科學教育學刊
2012, 第二十卷第五期, 393-412

Chinese Journal of Science Education
2012, 20(5), 393-412

網路化檔案評量環境下教學者評量之信度與效度

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學系
[2]臺北市立松山家商　資料處理科

(投稿日期：民國100年8月7日，修訂日期：民國101年5月28日，接受日期：民國101年6月4日)

摘要：本研究探討網路化檔案評量環境下，教學者評量的信度與效度。研究對象為某高職二年級修習「計算機概論」課程的兩個班級學生，共72人。學習者使用網路化檔案評量系統進行個人檔案的製作、觀摩、自評與同儕互評，教師則透過系統檢視學習者的檔案並評量其學習表現。結果顯示：(1)網路化檔案評量準規具足夠之信度；(2)網路化檔案教學者評量有很高的評分者間信度及評分者內部信度，是可以展示足夠信度的評量方式；(3)網路化檔案評量準規具足夠之效度；(4)教學者檔案評量分數與學生的測驗成績具有高度的一致性，顯示網路化檔案教學者評量方法具良好效度。建議未來研究可針對簡報軟體、文書處理軟體、網頁製作軟體等建置的自我建構式數位化學習檔案來製定可行的檔案評量準規並檢驗其信、效度。

關鍵詞：信度、效度、電子化學習歷程檔案、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)的使用已有其普遍性，對於學習過程的瞭解及學習品質的促進都有它正面的功效，同時亦可達到傳統評量方式所不能達到的目地(Zalatan, 2001)。傳統紙本式的學習歷程檔案在資料的查詢、修改、儲存、管理較費時耗力，尤其無法紀錄所有類型的資料，如聲音、影片等(Chang, 2008)。面對紀錄與儲存的問題，Chang與Tseng (2011)指出雖然數位化的檔案型式減少了傳統紙本式學習歷程檔案儲存上的問題，但是仍需要許多周邊的儲存設備，如硬碟、軟碟及光碟等，而這一個缺點可透過網路(web)來解決。網路科技對學習歷程檔案評量(簡稱檔案評量；portfolio assessment)的一項重大貢獻是可提供一個絕佳的評量機制，學習者透過網路可以遠距、即時觀摩與評量他人的作品，這一點是傳統紙本式檔案評量環境很難做到的。

學習歷程檔案已從紙本式檔案發展到與網路結合。網路化學習歷程檔案系統的使用

*通訊作者：張基成

03-張基成.indd 393　2012/10/29 上午 09:41:56

已趨普遍，對教師教學與學生學習亦有助益(Chang & Tseng, 2009, 2011)。然部分的系統並未提供評量準規；有者，其評量準規亦不完備或不嚴謹。對一個作為網路化檔案「評量」的系統而言，並不是一個恰當的現象。檔案評量準規本身除了是評分的工具，也有助於學生對檔案的建構(Oskay, Schallies, & Morgil, 2008)。Lynch與Purnawarman (2004)的研究結果顯示，美國有35%的大學師資培育機構有使用數位化檔案評量(e-portfolio assessment)，但大部分都缺乏評量準規。一些研究如Johnson, Fisher, Willeke與McDaniel (2003)、Rogers (2003)，已呼籲應建立明確的網路化檔案評量準規，這些準規確有迫切性！檔案評量時必須有良好信、效度的準規，評量才容易進行且評量的結果才值得信賴。目前見到的一些檔案評量準規多屬傳統紙本式檔案，較少見到針對網路特性所產生的網路化檔案評量準規。雖然兩種檔案的評量準規大同小異，但在製作上的評量必有不同，譬如多媒體製作、資料呈現、版面等。另外，目前見到的檔案評量準規有些圍繞在對檔案內容項目(entry)作檢視，或僅評估檔案製作的優劣，較少與學生的學習表現連結(Bonita & Duquence, 1999; Danielson & Abrutyn, 1997; Johnson et al.; Skawinski & Thibodeau, 2003)。如此，便失去評量的根本目的——為了要測得學生的學習成就！此外，對於評量準規的信、效度檢驗的實證研究較少。倘若要探討檔案評量，應該先建立能測出學生學習成就的準規，並探討評量結果的信、效度才具有意義！因此，建立有意義且嚴謹的網路化檔案評量準規並驗證其信、效度是此研究動機之一。

在檔案評量信度方面，Chang (2001)表示在實施學習歷程檔案評量時，評量準則並不易掌握。此外，盧貞穎(2001)指出檔案評量因為缺乏一致性的評分準規、評語敘述不夠精確、學生互評的能力不足、學生作品是否為親自完成等因素，常導致評量的信度受到質疑。岳修平與王郁青(2000)指出評量結果的信度易受質疑是檔案評量限制之一。在檔案評量效度方面，張麗麗(2002)表示直至目前為止，探討檔案效度方面的研究在質與量方面仍不足，多數是偏向以學生及教師自我陳述的方式來探討檔案在教學與學習上的效用，較缺乏以統計方法或系統性方式分析檔案評量效度的研究。Diperna與Derham (2007)分析了一些檔案信、效度的研究，發現大多是針對紙本式檔案，針對數位化檔案的幾乎沒有。因為兩種檔案信、效度的檢驗方法是類似的(Brickley, Schwartz, & Suen, 2000; Lenze, 2004)，所以紙本式檔案信、效度的研究結果仍可以做為數位化檔案信、效度研究的參考。Lynch與Purnawarman (2004)的研究結果顯示，在36個使用數位化檔案評量(e-portfolio assessment)的大學系所當中，25個沒有具體明確的評量準規，只有2個系所使用評量準規來提升檔案評量的信、效度。建立嚴謹的檔案評量準規使其能代表學生學習成就，固然能提升數位化檔案評量實施的信、效度，但它的信、效度究竟如何？是否足夠？綜上所述，評量準規的訂定及其信、效度的檢驗實為當務之急。

通常檔案評量包括教師評量、專家評量、學生自評、學生同儕互評。本研究雖同時包括教師評量、學生自評、同儕互評，但因受篇幅所限，本文僅對教師評量做分析。另則，聚焦於教師評量，對此議題可做單方面深度探討與分析。根據上述研究背景，本研究之目的為探討：一、網路化檔案教學者評量的信度；二、網路化檔案教學者評量的效度。研究問題為：一、不同教學者之間(授課教師、線上助教)的檔案評量結果是否

一致？二、每一教學者(授課教師、線上助教)檔案評量結果(對不同份)是否一致？三、各項準規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)？

貳、文獻探討

一、檔案評量

Alter與Spandle (1992)定義學習歷程檔案是一有目的蒐集學生一段時期之作品的評量工具，檔案中需包括讓學生參與選擇的內容、共同訂定的準規及自我反省的證據。學習歷程檔案可以說是個人作品的集錦(Bloom & Bacon, 1995)，同時也包含反省、選擇、評估的過程及其產物(Winsor & Ellefson, 1995)。張美玉(2001)認為學習歷程檔案要包含的內容項目應視評量之目的而決定；待評量目的決定之後，再決定評量項目的內容。檔案的大小與內容項目的多寡和評量的信、效度有密切的關係(郭生玉，2004)。由此可見，檔案中的內容項目應配合教學者的教學目的、教學上的需要、評量的目的與方便性來訂定相關要求。基於檔案評量之目的，研究者認為有一個必須要注意的原則是這些內容項目必須要能表現出學習者的學習成就，而且每一個學習者的檔案必須具備共同的及最低要求的內容項目，以方便於評量！

由於網路科技的進步，許多歐美國家的各級學校已將網路化學習檔案(web-based learning portfolio)融入教學與評量中，作為真實性評量與教學革新的工具(Batson, 2002; Lorenzo & Ittelson, 2005; Singh & Ritzhaupt, 2006)。Barker (2006)及e-Portfolio Consortium (2003)就都提到網路化學習檔案主要的功能之一是作為學生學習的線上評量工具，此時稱為網路化檔案評量(web-based portfolio assessment)。Rose-Hulman Institute of Technology使用自行發展的RosE Portfolio System (REPS)來評量學生的學習表現(Burnett & Williams, 2009)。Chang與Tseng (2009)的研究證實網路化檔案評量的使用對學生的設定學習目標、作品、反思、自我評估、同儕互評、同儕互動、蒐集與整理資料、持續改善、解決問題的表現或能力有提升作用。Dennis與Hardy (2006)也認為，網路化檔案評量的實施除了可讓學生展示學習成果，也方便教師線上檢視學生的表現及提供回饋。C. X. Wang (2009)研究數位化學習檔案的小組合作建構，對學生的學習是有幫助的。綜上所述，網路化檔案評量的使用對教師的教學及學生的學習，都有幫助。

二、檔案評量準規

C. I. Wang, Peng與Lu (2006)運用檔案評量準規針對學生做跨文化學習與溝通能力的評量。根據S. Wang (2009)，數位化檔案評量應有準規與回饋。但可惜的是，這些研究並未對評量準規做進一步的探討。Skawinski與Thibodeau (2003)在進行檔案評量時，主要分三個準規：(一)一般性準規：反思及進步證據、(二)特定評鑑準規：學習績效要求的重點、(三)綜合性的準規：整體品質；每個準規以四個等級來表示程度差別。在Johnson等(2003)所進行的檔案評量研究中，評分的準規可分為整體與分項。在整體部分中分為四個等級，分別為精通的、發展中、略顯徵兆及未合格。在分項的部分分為五個等級，以五等第量表方式呈現，在兩端(1與5分)及中點(3分)處分別敘寫績效水準；並且在分項評量準規下面提供所謂的徵兆舉例(look-fors)，列舉相關可見的行為證據與作品成果，不僅可作為被評量者蒐集資料的指引，同時更可作為評量打分數的依據。

Johnson等更強調，此方式將有助於評分者內部信度的達成。

鑒於有些績效水準並不是可以用詞語完全具體定義出來，在兩端及中點敘寫績效水準的方式預留給評量者在評分時彈性的拿捏空間，是個很不錯的方式，本研究中也將予以採用。但是，為避免此彈性空間可能對評量結果的效度造成不利影響，本研究對於績效水準的敘寫力求具體；同時亦提醒評量者(教師、同儕、學習者個人)要採取公平態度，切忌有偏袒、鄉愿及放水的情況產生。以Goldsby與Fazal (2001)評量學生的數位化學習歷程檔案例子而言，分為三個準規，分別是：(一)形式：主要在設計與美觀；(二)功能與使用性：主要在易於觀看性上；(三)內容成分：主要在於是否展現出要求的成品。對於每一個項目以四分等第的方式來進行評鑑。美國威斯康辛大學史坦道分校的評估準規是三分等第方式：差(poor; 1分)、好(good; 2分)、優(excellent; 3分)；該校教育、健康與人力科學學院及ePortfolio Portal Team的評估準規則是四分等第方式：不足(incomplete; 0分)、部分熟練(partially proficient; 1分或3分)、熟練(proficient; 2分或6分)、示範(exemplary; 3分或9分；ePortfolio Portal, 2009; Vandervelde, 2008)。University of New England的評估準規是四分等第方式：差(poor; 1分)、合理(fair; 2分)、好(good; 3分)、優(excellent; 4分；Rcampus, 2010)。美國Warren Community School (2000)的評估準規是四分等第方式：生手(novice; 1)、基本(basic; 2)、有能力(capable; 3)、熟練(proficient; 4)，以1、2、3、4表示等第，但未標示分數。Middlesex County College (2009)的評估準規是四分等第方式：不滿意(unsatisfactory; 5分)、可接受的(acceptable; 10分)、引人注目的(effective; 15分)、出眾的(exceptional; 20分)，但等第由低到高給5、10、15、20分。Microsoft Office的評估準規是五等第方式，但未標示等第之名稱，僅以1、2、3、4、5表示(Thompson, 2007)。San Diego State University (2008)的評估準規不是等第方式，而是某面向給一個分數比例(七個面向分別占3%、6%、36%、30%、12%、6%、7%)。以上縱然等第不同，等第的配分亦不同；但大部分都有一個共同點，就是除了有等第的名稱之外，每一個等第均有清楚的描述。上述不同等第量尺對評量的進行與評量結果有不同影響。

美國威斯康辛大學史坦道分校的評估準規包括技術(technical)、設計(design)、反思(reflection)、作品(artifact)四個面向(University of Wisconsin–Stout, 2010)。該校教育、健康與人力科學學院(College of Education, Health and Human Sciences)的評估準規包括作品選擇(selection of artifacts)、反思(reflection)、多媒體使用(use of multimedia)、版面與文字(layout and text elements)、寫作(writing)、註釋(annotations)、創造力(creativity)七個面向(Vandervelde, 2008)。ePortfolio Portal Team與University of New England採用類似的評估準規，包括作品選擇(selection of artifacts)、反思、多媒體使用、版面與文字(layout and text elements)、寫作、理論的解釋(rationale)或說明(caption)、容易導覽(ease of navigation)七個面向(ePortfolio Portal, 2009; Rcampus, 2010)。美國Warren Community School (2000)的評估準規包括反思、版面(layout)、檔案結構(folder structure)、圖像、聲音、合作、呈現、技巧(mechanics)八個面向，但未含作品面向。Morris (2007)的評估準規包括技巧(mechanics)、圖像、工具使用、內容相關、說明與反思(captions/reflections)、結

構(structure)六個面向，但未含作品面向。Middlesex County College (2009)的評估準規包括對作品的反應(response of assignment)、科技的創意使用(creative use of technology)、吸引力(attractiveness)、反思、文法五個面向。San Diego State University (2008)教育科技系評估碩士班學生檔案的準規與上述差別較大，包括個人簡介(3%)、作品的標準矩陣(standards matrix for artifact; 6%)、作品與反思(artifacts and reflections; 36%)、總括性的反思(comprehensive reflection; 30%)、使用性(usability; 12%)、表達(presentation; 6%)、整體品質(overall quality; 7%)七面向，括弧內百分比為評分比重。Microsoft Office的評估準規過於精簡，包括反思、內容(content)、連結與材料(link and material)，未明確指出作品面向(Thompson, 2007)。這些評估內涵是評估一個檔案本身的好壞，無法完整評量學生的學習表現，譬如學生的學習態度。上述評估準規中的設計、多媒體使用、版面與文字、版面、檔案結構、圖像、聲音、呈現、技巧、工具使用、連結與材料等都是屬於製作方面向。上述評估準規的一個缺點是過於強調檔案的製作，而漏掉了學習歷程檔案很重要的一項要件——對學習目標設定的評估與學生學習態度的評估。上述歸納出的作品與反思是學習歷程檔案內重要的項目，而製作是學習歷程檔案重要的活動，因此都將納入本研究的檔案評量準規。

Danielson與Abrutyn (1997)認為檔案評量準規可以包括目標設定、成長與成就、反思。這些準規將納入本研究中，做為檔案評量準規建立的參考。歐滄和(2002)表示，檔案評量的準規包括達成學習目標的程度、展現個人特色的程度、與他人合作的程度以及運用資源的能力。在本研究中，研究者將達成學習目標的程度歸類於量表內「學習目標」面向、與他人合作程度置於「態度」面向、展現個人特色與運用資源能力放置於「其他」面向中。針對上述各個與學習歷程檔案相關的評量準規，研究者將它們歸納成檔案製作、學習目標、作品、反思、態度與其他共六個面向(如表1)。每個面向皆有文獻的支撐，且文獻在上一段與此段文內皆已提到所列之面向。

在六個面向中，除了學習目標中「學科教育目標」的描述語會因為學科的內容而有

表1：數位化檔案評量準規之面向及文獻來源

面向	文獻來源
檔案製作	ePortfolio Portal (2009); Goldsby與Fazal (2001); Middlesex County College (2009); Morris (2007); Rcampus (2010); San Diego State University (2008); Thompson (2007); University of Wisconsin–Stout (2010); Vandervelde (2008); Warren Community School (2000); Worcester (2000)
學習目標	歐滄和(2002); Danielson與Abrutyn (1997)
作品	ePortfolio Portal (2009); Goldsby與Fazal (2001); Middlesex County College (2009); Rcampus (2010); San Diego State University (2008); University of Wisconsin–Stout (2010); Vandervelde (2008)
反思	Danielson與Abrutyn (1997); ePortfolio Portal (2009); Middlesex County College (2009); Morris (2007); Rcampus (2010); San Diego State University (2008); Skawinski與Thibodeau (2003); Thompson (2007); University of Wisconsin–Stout, (2010); Vandervelde (2008); Warren Community School (2000); Worcester (2000)
態度	歐滄和(2002); Worcester (2000)
其他	歐滄和(2002)

所差異外，其他的描述語及所有的準規皆能夠適用於所有學科的網路化檔案評量之用。由表1也可看出，學習目標與態度的文獻較少，較容易被忽略。製作方面是完成學習歷程檔案的基本功夫，而反思與作品是學習歷程檔案的最重要項目，都不可以忽略掉。

三、檔案評量的信效度

(一)檔案評量的信度

張麗麗(2002)強調，判斷檔案評量信度是否足夠，可以在不同檔案、不同評分者、不同時間或不同情境下，檢驗評量的結果是否一致。通常多次評量結果間的一致性即是指評量的信度(余民寧，2003；陳英豪、吳裕益，1991)。同一位評量者對相同標的物作數次評量結果之間的一致性，稱為內部信度或再測信度。不同評量者對相同標的物評量結果之間的一致性，稱為外部信度。Diperna與Derham (2007)、Gadbury-Amyot等(2003)的研究是檢視不同評分者之間的一致性。Rose-Hulman Institute of Technology使用的REPS即是採用評量者間信度方法(inter-rater reliability approach)以確保不同評量者之間的一致性(Burnett & Williams, 2009)。如果不同評量者之間的一致性低，則該系統不允許評量者繼續評量下去。本研究內的教學者評量即是採用此種方式檢驗評量的信度。但是，岳修平與王郁青(2000)表示檔案評量強調的是學生個人化的學習，而其最大的限制為評量結果往往只具備效度而不具備信度。張麗麗(2002)也表示學習歷程檔案由於受到主觀型計分方式、檔案內容項目數有限、檔案內容差異性大及資料為長時間的蒐集等因素的影響，分數變異的來源就變得多且複雜。一旦分數的變異來源受到影響，連帶地便會牽動到存在於傳統評量理論中信度所包含的誤差值，進而降低檔案評量的信度。

要如何提高檔案評量信度呢？透過明確、不複雜的計分規則可以使評分者間較容易達到一致性。另外，檔案準規的恰當與否、評分者對評量準規的熟悉度、評分者背景差異少、嚴謹的評分者訓練等因素可以提升評分的一致性。所以在進行檔案評量之前，應做好上述各項準備工作。本研究在實驗進行前將注意並控制這些因素，使其負面作用減至最低。

(二)檔案評量的效度

Russell與Butcher (1999)表示，檔案所蒐集的資料可能無法真正代表學生的學習過程或成果是檔案評量所會遇到的問題。顯示，檔案所蒐集資料的效度即是一個需要探討的問題。張麗麗(2002)表示直至目前為止，探討檔案效度方面的研究在質與量方面仍不足，多數是偏向以學生及教師自陳的方式來探討檔案評量在教學與學習上的影響，較缺乏系統性分析檔案效度的研究。她進一步指出：1. 透過明確界定檔案目標，直接將目標轉化為檔案活動及項目有助於提升檔案與所測特質間的關聯性；2. 將檔案目標直接轉換為計分規則有助於提升計分規則的效度。

Sweat-Guy與Buzzetto-More (2007)提到，數位化檔案評量團隊的建立可以確保評量的效度。亦即評量不能只由一人來擔任，最好由兩人以上來擔任。余民寧(2003)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為一個外在效度指標。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相關性。相關性越高，表示檔案評量分數的效度越高。Diperna與Derham (2007)使用同

時效度(或同步相關效度；concurrent-related validity)，檢驗師培生(preservice teachers)的學業表現與數位化檔案評量分數之間的相關性。由於本研究欲檢驗的效度為教師評量效度，故無法再以教師本身為外在效標；而是將學生測驗成績視為外在效標，以教師評分與學生測驗成績的相關性代表教師評分的效度。

參、研究方法

一、研究對象

研究對象為某高職二年級修習「計算機概論」課程的兩個班級學生，共72人(參與實驗79人，刪除檔案內容不全者七人)。其中男生34名，女生38名。課程學分數為三學分，每週上課三小時。教學實驗為期十二週，學生學習內容為「Word文書處理：版面設定及長文件編輯」兩個單元。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、自評與同儕互評，教師則透過系統檢視學習者的檔案並評量其學習表現。兩個單元的作業都是數位化作品，兼具技能操作與設計概念且方便上傳至檔案評量平臺內，適合網路化檔案評量的進行。期中與期末各有一次課堂考試，兩次平均為本研究的測驗成績。兩次課堂考試皆由授課教師自行出題，教師已具十餘年教學經驗。線上助教負責系統的維護、線上的引導、問題解答；另外，亦參與評分工作。兩班為同一教師、同一線上助教，以減少不同教學者的差異。線上助教由研究者擔任，亦為現職教師，具六年教學經驗。

二、實驗程序

實驗分兩階段進行，第一階段為課程單元一，第二階段為單元二。每個單元的時間為六週，各有一個作品。透過單元一的經驗，可以減低單元二時的困難，譬如提高評分者對評量準規的熟悉度與認知上一致性、充份的評分者訓練等。透過明確、不複雜的計分規則可以使評分者間較容易達到一致性。另外，檔案準規的恰當與否、評分者對評量準規的熟悉度、評分者背景差異少、嚴謹的評分者訓練等因素可以提升評分的一致性(Gadbury-Amyot, 2003; Oskay et al., 2008)。本研究在執行上，儘量控管這些要求。

Chang與Tseng (2009, 2011)的檔案評量活動內容完整，步驟具體明確，並已進行過數次教學實驗，受試者反應良好。故為本實驗採用，其教學活動包括：

(一) 教學者(一位授課教師與一位線上助教)在課堂上示範網路化檔案評量系統的操作，並說明各項評量準規的意涵、計分方式及如何評分，讓學習者進行檔案製作與評量時有所依循。

(二) 學習者課後在網路化檔案評量系統內，針對每一單元設定學習目標、撰寫反思、上傳作品(包括初期作品、完成作品、修正作品等不同歷程的作品)等活動，並對檔案的議題做線上討論。教學者透過該系統檢視學習者的學習狀況。

(三) 教學者在每個單元結束後數日內對每人的檔案做評量。授課教師與線上助教在評分前做充分的溝通與討論，以對檔案評量準規及評分有一致的認知。學習者在每個單元後數日內對自己的檔案進行自我評量，也對他人的檔案做同儕互評。互評方式是先由教師將學習者分為不同小組，再由小組成員跨組匿名互評(第一與第二組、第三與第四組……)。教學者每人評72份

檔案；學習者每人評一份自己的檔案及六份同儕的檔案，共計七份。

教學者點選網路化檔案評量系統左邊的「檔案評量區」，會出現所有學習者名字畫面供教學者挑選。教學者點選某位學習者後，如圖1所示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，會另外出現一個評分表視窗網頁。此時，教學者可點選每一項評量準規旁的分數及準規下面的文字方塊填寫回饋意見。評分時，教學者可同時瀏覽學習者的檔案內容與線上參與情形(如圖1)，以做為評分的參考。

三、資料蒐集與分析

本研究於兩個課程單元結束後，透過網路化檔案評量系統後端的資料庫，擷取教學者評分的資料。檔案評量結果為兩個課程單元的教學者評分之平均。測驗成績則為兩個課程單元的測驗分數之平均。針對研究問題所使用的統計方法及其意涵，如表2所示。



圖1：教師瀏覽學生的線上參與情形及評分

表2：研究目的、研究問題與統計方法對照表

研究目的	研究問題	統計方法	意涵
網路化檔案教學者評量的信度	不同教學者之間(授課教師、線上助教)的檔案評量結果是否一致？	Pearson積差相關	教學者之間評分一致性
	每一教學者(授課教師、線上助教)檔案評量結果(對不同份)是否一致？	同質性係數分析	評分者內部一致性
網路化檔案教學者評量的效度	各項準規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)？	Pearson積差相關	與效標分數一致性

四、檔案評量準規發展

透過文獻歸納出網路化檔案評量準規雛形，與授課教師討論修正後成為量表草稿，並建立表面效度。之後，將量表草稿製作成專家問卷調查表，郵寄給三位專家，進行專家問卷調查。回收後，召開專家焦點座談會議，討論專家問卷調查結果。於會議上，進行量表的修正，形成正式量表，並建立專家效度。評量準規分為檔案製作、學習目標、作品、反思、態度與其他共六個面向，共27項評量準規，詳如附錄。每個準規的滿分為5分，量尺為李克特五等第之間增加0.5的選項，評分者依據被評者表現給予1、1.5、2、2.5、3、3.5、4、4.5與5分。相較於1、2、3、4、5分方式，點數間以0.5為單位可增加評分時的選項，方便評分；相較於1、2、3、4、5、6、7、8、9、10分方式，該評分方式的間距較小，雖然評分結果(總分)的間接差異性無不同，但卻可減少評分結果(總分)的絕對差異性過大情況(余民寧，2003)。每個準規的點數之間皆定義不同層次的績效表現水準，以利評分。

五、檔案評量準規項目分析

首先針對評量準規進行項目分析(item analysis)。每一準規平均數高(27%)、低(27%)兩組之間的*t*檢定結果皆達顯著水準，顯示每一準規的鑑別力足夠，不必刪除。每一準規平均數與整體平均數的Pearson積差相關係數皆達顯著水準，顯示每一準規之間的一致性足夠，不必刪除。

六、檔案評量準規效度檢驗

評量準規各面向的抽樣適切性量數(Kaiser-Meyer-Olkin, KMO)值皆大於 .60(表3)，可進行因素分析並利用其主成分分析法(Principal Component Analysis, PCA)建構效度。因考量各因素面向之間具一定程度之相關性，故採斜交轉軸法(oblique rotation)；而Bartlett球形檢定之近似卡方分配皆達顯著，表示各準規之間有共同因素存在，適合進行因素分析。因素分析結果顯示，「態度」面向內有一準規的因素負荷量小於 .30，將此準規刪除。該準規是問到對課程的看法，與其他題項都是詢問對於線上檔案評量活動的看法有所不同，可能是被刪除的原因。第二回因素分析結果顯示，因素負荷量皆大於 .30，故保留所有準規。

最後共確定五個特徵值大於1的因素面向。整體解釋變異量達90%，顯示整體量表具高效度(表3)。各面向解釋變異量皆大於60%，顯示各面向皆具一定之效度。換言之，評量準規能測出檔案內容的品質。各面

表3：評量準規之因素分析

面向	KMO值	解釋變異量(%)
檔案製作	.61	61.50
學習目標	.65	82.12
作品	.83	72.82
反思	.67	68.50
態度	.66	83.26
其他	無	無
整體	.86	90.81

註：由於「其他」面向僅有一題，所以不做因素分析

向的解釋變異量以態度最高，學習目標次之，檔案製作面向最低。顯示態度最能測出其品質，檔案製作最難評測其品質。

七、檔案評量準規信度檢驗

如表4，量表各面向與整體Cronbach's α值皆大於 .70，代表各項準規之間具有高度的內部一致性，因此保留各項準規。各面向中以作品面向最高，反思面向最低。

表4：評量準規之Cronbach's α值

面向	Cronbach's α值
檔案製作	.71
學習目標	.81
作品	.85
反思	.70
態度	.81
其他	無
整體	.88

肆、結果與討論

一、教學者評量的信度

(一)不同評分者之間的檔案評量結果是否一致？

授課教師與線上助教對72位學生檔案評量結果的Pearson積差相關如表5。結果顯示授課教師與線上助教之間對不同學習者的檔案評量結果具有很高的相關性且達顯著，顯示授課教師與線上助教間的評量具有高度的一致性。其中以檔案製作面向最高，反思面向最低，顯示反思面向的評審較不容易達成一致。檔案製作面向評分的一致性最高，可能是因為檔案製作屬於技能面向，較容易評分；而反思面向最低，可能是因為反思屬於情意面向，較不容易評分。由於不同評量者間存在有不同的主觀判斷與意見，可能會導致評量結果的誤差而影響信度(余民寧，2003)。但本實驗授課教師與線上助教間的評量結果具有高度的相關性，顯示評量者雖有主觀判斷與意見的不同，但是評量結果卻相當一致。Gelinas (1998)的檔案評分者之間的整體一致性高(r = .81)，與本研究類似。Gadbury-Amyot等(2003)的研究對紙本式檔案評量信度做檢定，兩位評分者之間相關係數介於 .28至 .60之間；Rees與Sheard (2004)檢驗結果顯示，兩位檔案評分者之間相關係數介於 .36至 .69之間。此兩研究結果雖然都較本研究為低，但都達顯著水準。這兩個研究與本研究不同的是，這兩個研究的對象為大學生，使用紙本式檔案而非網路化檔案，且未交代評分者的身分。

(二)每一位評分者(含授課教師、線上助教)對不同份檔案的評量結果是否一致？

同質性係數可用來檢驗某評分者對不同標的物的評分是否一致，本研究用它來檢驗某評量者對不同份檔案的評分是否一致。同質性檢定(test for homogeneity)是指不同母體(教師與線上助教)的不同樣本(檔案)在某個變項(檔案分數)上的分配情形無差異，是獨立性考驗的一種變化(邱皓政，2008)。本研究採百分比同質性檢定，考驗四個群體(教師與線上助教所評檔案)在不同份檔案的評量分數的百分比是否一致。教師及線上助教評分的同質性係數分析如表6。兩人對72份檔案評分結果的同質性係數皆達顯著性，顯示這兩位評分者對不同份檔案評定結果皆具一致性(即評分者內部一致性)。整體同質性係數達顯著性，顯示兩位評分者對不同份檔案的評量結果依然是一致的。

表5：不同評分者評量結果之Pearson積差相關

檔案評量	相關係數	顯著性	教師評分		線上助教評分	
			平均數	標準差	平均數	標準差
檔案製作	1.00	.000***	4.03	0.41	4.06	0.43
學習目標	.98	.001***	3.68	0.25	3.72	0.28
作品	.84	.019**	3.77	0.27	3.81	0.30
反思	.75	.034**	4.01	0.32	4.03	0.35
態度	.94	.006***	3.48	0.38	3.52	0.41
其他	.93	.000***	3.62	0.39	3.67	0.39
整體	.91	.000***	3.76	0.28	3.80	0.34

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

表6：評分者對不同份檔案評量結果同質性係數分析

人員	個人同質性係數	個人z值	整體同質性係數	整體z值
教師	.97	1.91**	.95	1.87**
線上助教	.93	1.83**		

**$p < .010$

二、教學者評量的效度

(一)各項準規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)？

檔案評量分數與學生測驗成績的Pearson積差相關如表7。測驗成績為結果顯示，教學者(授課教師與線上助教的平均)的檔案評量分數與學生測驗成績結果具有高度的相關性且達顯著，表示教學者的檔案評量分數與學生測驗成績具有高度的一致性，也顯示出檔案評量的準規足以檢測出學習成就。此結果符合Gelinas (1998)提到的「檔案評量分數與被評者的學習表現有正相關」。不同的是，該研究對象是外語教師，使用紙本式檔案而非網路化檔案。其中以學習目標面向的相關性最高，其次為檔案製作面向，其他面向最低，顯示學習目標與學習成就最容易達成一致。

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	.78	.000***
學習目標	.79	.000***
作品	.67	.000***
反思	.63	.000***
態度	.59	.000***
其他	.57	.000***
整體	.84	.000***

***$p < .001$

三、綜合討論

就評量的信度而言，由表5授課教師與線上助教評量結果之Pearson積差相關表與表6評分者評量結果同質性係數分析表可以發現，不同評分者之間及其對不同份檔案評分結果皆具高度一致性。換言之，評量的信度足夠。就評量的效度而言，表7檔案分數與測驗成績之Pearson積差相關表顯示檔案評量分數與學生測驗成績具有高度的相關性且達顯著，表示檔案評量分數與學生測驗成績具有高度的一致性，足以反映出學生學習成就。綜上所述，網路化檔案教學者評量

是可以展現出足夠的信度與效度。Diperna與Derham (2007)的研究對師培生(preservice teachers)數位化檔案做信度檢定，兩位評分者之間相關係數低且未達顯著。他們也對數位化檔案做效度檢定，檔案評量分數與學生學業表現的相關係數低且未達顯著。這兩個結果都與本研究相反，原因可能為研究對象的教育層級、樣本大小、學科內容、評量準規、數位化檔案環境、評分者所受訓練等不同所致。

Oskay等(2008)綜合幾個檔案評量信、效度的研究結果，認為檔案評量是可以信賴的，這與本研究十分吻合。不同的是，這幾個研究的對象、樣本大小、學科內容、評量準規都不相同，且都是紙本式檔案而非網路化檔案。至目前為止，已有一些紙本式檔案評量與網路化檔案評量信、效度的研究。儘管這些研究的結果不一，但大部分是信、效度足夠的，少部分證明信、效度不足。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。但無論如何，要讓網路化檔案教學者評量具有足夠信、效度是可以做到的。

伍、結論與建議

網路化檔案評量準規施測之整體與各面向之Cronbach's α值皆大於 .70，顯示評量準規具足夠之信度。其中以作品面向最高，反思面向最低。就評分者之間信度而言，評分者之間對不同學習者的檔案評量結果具有很高度的一致性。就評分者內部信度而言，評分者對不同檔案評分結果具一致性(所有評分者皆如此)。因此，教學者網路化檔案評量同時具備評分者間一致性與評分者內一致性，是可以具足夠信度的。訓練有素的評分者、評分者對評量準規有共同的認知、具體明確的評量準規是可以提升評分的信度(Gadbury-Amyot, 2003; Oskay et al., 2008)。Diperna與Derham (2007)也提到評分者的訓練可以提升評分者之間的一致性。Johnson等(2003)提到，當增加第三個評分者之後，評分的信度自然會提高。Sulzen等(2008)的研究顯示，增加評分者人數是可以增加評分信度。本研究在實施前與過程中，研究者盡力讓評分者對評量準規有足夠的瞭解與討論(如研究方法小節內所述)，研究結果也顯示內部與外部信度皆達顯著。因此本研究綜合上述文獻與研究結果，認為如果評量者熟悉評量準規，對準規有一致的認知，並且以公正的態度進行評分，將可改善岳修平與王郁青(2000)所指出評量結果往往只具備效度而不具備信度的問題。以各面向來看，評量結果之間的一致性以檔案製作面向最高；反思面向最低，顯示反思面向的評審較不容易達成一致。

網路化檔案評量準規的因素分析之整體解釋變異量高，足以檢測出檔案內容的品質。檔案評量分數與學生的測驗成績具有高度的一致性，也顯示足以反應一定程度之學習成就。綜合兩項結果，此評量準規與評量方法具有良好的效度。張麗麗(2002)指出在檔案評量過程中須留意學習者多方面的學習狀態。本研究認為為了反應被評量者的學習狀況，評量準規的訂定應盡量做多元的考量，完整的評量準規能更有效衡量學習成就。本研究建構的網路化檔案評量準規較適用於學習者依照網路系統(或稱網路化檔案評量系統平臺)的表單填寫資料來製作的結構式(structured)學習歷程檔案(每人的檔案格式與內容相同)。但檔案有不同的類型，不同的類型可能有不同的評量準規。因此建議未來研究可針對簡報軟體、文書處理軟體、網頁製作軟體等建置的自我建構式(self-

constructed)數位化學習檔案來製定可行的檔案評量準規並檢驗其信、效度。此網路化檔案評量準規並未牽涉到學科的內容與特性，故大體上適用於不同學科的檔案評量。僅在學習目標設定面向內的「學科教學目標：學習者達到學科所訂立的教學目標」內，可以加入對某學科教學目標的描述，以方便評分者評分。此網路化檔案評量準規與紙本式檔案評量不同處有：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指標；在情意、態度、互動面向上，增加了線上觀摩、欣賞、互評與回饋的表現，網路資源的分享，線上互動的數量及品質等評量項目。

本研究中所建立的評量準規並沒有將學習者的測驗成績納入，原因是不讓檔案評量複雜化並利於之後的統計分析。若將測驗成績加入檔案評量準規中，則檔案評量成績就幾乎代表了所有學習者表現的成績。但在進行譬如檔案評量成績與測驗成績之相關性檢驗時，必須將測驗成績自檔案成績中剔除，以避免重複計算。本研究對於涉及技能及認知成分較多的評量準規，嘗試對它的績效表現遵循ABCD (對象、行為、條件、程度)模式來敘寫。對於較具情意態度色彩的準規則嘗試用舉出具體事例的方式，以求能將情意態度化為可具體觀察的行為表現。雖然如此，並未能完全解決評量者在使用這些準規進行評量時所遭受到的評分問題，但是對於評量效度的提升仍多少會有所幫助。如果能佐以參考的實際範例，或者在文字遣詞上予以簡化會更佳。就如同Oskay等(2008)提到的「為了達到高度的評分者一致性，應該讓學生對檔案有充分的瞭解，且學生的檔案內容與格式應儘量一致。此外，評分者必須訓練有素且富評量經驗，並使用具體明確的評量準規；所有評分者對評量目的有共同的認知，而且對預期的學生學習表現有深入瞭解」。上述行為對評量信、效度有提升作用，也是本研究過程中一再注意的細節。理想上，授課教師之間的評分結果相比，較能回答原本研究的問題。但因礙於現實教育環境，一門課無法安排由兩位授課教師任教；因此，另一位評分者是由線上助教擔任。不過線上助教也是現職教師，只不過不是該課授課教師罷了，但等同於由兩位教師評分。這雖然是本研究限制之一，但已將限制減至最低。另外，本研究樣本數稍少，也可能影響研究結果。未來類似的研究宜採更大樣本，使研究的信、效度更為穩當。

誌謝

本研究由國科會經費補助(NSC 98-2511-S-003-032-MY3)，特此致謝。

參考文獻

1. 余民寧(2003)。**教育測驗與評量**。臺北市：心理。
2. 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。
3. 邱皓政(2008)。**量化研究與統計分析——SPSS中文視窗版資料分析範例解說**。臺北市：五南。
4. 郭生玉(2004)。**教育測驗與評量**。臺北市：精華。
5. 張美玉(2001)。從多元智能的觀點談歷程檔案評量在教育上的應用。**教育研究資訊**，**9**(1)，32-54。
6. 張麗麗(2002)。檔案評量信度與效度的分析——以國小寫作檔案為例。**教育與心理研究**，**25**，1-34。

7. 陳英豪、吳裕益(1991)。**測驗與評量**。高雄市：復文。
8. 歐滄和(2002)。**教育測驗與評量**。臺北市：心理。
9. 盧貞穎(2001)。讓學生做學習的主人——檔案評量的應用與省思。**敦煌英語教學雜誌，30**，15-21。
10. Alter, J. A., & Spandle, V. (1992). NCME instructional module: Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, *11*(1), 36-44.
11. Barker, K. C. (2006). *Environmental scan: Overview of the ePortfolio in general and in the workplace specifically*. Retrieved July 12, 2011, from http://www.futured.com/documents/LIfIABackgroundePforSkilled-Immigrants.pdf
12. Batson, T. (2002). The electronic portfolio boom: What's it all about? *Syllabus*, *16*(5). Retrieved September 16, 2009, from http://campustechnology.com/articles/2002/11/the-electronic-portfolio-boom-whats-it-all-about.aspx
13. Bloom, L., & Bacon, E. (1995). Using portfolios for individual learning and assessment. *Teacher Education and Special Education*, *18*(1), 1-9.
14. Bonita, W. L., & Duquence, L. (1999). Assessing teacher performance with a portfolio rubric. *English Leadership Quarterly*, *22*(2), 6-11.
15. Brickley, D., Schwartz, A. I., & Suen, H. K. (2000). *Web-based assessment: Validation of electronic portfolios*. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA.
16. Burnett, M. N., & Williams, J. M. (2009). Institutional uses of rubrics and e-portfolios: Spelman college and Rose-Hulman institute. *Peer Review*, *11*(1), 24-27.
17. 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.
18. Chang, C. C. (2008). Enhancing self-perceived effects using web-based portfolio assessment. *Computers in Human Behavior*, *24*(4), 1753-1771.
19. Chang, C. C., & Tseng, K. H. (2009). Use and performances of web-based portfolio assessment. *British Journal of Educational Technology*, *40*(2), 358-370.
20. Chang, C. C., & Tseng, K. H. (2011). Using a web-based portfolio assessment system to elevate project-based learning performances. *Interactive Learning Environments*, *19*(3), 211-230.
21. Danielson, C., & Abrutyn, L. (1997). *An introduction to using portfolios in the classroom*. Alexandria, VA: Association for Supervision and Curriculum Development.
22. Dennis, C., & Hardy, J. (2006). Development of a model to advance the uptake of e-portfolios for undergraduates in teacher education and registered nurse preparation: An exemplar of best practice. In E. Pearson & P. Bohman (Eds.), *Proceedings of world conference on educational multimedia, hypermedia and telecommunications 2006* (pp. 248-253). Chesapeake, VA: AACE.
23. Diperna, J., & Derham, C. (2007). Digital

03-張基成.indd 406　2012/10/29 上午 09:41:58

professional portfolios of preservice teaching: An initial study of score reliability and validity. *International Journal of Technology and Teacher Education*, *15*(3), 363-381.

24. e-Portfolio Consortium. (2003). *Electronic portfolio white paper*. Retrieved October 3, 2009, from https://portfolio.du.edu/portfolio/getportfoliofile?uid=147385
25. ePortfolio Portal. (2009). *A generic rubric for evaluating ePortfolios*. Retrieved July 22, 2011, from http://www.danwilton.com/eportfolios/rubric.php
26. 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 City, MO.
27. 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.
28. Gelinas, A. M. (1998). *Issuse of reliability and validity in using portfolio assessment to measure foreign language teacher performance*. Unpublished doctoral dissertation, Ohio State University, Columbus, OH.
29. Goldsby, D., & Fazal, M. (2001). Now that your students have created web-based digital portfolios, how do you evaluate them? *Journal of Technology and Teacher Education*, *9*(4), 607-616.
30. Johnson, R. L., Fisher, S., Willeke, M. J., & McDaniel, F., II. (2003). Portfolio assessment in a collaborative program evaluation: The reliability and validity of a family literacy portfolio. *Evaluation and Program Planning*, *26*(4), 367-377.
31. 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. 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.
32. Lorenzo, G., & Ittelson, J. (2005). *An overview of ePortfolios*. Retrieved on October 2, 2009, from http://net.educause.edu/ir/library/pdf/ELI3001.pdf
33. Lynch, L. L., & Purnawarman, P. (2004). Electronic portfolio assessments in U.S. educational and instructional technology programs: Are they supporting teacher education? *TechTrend*, *48*(1), 50-56.
34. Middlesex County College. (2009). *Grading rubric*. Retrieved July 22, 2011, from http://eportfolio.middlesexcc.edu/control.cfm/ID/3023
35. Morris, J. L. (2007). *Rubric for assessing electronic portfolios*. Retrieved July 22, 2011, from http://www.uvm.edu/~jmorris/rubricep.html
36. Oskay, O. O., Schallies, M., & Morgil, I. (2008). Reliability of Portfolio: A closer look at findings from recent publications. *Hacettepe Unuversity Journal of Education*, *35*, 263-272.
37. Rcampus. (2010). *iRubric: ePortfolio rubric–University of New England*. Retrieved July 22, 2011, from http://www.rcampus.

com/rubricshowc.cfm?code=W38X23&sp=true&

38. Rees, C. E., & Sheard, C. E. (2004). The reliability of assessment criteria for undergraduate medical students' communication skills portfolios: The Nottingham experience. *Medical Education*, *38*(2), 138-144.
39. Rogers, G. (2003). *RosE-portfolio*. Retrieved July 9, 2011, from http://www.rose-hulman.edu/reps/
40. Russell, J. D., & Butcher, C. (1999). Using portfolios in educational technology courses. *Journal of Technology and Teacher Education*, *7*(4), 279-289.
41. San Diego State University. (2008). *EDTEC masters professional e-portfolio*. Retrieved July 22, 2011, from http://edtec.sdsu.edu/portfolio.htm
42. Singh, O., & Ritzhaupt, A. (2006). Student perceptive of organizational uses of ePortfolios in higher education. In E. Pearson & P. Bohman (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2006* (pp. 1717-1722). Chesapeake, VA: AACE.
43. Skawinski, S. F., & Thibodeau, S. J. (2003). A journey into portfolio assessment. *The Education Forum*, *67*(1), 81-88.
44. Sulzen, J., Young, M., & 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.
45. Sweat-Guy, R., & Buzzetto-More, N. A. (2007). A comparative analysis of common e-portfolio features and available platforms. *Issues in Informing Science and Information Technology Education*, *5*(1), 327-342.
46. Thompson, S. (2007). *ePortfolio rubric that can be used with office live workspace*. Retrieved July 22, 2011, from http://office.microsoft.com/en-us/templates/
47. University of Wisconsin–Stout. (2010). *Evaluating your portfolio*. Retrieved July 22, 2011, from http://www2.uwstout.edu/content/art/artedportfolios/evaluating/portevaluatiorubric.html
48. Vandervelde, J. M. (2008). *Eportfolio (digital portfolio) rubric*. Retrieved July 22, 2011, from http://www.uwstout.edu/soe/profdev/eportfoliorubric.html
49. Wang, C. I., Peng, H. Y., & Lu, W. H. (2006). A framework for the assessing of intercultural communicative competence: A mixed approach. In E. Pearson & P. Bohman (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2006* (pp. 2979-2984). Chesapeake, VA: AACE.
50. Wang, C. X. (2009). Comprehensive assessment of student collaboration in electronic portfolio construction: An evaluation research. *TechTrends*, *53*(1), 58-66.
51. Wang, S. (2009). E-portfolios for integrated reflection. *Issues in Informing Science and Information Technology*, *6*, 449-460.
52. Warren Community School. (2000). *Electronic portfolios scoring rubric*. Retrieved July 22, 2011, from http://seuss.wcs.sad40.k12.me.us/ep/rubric.htm

53. Winsor, P. J., & Ellefson, B. A. (1995). Professional portfolios in teacher education: An exploration of their value and potential. *Teacher Educator*, *31*(1), 68-81.
54. Zalatan, J. 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. 37-43). Washington, DC: American Association for Higher Education Press.

附錄

面向	準規
檔案製作	1-1 資料完整性： 完整填寫網路化檔案評量系統要求之資料。
	1-2 內容恰當程度： 檔案中內容資料與檔案製作目的相關。
	1-3 內容豐富程度： 除了規定項目外，另有蒐集其他相關資料。 如：測驗結果、觀察記錄、文章檔案、講義檔案、課堂軼事趣聞、活動過程與結果、好站分享、好書分享……有益於增加檔案豐富性者(註：視需要可以上傳不同格式之檔案，如掃描圖檔、Word檔、PowerPoint檔、影音檔……等)。
	1-4 組織與呈現： 資料呈現清楚、美觀，內容層級安排恰當。
學習目標	2-1 個人學習目標設定： 學習者針對自身情況(如能力、自我期許……等)訂立明確而適當的學習目標及預期成果。
	2-2 學科教學目標： 學習者達到學科所訂立的教學目標。 如：以「高職計算機概論」學科為例，學科教學目標有展現操作套裝軟體的基本知識與能力、瞭解資訊網路之建構與應用、科技新知的發展在電子計算機概論上的效益、展現團體合作精神、正確使用電子計算機的態度等相關證據。
	2-3 成長與成就： 學習者顯示在學習內容、成果、態度等方面的改善與進步。
	2-4學習目標達成程度： 學習者達成個人學習目標的程度。
作品	3-1 作品正確性： 學習者作品的正確性。
	3-2 作品恰當性： 學習者作品的恰當性。
	3-3 作品豐富性與完整性： 學習者作品的豐富性與完整性。
	3-4 作品難度： 學習者作品的難度。
	3-5 作品創意與創新： 作品展現創意與創新。
	3-6 對學習內容的理解： 學習者展現對學習內容的理解程度。
	3-7 作品產生的歷程： 學習者上傳其他可以代表作品產生歷程的相關事物，如練習作業、原始作品、修改的作品……等。

反思	4-1 學習目標的反思： 對自己設定的學習目標進行反思。
	4-2 作品的反思： 學習者對作品產生過程與結果的心得闡述。
	4-3 學習成就的反思： 學習者在反思中舉出自己的學習成就、缺點、進步情形。
	4-4 學習態度的反思： 學習者對自己的學習態度優缺點、進步情形進行反省思考。
	4-5 觀摩同儕表現的反思： 觀察同儕表現(含檔案製作、學習目標、作品、反思、學習成就、學習態度等)所做出的自我期許。(註：網路化檔案評量系統提供同儕觀摩與互評)
	4-6 針對回饋的反思： 對於他人(教師、同儕等)所給予的回饋進行反思，且提出改進或解釋。(註：網路化檔案評量系統提供同儕觀摩與互評)
態度	5-1 線上觀摩、欣賞、互評、回饋： 對同儕表現(含檔案製作、學習目標、作品、反思、學習成就、學習態度等)作出評論，並給予回饋意見。(註：網路化檔案評量系統提供同儕互評)
	5-2 資源分享： 運用網路化檔案評量系統功能提供相關的資源。
	5-3 線上溝通討論、知識分享、意見交換、問題解決： 運用網路化檔案評量系統功能進行溝通討論、知識分享、意見交換、問題解決。 如：在系統的討論區中與他人討論、給予同儕鼓勵或協助問題解決、踴躍回覆問題等足以展現學習者在溝通討論、知識分享、意見交換、問題解決等方面事跡者。
	5-4線上多元互動數量： 此項根據網路化檔案評量系統線上行為紀錄(作品觀摩次數與時間、發表與回覆文章次數與字數、檔案瀏覽次數與時間……等)列出相對排名，再自動轉換成等第並給予點數。評量者無需評分。
	5-5線上多元互動品質： 與其他人有正向的互動(發表與回覆文章的內容、互評與回饋的內容……等)。 如：在系統討論區或同儕互評回饋中沒有惡意質疑或批判其他學習者的行為等足以展現學習者與同儕間有正向互動事跡者。
其他	6-1 其他事蹟或證據： 如除上述以外，其他足以展示或證明學習者學習歷程或進步情況之證據者。

註：1. 每項評分為1、1.5、2、2.5、3、3.5、4、4.5與5分，點數間皆定義不同的績效表現水準(因受限於篇幅，未列入)。

2. 評量準規中不含測驗成績。

科學教育學刊
2012, 第二十卷第五期, 393-412

Chinese Journal of Science Education
2012, 20(5), 393-412

Reliability and Validity of Instructor Assessment in a Web-Based Portfolio Assessment Environment

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 Municipal Sung-Shan Junior High School

Abstract

This study investigates the reliability and validity of instructor assessment in a web-portfolio assessment environment. The research target group consists of seventy-two students who take "Introduction to Computer" course of the second grade in some vocational high school. During experimental period, the students use the Web-based portfolio assessment system for creating their portfolio, emulating peers' portfolio, and doing self-assessment and peer-assessment. The teacher uses the system for inspecting and assessing students' learning performances. The results reveal that the rubric of the Web-based instructor assessment has an adequate reliability. The web-based instructor assessment has a high reliability between assessors as well as a high internal reliability within one assessor. Therefore, it is an assessment method with high reliability. In regard to validity, the rubrics of the web-based portfolio assessment have an adequate validity. The consistence between portfolio instructor assessment score and test score is high, which means the web-based instructor assessment has a high validity. Finally, the study recommends to develop the rubric of self-constructed e-portfolio created by PowerPoint, Word and FrontPage software and to investigate its reliability and validity.

Key words: Reliability, Validity, e-Portfolio, Web-Based Portfolio, Web-Based Portfolio Assessment

* Corresponding author: Chi-Cheng Chang

03-張基成.indd 412 2012/10/29 上午 09:41:59