「其他」面向僅有一題，不做因素分析。

表3：量表兩次施測Cronbach's α值

面向	預試	正式施測
檔案製作	0.91	0.91
學習目標	0.70	0.94
作品	0.77	0.93
反思	0.84	0.90
態度	0.73	0.74
其他	無	無
整體	0.92	0.97

註：「其他」面向僅有單獨一項，無法計算Cronbach's α值。

有利於評量者評分。

系統功能(如圖1左邊按鈕)主要有：(一)檔案指引區。(二)檔案製作區(僅供學生使用)：含基本資料、學習目標設定、作品繳交、自我反思、其他等。(三)檔案評量區：學生與教學者身份不同，若爲教師進入則爲教師評分；若爲學生登入則爲自評與同儕互評(內含檔案評量指標之等第勾選)。評分的同時，也可以瀏覽被評者的檔案內容。(四)檔案成績區：可查閱學習者自評分數、同儕互評分數、教學者評分分數，及觀摩優良作品等。(五)課程說明區：包括課程基本資料及教師的教案。

上方的工具列(如圖1上方按鈕)功能有：(一)系統公告：觀看系統最新公告。(二)討論專區：可進行課程相關及檔案製作的討論。(三)基本資料：可查看個人基本資料。(四)網頁式個人檔案總覽：顯示每個人的檔案內容。(五)檔案分項瀏覽：依據分類的內容項目顯示所有成員學習檔案中的該項目內容。(六)發送信件：可發送信件給每一位參與者。(七)單元切換：可切換不同的授課單元，讓每個單元呈現不同的檔案內容。

自評時，學習者點選系統左邊的「檔案評量區」，系統便會出現所有學習者姓名，如圖1。按自己名字旁邊的「自評」，在畫面中間上面會出現一整列的檔案內容選項(基本資料、學習目標、反思內容、單元作品、其他內容、「評分」、教師回饋、同儕回饋、參與情形等)。點選中間的「評分」按鈕，會另外出現一個網頁視窗評分表。此時，學習者可點選每一項評量指標旁的分數，及指標下面的文字方塊填寫回饋意見。評分完畢後，可以選擇預覽評分結果、重新評分、或者確定送出，如圖2。

七、教學實驗

實驗依課程單元分兩次進行，第一單元爲預試，而第二單元爲正式施測。預試之目的爲減低正式施測時的不確定性，提高學習者對評量準規的熟悉度與認知上的一致性，以提升正式施測的信、效度。如表4，參考Chang與Tseng (2009a, 2009b)設計的檔案評量活動，該活動完整且豐富並已經過數次的實驗，成效良好。實驗活動內容主要爲學習者課後使用網路化檔案評量系統進行個人檔案的製作(含設定學習目標、撰寫反思、上傳作品等)、觀摩與自我評量，而教學者透過該系統查看學習者的學習狀況與評量學習者的學習表現。上傳的作品包括初期、完成、修正等不同歷程的作品。

第一週主要是提供學習者自評、檔案評量標準及系統操作的訓練。由教師在課堂上對學生說明學習檔案的原理及如何製作等相關理論，並逐條解釋評量標準及評分時該注意的事項。之後，再由助教們示範如何使用

網路化檔案評量系統，並要求學生進入個人自評畫面，進行自評的練習。在學生練習自評的期間，由助教們個別輔導對評分及檔案評量標準有問題的學生。最後，由教師隨機選出兩位學生自評的結果，在課堂上與大家討論並說明應改善的地方。透過上述的訓練活動，協助學生進行自評。

爲檢測學習者自評效度，學習者課後在兩個單元結束後數日內對自己的學習歷程檔案進行2次的自我評量(共4次)，以確認再測效度。同一單元的兩次自評間隔一週，在這期間學生不得再修改自己的檔案內容(系統不接受)，以維護受評內容的相同性。同樣地，教學者(一位授課教師與三位線上助教)亦在兩個單元結束後數日內對每人的檔案做評量。

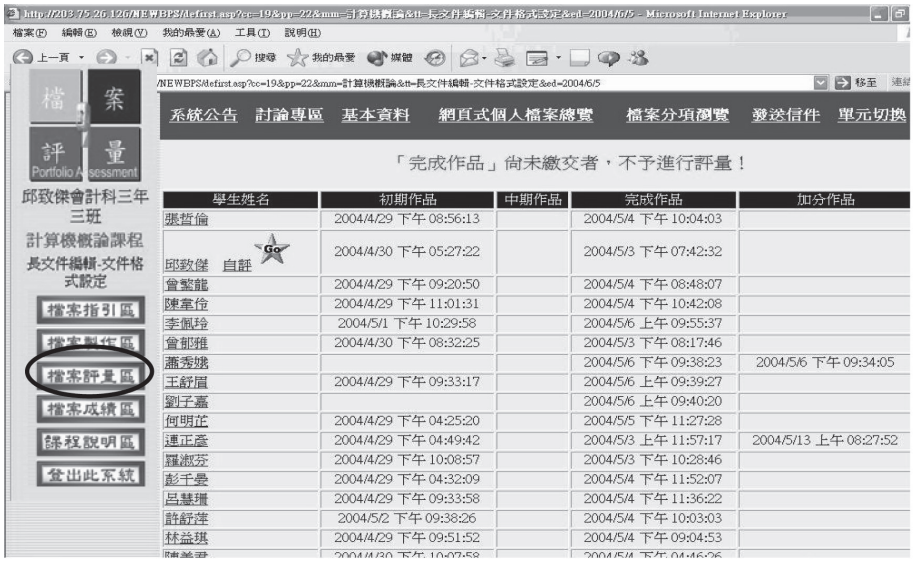

圖1：自評

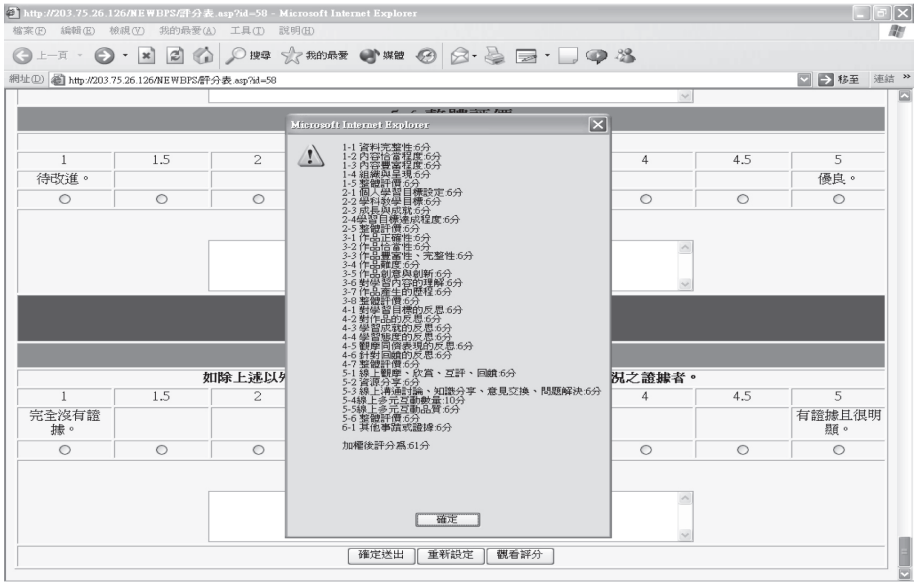

圖2：評量結果確認

表4：實驗活動

週次		活動內容
第一週 (預試階段)	教師	1.在課堂上說明如何製作學習檔案。 2.在課堂上解釋評量標準。 3.準備第一單元課程內容與教學活動。
	助教	1.系統測試。新增學生帳號。 2.在課堂上示範網路化檔案評量系統之使用。
	學生	1.學習如何製作學習檔案。 2.學習如何評分。 3.學習如何使用系統。
第二、三週	教師	1.進行第一單元課程教學。 2.上網檢視學生檔案製作狀況。
	助教	1.進行系統的管理與維護。 2.上網檢視學生檔案製作狀況。
	學生	1.課後使用檔案評量系統進行檔案製作。
第四週	教師	1.進行第一單元課程教學。 2.上網檢視學生檔案製作狀況。
	助教	1.上網檢視學生檔案製作狀況。
	學生	1.課後使用檔案評量系統進行檔案製作。 2.上傳初期作品與完成作品。
第五週	教師	1.上網檢視學生作品繳交狀況。
	助教	1.利用系統發出作品催繳通知。
	學生	1.上傳修正作品。
第六週	教師	1.進行教學者評分。
	助教	1.進行教學者評分。
	學生	1.進行自評(兩次)、同儕互評。
第七週 (正式施測)	教師	1.準備第二單元課程與教學活動。 2.在課堂上說明第一單元實施需要改進的事項。
	助教	1.整理各項評量數據，並檢驗信效度及修正評量標準。 2.說明修正的評量標準。
	學生	1.瞭解第一單元實施需要改進的事項。 2.瞭解修正的評量標準。
第八、九週	同第二、三週。	
第十週	同第四週。	
第十一週	同第五週。	
第十二週	同第六週。	
第十三週	助教	1.整理各項評量數據，並檢驗信效度。

註：「其他」面向僅有單獨一項，無法計算Cronbach's α值。

肆、結果與討論

一、學習者兩次自我評量結果是否一致？

兩次自評結果的Pearson積差相關如表5。結果顯示學習者兩次的檔案自評結果具有高度的相關性且達顯著，表示兩次自評結果具有高度一致性，此結果符合Zalatan (2001)指出「學習者對於自己的學習結果評鑑會用心投入，並且發展出穩定的反省思考能力」。各面向中，以檔案製作面向最高，

學習目標面向居次；反思面向最低，顯示反思面向的評審較不容易達成一致。Gadbury-Amyot等(2003)的研究對紙本式檔案評量效度做檢定，兩位教師評分者之間相關係數介於0.28至0.60之間，較本研究結果低。Gadbury-Amyot等的研究評分者是教師，但本研究是學生自評，評分者的身份不同。

二、學習者自我評量與教學者評量結果是否一致或有顯著差異

學習者自評與教學者評量結果的Pearson積差相關及 *t* 考驗如表6。結果顯示在作品面向及整體上，兩者具有高度的相關且達顯著，表示該兩種評量方式在作品面向及整體上皆具高度一致性。但在其他的四個面向上，兩者皆未達顯著相關，顯示這兩種評量方式在這四個面向上的一致性低。在假設教學者評量是具有效度而做爲效標的前提之下，整體而言，學習者自評的評量方式是具有效度的。

自評與教學者評量在檔案製作、作品、態度、整體面向上皆未達顯著差異，顯示兩

表5：學習者兩次自我評量結果之Pearson積差相關

檔案評量	相關係數(顯著性)
檔案製作	0.99 (0.003***)
學習目標	0.93 (0.022**)
作品	0.87 (0.011**)
反思	0.78 (0.023**)
態度	0.85 (0.033**)
其他	0.82 (0.028**)
整體	0.81 (0.001***)

***p* < .010, ****p* < .001

表6：學習者自評與教學者評量結果之Pearson積差相關及 *t* 考驗

面向	相關係數(顯著性)	評分方式	平均數	標準差	*t* 值(顯著性)	效果量
檔案製作	0.19	自評	4.02	0.57	-0.309	0.061
	(0.270)	教學者評量	4.05	0.40	(0.759)	
學習目標	0.28	自評	3.93	0.53	2.860	0.583
	(0.102)	教學者評量	3.69	0.24	(0.007**)	
作品	0.49	自評	3.70	0.44	-0.775	0.138
	(0.002**)	教學者評量	3.75	0.26	(0.443)	
反思	0.32	自評	3.82	0.52	-3.289	0.672
	(0.061)	教學者評量	4.11	0.32	(0.002**)	
態度	-0.07	自評	3.62	0.54	1.555	0.379
	(0.685)	教學者評量	3.44	0.40	(0.129)	
其他	0.18	自評	3.54	0.53	0.867	0.105
	(0.373)	教學者評量	3.59	0.42	(0.065)	
整體	0.83	自評	3.82	0.46	1.058	0.214
	(0.001**)	教學者評量	3.74	0.26	(0.297)	

註：1.「其他」面向僅有單獨一項，無法計算Pearson積差相關係數。

2.學習者兩次自評之平均後再與教學者評量的平均做Pearson積差相關的檢定。

***p* < .010

種評量方法的結果在這四個面向上的落差不大。但在學習目標、反思兩個面向上，自評與教學者評量的結果有顯著差異，顯示兩種評量結果在這兩個面向上的落差很大。以差異大小來看(效果量)，反思面向差異最大，其次爲學習目標，檔案製作面向最小。就整體平均分數而言，學習者自評略高於教學者評量，顯示學習者自評較爲寬鬆，教學者評量較爲嚴格。以各面向平均數而言，除了學習目標、態度之外，其餘三個面向皆爲教學者評量高於自評分數。顯示在檔案製作、作品、反思這三個面向上，教學者評量反而較爲寬鬆，其原因需要進一步探討。

三、各項指標是否足以檢測出學習成就(即檔案分數是否與測驗成績一致)

學習者檔案自評分數與其測驗成績的Pearson積差相關如表7。測驗內容主要爲該單元的學科知識，與檔案評量標準無重覆地方，以提高統計檢驗的可信度。如果兩者有統計上的顯著相關，表示學習者的檔案自評分數可以反應學習成就。結果顯示，學習者的檔案自評分數與其測驗成績具高度一致性且達顯著，個別面向亦具高度一致性且達顯著。顯示檔案自評結果可以反應學習成就，符合Gelinas (1998)提列的「檔案評量分數與被評者的學習成果表現有正相關」。其中以作品面向的相關性最高，檔案製作面向居次，其他面向最低且未達顯著。顯示作品最能檢測出學習成就。

四、綜合討論

就學習者自評結果的信度而言，表5學習者兩次自評結果之Pearson積差相關表顯示學習者兩次自評結果具高度一致性。換言之，學習者自評的信度足夠，這與Knowles等(2005)的看法吻合。就自評結果的效度而言，表6學習者自評與教學者評量結果之Pearson積差相關及 *t* 考驗表顯示學習者自評與教學者評量的結果之間具有高度的一致性且無差異。表7檔案分數與測驗成績之Pearson積差相關表顯示，學習者的檔案自評分數與其測驗成績結果具有高度的一致性。上述兩項結果顯示，學習者自評的效度足夠。這與Sung等(2005)、Sadler與Good (2006)的研究結果吻合，但與Knowles等的看法不一致。

綜上所述，學習者自評是一種同時具有信度與效度的檔案評量方式。各面向的信、效度高低順序並不一致；整體觀之，以作品面向較高，態度與其他面向較低。不同類的學習者自評相關係數之比較如表8所示。

至目前爲止，已有一些透過紙本式與網路化檔案進行自評信、效度的研究。儘管這些研究的結果不一，大部分是信、效度

表7：檔案分數與測驗成績之Pearson積差相關

面向	相關係數	顯著性
檔案製作	0.67	0.000***
學習目標	0.64	0.003**
作品	0.68	0.000***
反思	0.66	0.000***
態度	0.61	0.006***
其他	0.24	0.008*
整體	0.71	0.023**

* $p < .050$, ** $p < .010$, *** $p < .001$

表8：學習者自評相關係數比較表

面向	兩次自評	自評與教學者評量	與學習成就相關性
檔案製作	1***	4	2***
學習目標	2***	3	4***
作品	3***	1**	1***
反思	6**	2	3***
態度	4**	6	5**
其他	5**	5	6*

註：1表示最高，依序爲2, 3, 4, 5，最低爲6。
* $p < .050$, ** $p < .010$, *** $p < .001$

足夠的，少部分證明信、效度不足。無論信、效度是足夠或不足夠，都各有其支持的理論(Cho et al., 2006)。評量信、效度的檢驗最好是經過大樣本、多課程、與不同教育層級研究對象的實證研究，所獲結果才完整且可靠(Bouzidi & Jaillet, 2009)。Sulzen, Young與Hannifin (2008)的研究顯示，數位化檔案評量的信度不足，效度足夠；而增加評分者人數是可以增加評分信度。許多研究提到，線上評量份數多時較耗時費力，容易影響評量的結果(Ballantyne, Hughes, & Mylonas, 2002; Davies, 2000; Lin et al., 2001; Miller, 2003; Sung et al., 2005; Sung, Lin, Lee, & Chang, 2003; Topping, Smith, Swanson, & Elliot, 2000; Tsai & Liang, 2009; Tsai et al., 2002; Wen & Tsai, 2006; Yang & Tsai, 2010)。但Sung等(2003)及Sung等(2005)研究發現，學生線上評量的次數多一些，則評量結果與教師評量結果的差異會漸漸減少。這個發現顯示，學生經過多次評量的歷練其評分能力提升，使得評量效度也增加。因此，學生自評應兼顧次數與負擔，太多與太少都可能影響評量結果。

Falchikov與Goldfinch (2000)的研究顯示，在設計良好的評量標準下，整體上來看，學生評量與教師評量結果頗爲接近；但從不同評量標準的面向來看，學生評量與教師評量結果不一定在每一個面向都接近。這與本研究結果十分吻合。然而Bouzidi與Jaillet (2009)的研究結果顯示，從不同評量標準的面向來看，學生評量與教師評量結果具一致性是可能的，但本研究並未達到此結果。但無論如何，要讓網路化檔案自我評量具有足夠信、效度是可以做到的。不管網路化學生自評檔案的信、效度如何，都不宜讓它來取代教師評量或網路化教師評量檔案。

伍、結論與建議

高職學生兩次的檔案自評結果具有高度的一致性，因此本研究認爲網路化檔案自評是具有足夠信度的評量方式。郭怡雯(2001)認爲小學生檔案自評是不具信度，但本研究的高職學習者自評是具信度。訓練有素的評分者、對檔案有充分的瞭解可以提升評分的信度(Gadbury-Amyot, 2003; Oskay et al., 2008)。Derham與DiPerna (2007)、Sulzen等(2008)的研究建議，給評分者足夠的訓練可以提升評分的信度。本研究認爲年齡所造成的對檔案認知能力差異或心理成熟度可能會對自評信度有所影響。學生自評的信、效度除了會受到檔案評量標準這個工具的信、效度的影響之外，其他如學生人格成熟度、學生對檔案評量標準的熟悉度、學生對給分的判斷、學生評分的認真程度、學生對網路科技的熟悉程度等，都有可能影響自評的信、效度。因此，提高評量標準的品質與提供足

夠的評量訓練應可減低上述因素的影響。就面向而言，學習者兩次檔案自評結果的一致性以檔案製作面向最高，顯示檔案製作面向的評審較容易達成一致；反思面向最低，顯示反思面向的評審較不容易達成一致。

整體而言，學習者自評與教學者評量結果具有一致性且無顯著差異，因此在教學者評量是具有效度的前提之下，高職學生自評是具有足夠效度的評量方式。但就各個評量面向而言，部分評量面向的效度不足。在檔案製作、學習目標、反思、態度上，自評與教學者評量一致性低。在學習目標、反思面向上，自評與教學者評量有顯著差異。Bouzidi與Jaillet (2009)的研究建議，評量前應注意計分方式、評量標準的品質、標準的清楚度、科技的支援、評量範例是否足夠。具體明確的評量標準是可以提升評分的效度(Gadbury-Amyot, 2003; Oskay et al., 2008)。本研究認爲學習者能運用較爲具體的評量標準(作品)，但對於較爲抽象評量標準(學習目標、反思)的掌握仍須加強。學習者的檔案自評分數與其測驗成績具高度一致性且達顯著，顯示網路化檔案自評可以檢測出學習成就。上述兩項結果顯示，網路化檔案自評是具有足夠效度的評量方式。

由於受限於參與實驗的學習者校內學期進度、畢業考試及升學考量等因素，實驗時間不足三個月，無法進行整學期甚至更長時間的實驗研究。由於學習歷程檔案運用於評量上的觀察應以長時間來檢視較佳，因此建議後續的研究可將整個實施時程再予拉長，以增加檔案評量的真實性。同樣地，受限於實驗時間，只採用兩份作品分別進行網路化檔案評量的預試與正式實驗。建議後續研究可以增加作品數量，如此一來每一位學習者可以呈現更多的歷程資料，供評量方式的信、效度檢視。大樣本評量的信度會較高(Gadbury-Amyot, 2003)。本研究的參與對象屬小樣本，在統計結果的推估上較易有誤差，建議未來研究能增加樣本人數，以提升研究結果的信、效度。

在反思面向中，考核的重點包含學習者對學習目標、作品、學習成就、學習態度等的反省思考。倘若某位學習者反省內容爲「自己是以虛應故事的態度來製作作品，下一次必須更爲積極，不可以敷衍了事」。根據評量標準，學習者在反思中明確指出自己的缺點及需改進的地方，所以可獲致較高評價；但是反思內容呈現的是該位學習者學習上的負面缺失，是否應予適當扣分？或者可另建立哪些指標或機制使評分更加合理，此爲後續研究可進一步思考的議題。

參考文獻

1. 王家玲(2002)。**甄選工具之效度驗證與運用——以某高科技公司為例**。未出版之碩士論文，國立中央大學人力資源管理研究所，桃園縣。
2. 王等元(2000)。「歷程檔案」概念及其在學習者支持系統之意涵。**社會教育學刊**，**29**，249-274。
3. 余民寧(2002)。**教育測驗與評量——成就測驗與教學評量(第二版)**。臺北市：心理。
4. 岳修平、王郁青(2000)。電子化學習歷程檔案實施之態度研究。**教育心理學報**，**31**(2)，65-84。
5. 張麗麗(2002)。檔案評量信度與效度的分析——以國小寫作檔案爲例。**教育與心理研究**，**25**(1)，1-34。
6. 郭怡雯(2001)。**融入課程的數學檔案評量——以四年級為例**(國科會大專生參與專題研究計畫報告NSC89-2815-C-024-011R-H)。臺北市：行政院國家科

學委員會。
7. 陳英豪、吳裕益(1990)。**測驗與評量**。高雄市：復文書局。
8. 陳嘉鴻(2002)。**高職資訊性向量表信效度與結果運用之研究**。未出版之碩士論文，國立高雄師範大學資訊教育研究所，高雄市。
9. 劉旨峰、林珊如、袁賢銘(2002，5月)。網路化學習歷程檔案評量系統、學習成效分析及未來展望。發表於**第三屆電子化企業經營管理理論暨實務研討會**。彰化縣：大葉大學。
10. Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education*, *27*(5), 427-441.
11. Bloxham, S., & West, A. (2004). Understanding the rules of the game: Marking peer assessment as a medium for developing students' conceptions of assessment. *Assessment & Evaluation in Higher Education*, *29*(6), 721-733.
12. Bouzidi, L., & Jaillet, A. (2009). Can online peer assessment be trusted? *Educational Technology & Society*, *12*(4), 257-268.
13. Chang, C.-C. (2001). A study on the evaluation and effectiveness analysis of web-based learning portfolio (WBLP). *British Journal of Educational Technology*, *32*(4), 435-458.
14. Chang, C.-C. (2008). Enhancing self-perceived effects using web-based portfolio assessment. *Computers in Human Behavior*, *24*(4), 1753-1771.
15. Chang, C.-C., & Tseng, K.-H. (2009a). Using a web-based portfolio assessment system to elevate project-based learning performances. *Interactive Learning Environments*, *16*(2), 1-20.
16. Chang, C.-C., & Tseng, K.-H. (2009b). Use and performances of web-based portfolio assessment. *British Journal of Educational Technology*, *40*(2), 358-370.
17. Cho, K.-S., Schunn, C. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, *98*(4), 891-901.
18. Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Training International*, *37*(4), 346-355.
19. Derham, C., & DiPerna, J. (2007). Digital professional portfolios of preservice teaching: An initial study of score reliability and validity. *Journal of Technology and Teacher Education*, *15*(3), 363-381.
20. Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287-322.
21. Gadbury-Amyot, C. C. (2003). *Validity and reliability of portfolio assessment of competency in a baccalaureate dental hygiene program*. Unpublished doctoral dissertation, University of Missouri, Kansas, MO.
22. Gadbury-Amyot, C, C., Kim, J., Palm, R. L., Mills, G. E., Noble, E., & Overman, P. R. (2003). Validity and reliability of portfolio assessment of competency in a

baccalaureate dental hygiene program. *Journal of Dental Education*, *67*(9), 991-1002.

23. Gelinas, A.-M. (1998). *Issue of reliability and validity in using portfolio assessment to measure foreign language teacher performance*. Unpublished doctoral dissertation, Ohio State University, Columbus, OH.
24. Hult, C. (2001). Using on-line portfolios to assess English majors at Utah State University. In B. L. Cambridge, S. Kahn, D. P. Tompkins, & K. B. Yancey (Eds.), *Electronic portfolios: Emerging practice in student, faculty, and institutional learning* (pp. 60-70). Washington, DC: American Association for Higher Education Press.
25. Knowles, M. S., Holton, E. F., III, & Swanson, R. A. (2005). *The adult learner: The definitive classic in adult education and human resource development* (6th ed.). Burlington, MA: Elsevier.
26. Lenze, J. (2004). Inter-rater reliability in the evaluation of electronic portfolios: A survey of empirical research results. In R. Ferdig, C. Crawford, R. Carlsen, N. E. Davis, J. Price, R. Weber, & D. A. Willis (Eds.), *Proceedings of society for information technology & teacher education international conference 2004* (pp. 164-169). Chesapeake, VA: AACE.
27. Lin, S. S. J., Liu, E. Z.-F., Yuan, S.-M. (2001). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, *17*(4), 420-432.
28. Liu, E. Z.-F., Lin, S. S. J., Chiu, C.-H., & Yuan, S.-M. (2001). Web-based peer review: The learner as both adapter and reviewer. *IEEE Transactions on Education*, *44*(3), 246-251.
29. Luca, J., & McMahon, M. (2006). Developing multidisciplinary teams through self-assessment, supported with online tools. In E. Pearson & P. Bohman (Eds.), *Proceedings of ed-media* (pp. 1855-1860). Norfolk, VA: AACE.
30. Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment & Evaluation in Higher Education*, *28*(4), 383-394.
31. Oskay, O. O., Schallies, M., & Morgil, I. (2008). Reliability of portfolio: A closer look at findings from recent publications. *Hacettepe University Journal of Education*, *35*, 263-272.
32. Price, M., & O'Donovan, B. (2006). Improving performance through enhancing student understanding of criteria and feedback. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 100-109). London: Routledge.
33. Prins, F. J., Sluijsmans, D. M. A., Kirschner, P. A., & Strijbos, J.-W. (2005). Formative peer assessment in a CSCL environment: A case study. *Assessment & Evaluation in Higher Education*, *30*(4), 417-444.
34. Russell, J. D., & Butcher, C. (1999). Using portfolios in education technology courses. *Journal of Technology and Teacher Education*, *7*(4), 279-290.
35. Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, *11*(1), 1-31.

36. Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer-, and co-assessment. *Learning Environments Research*, *1*(3), 293-319.
37. Springfiled, E. (2001). Comparing electronic and paper portfolios. In B. L. Cambridge, S. Kahn, D. P. Tompkins, & K. B. Yancey (Eds.), *Electronic portfolios: Emerging practice in student, faculty, and institutional learning* (pp. 76-82). Washington, DC: American Association for Higher Education Press.
38. Sulzen, J., Young, M. F., & Hannifin, R. (2008). Reliability and validity of an ecologically-grounded student teacher electronic portfolio rubric. In K. McFerrin, R. Weber, R. Carlsen, & D. A. Willis (Eds.), *Proceedings of society for information technology & teacher education international conference 2008* (pp. 153-159). Chesapeake, VA: AACE.
39. Sung, Y.-T., Chang, K.-E., Chiou, S.-K., & Hou, H.-T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers & Education*, *45*(2), 187-202.
40. Sung, Y.-T., Lin, C.-S., Lee, C.-L., & Chang, K.-E. (2003). Evaluating proposals for experiments: An application of web-based self-assessment and peer-assessment. *Teaching of Psychology*, *30*, 331-334.
41. Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment & Evaluation in Higher Education*, *25*(2), 149-169.
42. Tsai, C.-C., & Liang, J.-C. (2009). The development of science activities via on-line peer assessment: The role of scientific epistemological views. *Instructional Science*, *37*(3), 293-310.
43. Tsai, C.-C., Lin, S. S. J., & Yuan, S.-M. (2002). Developing science activities through a networked peer assessment system. *Computers & Education*, *38*(1-3), 241-252.
44. Wen, M. L., & Tsai, C.-C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education*, *51*(1), 27-44.
45. Yancey, K. B. (2001). General patterns and the future. In B. L. Cambridge, S. Kahn, D. P. Tompkins, & K. B. Yancey (Eds.), *Electronic portfolios: Emerging practice in student, faculty, and institutional learning* (pp. 83-87). Washington, DC: American Association for Higher Education Press.
46. Yang, Y.-F., & Tsai, C.-C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction*, *20*(1), 72-83.
47. Zalatan, K. A. (2001). Electronic portfolios in a management major curriculum. In B. L. Cambridge, S. Kahn, D. P. Tompkins, & K. B. Yancey (Eds.), *Electronic portfolios: Emerging practice in student, faculty, and institutional learning* (pp. 44-52). Washington, DC: American Association for Higher Education Press.

科學教育學刊
2011, 第十九卷第二期, 169-184

Chinese Journal of Science Education
2011, 19(2), 169-184

Reliability and Validity of Learner Web-Based Portfolio Self-Assessment

Chi-Cheng Chang[1,*] and Bing-Hong Wu[2]

[1]Department of Technology Application and Human Resource Development, National Taiwan Normal University

[2]Department of Data Processing, Taipei City Song-Shan High School

Abstract

This study examines the reliability and validity of Web-based portfolio self-assessment. The research sample consisted of seventy-two students taking an "Introduction to Computers" course in the second grade of a vocational high school. The students used a Web-based portfolio assessment system to create their portfolio, emulate peers' portfolios, and perform self-assessment. The results demonstrate that there exists a high consistency between the two self-assessments. There exists a high consistency and no significant difference between student self-assessment and teacher assessment. The consistency between peer-assessment and test score was high, implying that the Web-based portfolio self-assessment may reflect learning achievement. In short, Web-based portfolio self-assessment has adequate reliable and validity.

Key words: Self-Assessment, Portfolio, Web-Based Portfolio, Web-Based Portfolio Assessment

* Corresponding author: Chi-Cheng Chang