

教育資料與圖書館學

*Journal of Educational Media & Library Sciences*

<http://joemls.tku.edu.tw>

---

Vol. 49 , no. 1 (Autumn 2011) : 135-170

網路化檔案評量環境下教學者評、學生自評與同儕 互評  
之信效度比較

**Comparisons of Reliability and Validity among  
Instructor Assessment, Student Self-Assessment and  
Peer-Assessment under Web-Based Portfolio  
Assessment Environment**

張基成 Chi-Cheng Chang\*

Professor

E-mail: [samchang@ntnu.edu.tw](mailto:samchang@ntnu.edu.tw)

吳明芳 Ming-Fun Wu

Teacher

E-mail: [wumingfang@ms26.url.com.tw](mailto:wumingfang@ms26.url.com.tw)

**[English Abstract & Summary see link](#)**

**[at the end of this article](#)**

---



# 網路化檔案評量環境下教學者評、學生自評與同儕互評之信效度比較

張基成\*

教授  
台灣師範大學科技應用與人力資源發展系  
E-mail: samchang@ntnu.edu.tw

吳明芳

教師  
台北市立松山家商資料處理科  
E-mail: wumingfang@ms26.url.com.tw

摘要

本研究比較網路化檔案評量環境下教學者評、學生自評與同儕互評的信效度。研究樣本為某高職二年級修習「計算機應用」課程的72位學生。學生使用網路化檔案評量系統進行檔案的製作、觀摩、自評與同儕互評，教師透過系統檢視學生的檔案並評量其學習表現。在三種評量結果之間差異上，教學者評、學生自評、同儕互評三種評量結果有顯著差異，其中教學者評量最為嚴格、其次為學生自評，同儕互評最為寬鬆。在三種評量結果之間一致性上，僅自評與教師評之間具一致性。自評與同儕互評之間、同儕互評與教學者評之間，都不具一致性。在三種評量結果與測驗成績一致性上，教學者評分結果與測驗成績具高度一致性，學生自評與測驗成績亦具高度一致性；而同儕互評結果與測驗成績一致性低。

**關鍵詞：**網路化學習歷程檔案，網路化檔案評量，自評，同儕互評

## 研究背景

學習歷程檔案(portfolio)最主要作用是做評量用，此時稱檔案評量(portfolio assessment)。網路化檔案評量(web-based portfolio assessment)是指學生在網

---

\*本文主要作者兼通訊作者。

路上呈現學習歷程與成果資料，讓教師或學生線上檢視與評估學生努力、進步與成就的一種評量方法(Chang, 2008)。網路化檔案評量的方法主要包括教師評量、學生自評、同儕互評。教師進行檔案評量時，可根據教學上的需要，採用一或兩種或全部評量方法。Chang與Tseng(2009a)一些網路化檔案評量的實驗研究，都是同時使用這三種評量方式，以達到真實性評量。其中，教師評是大部分研究最常使用的方式。但若只有教師評，可能忽略了學生的想法。Jenkins(2004)就提到，同儕意見對學生影響高過教師意見。因此，若有學生自評或同儕互評，可增加學生的參與度，提昇學生的學習興趣與動機。

學習歷程檔案做為評量工具，有如下優點：為動態、真實與整體的評量；注重過程與結果的評量；評量權由教師轉到學生身上，促使學生不斷自我改進；與自己先前結果比較，尊重學生個別差異；資料來源多重與多樣，提高了評量品質(張振成，1997；曾吉雄，1999)。這些優勢必需透過教師評量、學生自評、同儕互評才得以達到。教師評固然不可或缺，另兩種方法若能配合使用，更能增加檔案評量的真實性。學生自評除可增加學生的參與，提昇學習興趣與動機，也可促進其自我反省及增進自我負責的態度。Chen(2010)研究顯示，同儕互評時，被評者可從同儕評量結果獲得比來自於教師更多的靈感，因同儕比教師更了解自己。她的研究亦顯示，同儕評分時可觀摩到他人的作品，會將他人表現與自己做比較，進而對自己作品做反省；亦可從被評者作品獲得啟發，進而改善自己作品。因此，同儕互評除具有前述學生自評的作用之外，尚有觀摩、比較、見賢思齊，見不賢而自省的作用。

Russell與Butcher(1999)表示在教師使用學習歷程檔案來進行評量時，學生常會因不明了其好處，只知會花更多時間而排斥。譬如學生會認為自評、同儕互評會花掉他們一些時間，從而增加負擔。因而使得自評、同儕互評的結果受到質疑。此外，評量結果可能無法真正代表學生的學習成就，亦是一個檔案評量常碰到的問題。因此這時，教師評量、學生自評、同儕互評的信效度就顯得格外重要。Oskey、Schallies與Morgil(2008)綜合幾個檔案評量信效度的研究結果，結論認為檔案評量是一個恰當且可靠的評量方法。但這些研究的評分者大多是教學者而非學生，較少探討同儕檔案互評與學生檔案自評的信效度。Lin、Liu與Yuan(2001a, 2001b)及劉旨峰、林珊如與袁賢銘(2000)的研究顯示，在網路同儕互評活動中，同儕互評的效度高於學生自評的效度。上述結果顯示，教師評量、同儕互評的信效度優於學生自評。這三種評量方法結果若欲納入成績計算，比例之大小應根據其信、效度而定。通常教師評信效度較高，所佔比例可較高。

另外，有關於三種評量方式的嚴格或寬鬆，Lin、Liu與Yuan(2001a, 2001b)及劉旨峰、林珊如與袁賢銘(2000)的研究指出，學生自評的分數最

高、同儕互評的次之，教師評量的分數最低，並進一步建議網路評量可採同儕互評方式。上述結果顯示，教師評分最嚴格，同儕互評次之，學生自評最寬鬆。Sadler & Good (2006) 亦有類似的結果顯示，高中學生同儕互評分數會低於自評分數。這些結果顯示，教師評分最嚴格，同儕互評次之，學生自評最寬鬆。上述研究大多是針對紙本式的檔案評量，而電腦化評量因受到科技的支援、人在科技環境下的心智等因素影響，其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（Bouzidi & Jaillet, 2009）。因此在網路化檔案評量環境下是否也相同，抑或有不同結果出現？

基於上述背景，本研究欲比較在網路化檔案評量環境下，教師評量、學生自評、同儕互評三種評量方法的信、效度。研究問題包括三種評量方法的結果是否一致？是否有顯著差異？哪一個評量方法的信度較高？哪一個評量方法的效度較高？哪一個評量方法較能夠檢測出學習成就？哪一個評量方法較嚴格？哪一個評量方法較寬鬆？這些問題或所獲結論可供相關研究及教師進行網路化檔案評量的參考。

## 二、文獻探討

### (一) 檔案評量方法

檔案評量(portfolio assessment)是教師依據「學生配合教學目標與計畫而持續一段時間主動蒐集、組織與省思學習成果的學習歷程檔案」來評定其努力、成長、進步的情形（鄒慧英，2000）。檔案評量的方式有哪些呢？Burch (1999) 提出雙尾式評量(two-tailed assessment)方式，由教師與學生共同擬定評量標準，並共同參與評量的進行。檔案評量的方式有哪些呢？傳統評量通常只有教師評量，較少學生參與。但檔案評量不同，須強調學生的自主及參與，以達到真實性評量的目的。因此，除教師評量外，還必須有學生自我評量、同儕互評。尤其網路化檔案評量環境下，教師線上評量、學生線上自評、同儕線上互評可以遠距、即時進行，分數可立即獲得。因此，評量的進行更為方便，也可減少紙面式評量的缺點（Chang & Tseng, 2009b）。

Burch (1999) 提到，檔案評量整個過程對於培養學生成為一個負責任學生，及教室民主氣氛的營造有很大的幫助。學生透過自我評量，可對自己的學習有更深的反省，並對自己的學習與評量行為負責。Chang (2008) 提到，如果對作品做自評，學生可對自己作品的優缺點與學科知識技能有更進一步了解，進而促進其對作品做持續改善的動力。透過同儕互評，學生可對同儕的學習有所了解，並進而轉化成對自己學習的反省與鞭策。如果對作品做同儕互評，學生可透過觀摩同儕的作品，與自己作品比較的過程，更了解自己作品的優缺點；也可從同儕所給的回饋意見，做為作品改善的參考（Chang & Tseng,

2009a)。Chang & Tseng (2009b) 的研究就證實網路化檔案評量的使用對學生的自我評量、同儕互評能力有提昇作用。綜上所述，在網路化檔案評量環境下，學習者線上自評、同儕互評與教學者評量同樣都扮演不可或缺的角色。

## (二) 檔案評量標準

檔案評量應包括哪些項目呢？Reckase (2002) 認為檔案的評量項目應包括對學科內容了解的展現程度；內容所欲傳達的意思；自我反省思考；內容豐富性與難度；組織與呈現。Burch (1999) 提到檔案評量標準的構面，可包括檔案呈現（風格特殊、富有創造力的表現方式）、內容組織（井然有序的呈現，且在檔案內移動是方便的）、反省思考（深入思考且察覺自己的表現情況）、內容紀錄（豐富且解說詳盡）、內文鋪陳技巧（拼字、標點符號及用字遣詞均恰當）。兩位提到的檔案評量標準有一些共同的交集，譬如反省思考、內容組織、內容豐富等。大部分指標較傾向於檔案本身好壞的評斷，較少涉及對學科知識的了解程度（只Reckase第一項有關於此）。因此，檔案內容要項之一的作品表現可納入評量標準內。上述「內容豐富性與難度」就可衍申為「作品內容豐富性與難度」，而非只狹隘的「資料豐富性與難度」。前者牽涉到對學科知識與技能的了解，而後者僅是單純的資料整理與呈現。上述評估標準除可做為作品的量測題項，亦可做為檔案製作、反思的量測題項。

Azusa Pacific 大學教育科技系評估檔案的標準包括教學設計、整體學習物件 (global learning object)、成長評量、超連結、應用 (use of applications)、課程應用 (application of coursework) (High Plains Regional Technology in Education Consortium, 2005)。其中的成長評量、課程應用直接與學生的學習表現有關。美國 Milford School District 學區的國中以促進學生資訊與傳播科技應用能力為出發點，實施檔案評量，要求學生製作數位化檔案並達到以下的能力標準才可畢業：創造力與創新 (creativity and innovation)、溝通與合作 (communication and collaboration)、研究與資訊流暢 (research and information fluency)、批判思考、問題解決與決策 (critical thinking, problem solving and decision making)、數位國民素養 (digital citizenship)、科技操作與概念 (technology operation and concepts)、內容 (content)、反思 (Milford School District, 2009)。馬里蘭大學 Eastern Shore 校區商業、管理與會計系的評量標準分四部分，學習目的、反思、檔案品質、整體表現。學習目的又包括批判思考與問題解決、溝通、量化推理及科學化知識與技能、對道德與社會責任行為的反省、科技使用、商業核心知識的深度與廣度、職業所需綜合知識與技能 (Sweat-Guy & Buzzetto-More, 2007)。學生必須具備上述能力，始得畢業。上述能力標準不只可用來評估檔案本身的好壞，也可用來評量學生的學習表現。可做為學習目標、檔案製作、作品、反思、態度等的量測題項。

Barrett (2006) 提出評估檔案的重點應包括：具有一定程度的後設認知或反思；易於導覽；使用者導向的導覽設計；整合評估標準、作品、反思於檔案；適當的多媒體使用以及基於檔案評估者之評估標準。Burch (1999) 提到檔案評量前，教師需與學生討論檔案應包括哪些要項及其所佔分數比例。譬如可規定四份寫作成品 (佔 40 分)、兩份反思紀錄 (佔 15 分)、兩份學生觀摩的同儕作品 (佔 5 分)、兩份學生自選的相關作品 (佔 5 分)。上述標準綜合起來，可做為本研究量表設計的參考，譬如可包括檔案製作與呈現 (內容組織、內容豐富、難度、創意、導覽設計、多媒體使用等)、作品 (對學科內容了解程度)、反省思考等，這些也是教師評量、學生自評、同儕互評的重點。

學習歷程檔案評量的理念是透過檔案內容，檢視學生的學習表現，因此檔案評量的構面至少可包括檔案製作與學習表現；但這樣區分，構面太少。學習歷程檔案可擺放的項目很多，最重要的項目為學習目標、作品與反思 (Chang, Tseng, Yueh, & Lin, 2011)，因此可將這三項做為評量的構面。加上前述各個與學習歷程檔案相關的評量標準，研究者將之歸納成檔案製作、學習目標、作品、反思、態度五個構面 (如表 1)。有些標準的概念可在不同構面內出現，每個構面皆有文獻支援。

表 1 數位化檔案評量標準之構面與題項之文獻來源

| 構面   | 文獻來源  | 題 項   |
|------|---|---|
| 檔案製作 | Reckase (2002) ; Burch (1999) ;<br>Milford School District, 2009) ;<br>Sweat-Guy & Buzzetto-More (2007) ;<br>Barrett (2000) | 內容所欲傳達的意思、內容豐富性與難度、<br>組織與呈現、內容組織 (井然有序的呈現，<br>且在檔案內移動是方便的)、檔案呈現 (風格<br>特殊、富有創造力的表現方式)、內容紀錄<br>(豐富且解說詳盡)、內文鋪陳技巧 (拼字、<br>標點符號及用字遣詞均恰當)、超連結、研<br>究與資訊流暢、數位國民素養、檔案品質、<br>科技使用、導覽設計、多媒體使用 |
| 學習目標 | Sweat-Guy & Buzzetto-More (2007)  | 成長評量、批判思考、問題解決  |
| 作品   | High Plains Regional Technology<br>in Education Consortium (2005) ;<br>Milford School District (2009) ;<br>Barrett (2000)   | 成長評量、應用、創造力與創新、問題解決<br>與決策、科技操作與概念、內容豐富性與難<br>度、  |
| 反思   | Reckase (2002) ; Burch (1999) ;<br>Milford School District, 2009) ;<br>Sweat-Guy & Buzzetto-More (2007) ;<br>Barrett (2000) | 成長評量、批判思考、內容所欲傳達的意<br>思、內容豐富性與難度、組織與呈現、內容<br>組織 (井然有序的呈現，且在檔案內移動是<br>方便的)、內容紀錄 (豐富且解說詳盡)、內<br>文鋪陳技巧 (拼字、標點符號及用字遣詞均<br>恰當)   |
| 態度   | Milford School District (2009) ;<br>Sweat-Guy & Buzzetto-More (2007)  | 溝通與合作、溝通  |

學習目標的設定與學習態度的檢視是檔案的要項之一，但從表 1 可看出文獻較少，且大多未針對量測的細項做探討。另外，與網路有關的特性也可納入各構面 (譬如網路多媒體的應用可加入檔案製作、線上觀摩，及同儕互動可加入學習態度等)，成為一個題項，使成為名符其實的網路化檔案評量表。

### 三、研究方法

#### (一)研究對象

以某高職會計事務科二年級修習「計算機應用」課程的一個班級學生為研究對象，共72人。其中女生34名，男生38名(參與實驗者為79人，刪除檔案內容不全者7人)。教學實驗為期12週，每週三小時。學生是藉由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(含設定學習目標、撰寫反思、上傳作品等)、觀摩、自評及同儕互評；教學者(一位授課教師與三位線上助教)則對學生的檔案進行評量。擔任教學者的授課教師曾與研究者合作進行過網路化學習歷程檔案研究的教學實驗，熟悉網路化檔案的實施。此外，任教本研究的課程已逾十二年，對課程內容與評量都十分熟悉。另三位教學者(線上助教)皆有Office應用軟體證照，對於課程所需操作技能(版面設定及文件編輯)十分熟悉。其中兩位曾擔任過網路化學習歷程檔案研究的教學實驗助教，熟悉網路化檔案的實施。教學內容為動畫製作、網頁製作兩個單元。兩個單元的作業是數位化檔案的主要項目。此課程的特性是作品的設計與製作，且作品是數位化存儲與呈現，很適合數位化檔案的實施。

#### (二)網路化學習歷程檔案評量系統

此網路化檔案評量系統內有檔案項目按鈕及輸入資料的表單，使用者只要點選按鈕及輸入資料於表單內，即可完成個人檔案的製作。每人的檔案項目名稱、檔案格式一致，方便評量者評分。

系統功能(如圖1左邊按鈕)有：1. 檔案製作指引。2. 檔案製作區：含基本資料、學習目標設定、作品繳交、反思、其它等。3. 檔案評量區：若為教師進入則為教師評分，若為學習者登入則為自評與同儕互評。評分時，也可瀏覽被評者的檔案內容。4. 檔案成績區：可查閱學習者自評、同儕互評、教學者評分的分數，及觀摩優良作品。5. 課程說明區：含課程基本資料及教師的教案。

如圖1所示，點選系統左邊的「檔案評量區」，會出現隨機產生學習者編號供挑選。學習者自己編號旁邊會有「自評」字樣，按「自評」可進行學習者自評。學習者點選同儕的編號，可進行同儕互評，教學者點選學習者的編號，可進行教學者評分。

之後，如圖2所示，在畫面中間上面會出現一整列檔案選項(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)。評分時，可同時瀏覽學習者的檔案內容與線上參與情形(如圖2)，以做為評分的參考。

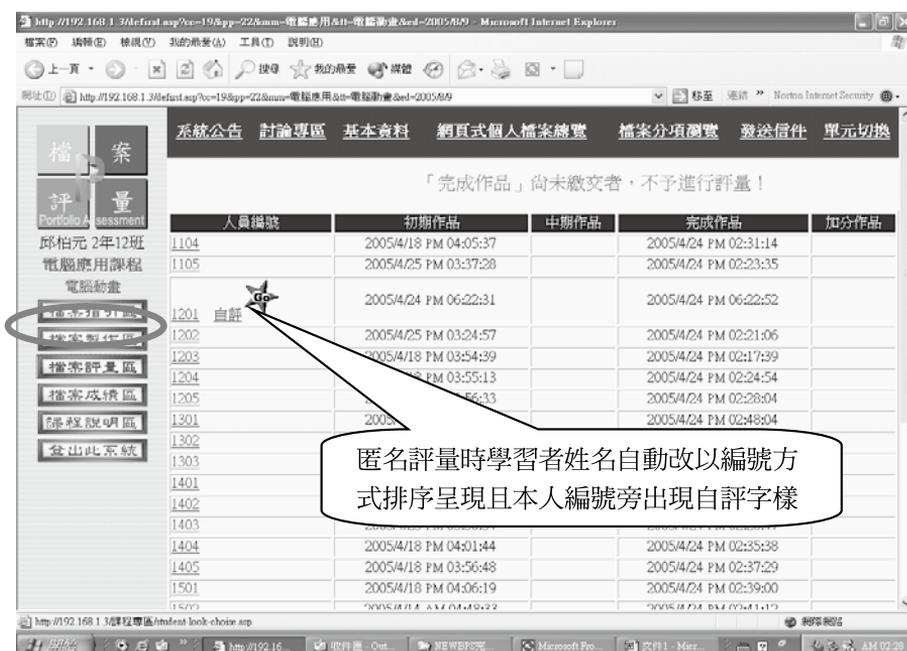


圖1 檔案評量區內之學習者編號名單

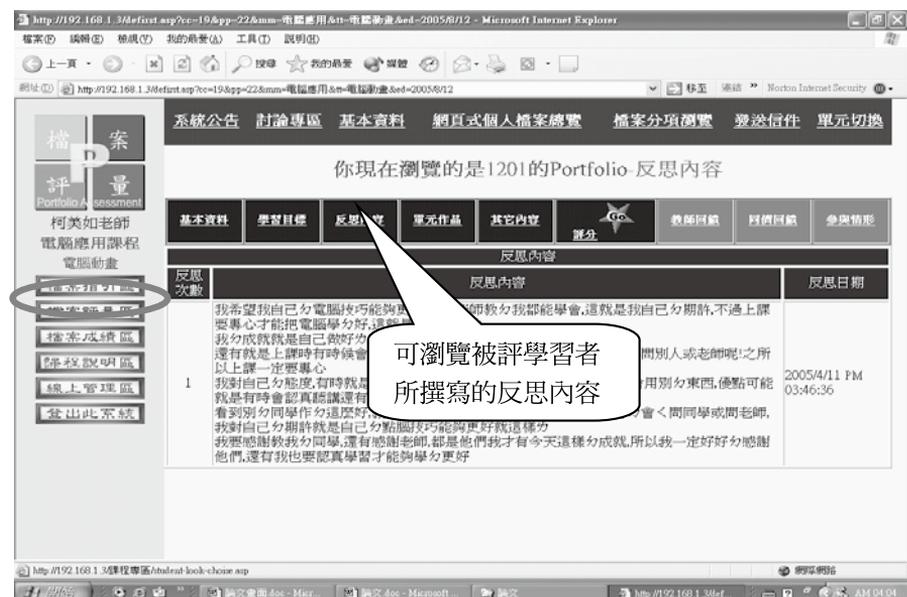


圖2 檔案評量區內之反思內容

點選中間的「評分」按鈕，如圖3所示，會另外出現一個評分表網頁。此時，可點選每一項評量標準的分數，及填寫回饋意見。

The screenshot shows a web browser window with the address bar displaying 'http://127.0.0.1/評分表.asp?id=204'. The main content area is titled '檔案製作' (Archive Making) and contains two evaluation sections:

**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                     | 1.5                   | 2                     | 2.5                   | 3                                | 3.5                   | 4                     | 4.5                   | 5                     |
| 完全無依系統要求填寫資料或亂填。      |                       |                       |                       | 對於系統要求填寫之資料有缺漏。                  |                       |                       |                       | 完整填寫系統要求填寫之資料。        |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

評分意見：

**1-2 內容恰當程度**  
 文章檔案、檔案中內容資料與檔案製作目的相關。

|                       |                       |                       |                       |                                  |                       |                       |                       |                       |
|-----------------------|-----------------------|-----------------------|-----------------------|----------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1                     | 1.5                   | 2                     | 2.5                   | 3                                | 3.5                   | 4                     | 4.5                   | 5                     |
| 毫無相關                  |                       |                       |                       | 大致符合                             |                       |                       |                       | 非常相關                  |
| <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input checked="" type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

評分意見：

圖3 檔案評量表(此為彈出視窗,故左邊按鈕看不到)

### (三)實驗程序

參考 Chang 與 Tseng (2009a, 2009b) 設計的檔案評量活動, 實驗分兩階段進行。該評量活動的內容完整, 符合網路化檔案評量該有的活動項目, 並已進行過數次教學實驗, 受試者反應良好。第一階段為預試(第一單元「動畫製作」), 第二階段為正式實施(第二單元「網頁製作」)。透過預試訓練, 可增加評分者對評量標準的熟悉度與認知上一致性, 提升正式施測的信效度。

#### 1. 第一單元(預試)

##### (1)第一週

教師在課堂上對學生說明檔案評量表及如何評分, 並經過討論與修正, 以增進學生對檔案評量的了解與參與度。正如 Tsai et al. (2002)、Bloxham 與 West (2004)、Price 與 O'Donovan (2006)、Prins et al. (2005) 提到, 讓學生參與檔案評量表的討論, 可提升甚對檔案評量的動機。另外, 由助教教導學生試用網路化檔案評量系統, 使學生能熟悉系統的操作。學生則在課堂上練習使用系統。

##### (2)第二至第五週

學生於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, 並對檔案的議題做線上討論。上傳的作品包括初期、完成、修正等不同歷程的作品。教學者透過該系統查看學生的學習狀況與評量學生的學習表現。

##### (3)第六週

教學者(一位授課教師與三位線上助教)在單元結束後數日內對每位學生的檔案做評量。學生課後則對自己的學習歷程檔案進行兩次自我評量, 以確認

再測效度。同樣地，學生於課後亦對他人的檔案做同儕互評。在考量學生負荷量之後，由教師隨機將學生分為六人一組，進行兩組成員之間的跨組匿名互評（第一與第二組、第三與第四組、第五與第六組）。某小組中的每一位成員對另一組所有成員的檔案作同儕互評，反之亦然。即每位學習者評六位同儕，這與 Cho、Schunn 與 Wilson (2006) 每位學習者評四位同儕很接近。評分時，教學者可同時瀏覽學生的檔案內容，以做為評分的參考。

## 2. 第二單元(正式實施)

### (1)第七週

教學者在課堂上對學生說明前幾週需要改進的事項，尤其加強自評及同儕互評的說明。另外，教學者整理三種評量的數據及進行信效度檢驗，並修正量表內的部分題項。最後，教學者針對修正的題項對學生做說明。

### (2)第八至第十二週

重覆第二至第六週的活動。

## (四)量表發展

本研究首先透過文獻歸納出網路化檔案評量表雛形，並與授課教師多次討論及透過三位專家協助檢視與修正，以提昇量表內容效度。之後，根據量表預試的因素分析效度檢驗及 Cronbach's  $\alpha$  值信度考驗結果，修改評量表部分題項的語意及內容。

量表包括檔案製作、學習目標、作品、反思、態度六個構面，共 26 個題項，詳見附錄。檔案製作構面的評量項目包括：1. 資料完整性；2. 內容恰當程度；3. 內容豐富程度；4. 組織與呈現。作品構面的評量項目包括：1. 作品正確性；2. 作品恰當性；3. 作品豐富性、完整性；4. 作品難度；5. 作品創意與創新；6. 對學習內容的理解；7. 作品產生的歷程。學習目標構面的評量項目包括：1. 個人學習目標設定；2. 學生達到學科所訂立的教學目標；3. 成長與成就；4. 學習目標達成程度。反思構面的評量項目包括：1. 對學習目標的反思；2. 對作品的反思；3. 學習成就的反思；4. 學習態度的反思；5. 觀摩同儕表現的反思；6. 針對回饋的反思。態度構面的評量項目包括：1. 線上觀摩、欣賞、互評、回饋；2. 資源分享；3. 線上溝通討論、知識分享、意見交換、問題解決；4. 線上多元互動數量；5. 線上多元互動品質。

計分方式為李克特五等第之間增加 0.5 的選項，為 1、1.5、2、2.5、3、3.5、4、4.5 與 5。此種方式可增加評分的方便性，減少評分者不易給分的困擾。點數越高，表示該評量標準的表現越高。每項標準的點數間皆定義不同的績效表現水準，以方便評分者能精準地給分。Reckase (2002)、余民寧 (2003) 提出將評量結果轉換為百分制 (滿分為 100 分) 的計分方式，亦即將評量所獲原

始總分除以量表總分再乘100。因此，本研究將所獲原始總分除以量表總分160再乘100，轉化為百分制的分數，供教師作為學生的學期成績參考。

### (五)量表項目分析

首先針對教學者預試量表的題項進行項目分析 (item analysis)。每一題項平均數高 (27%)、低 (27%) 兩組的  $t$  值 (決斷值) 皆達顯著水準，顯示每一題項的鑑別力足，應予保留。每一題項平均數與整體平均數的 Pearson 積差相關係數皆達顯著水準，顯示每一題項之間的一致性足夠，應予保留。

### (六)量表效度檢驗

教學者量表預試的 Bartlett 球形檢定之近似卡方分配皆達顯著，表示各題項之間有共同因素存在，適合進行因素分析。各構面的取樣適切性量數值 (Kaiser-Meyer-Olkin, KMO) 值皆大於 0.5 (表 2)，可進行因素分析。利用主成份分析法 (Principal Component Analysis, PCA) 建構效度，採斜交轉軸法 (oblique rotation) 以容許因素構面之間具相關性。因素分析結果顯示，「態度」構面內有一題項的因素負荷量小於 0.5，將此題項刪除 (王保進，2006；邱皓政，2008；吳明隆與涂金堂，2009；Tabachnick & Fidell, 2007)。第二次因素分析結果顯示 (表 2)，因素負荷量皆大於 0.5，故保留所有題項。沒有任何一個題項在不同因素上的負荷量都很高 (皆大於 0.5) 的情況，因此沒有跨因素負荷量 (cross-loading) 問題，可確保量表構面之區別效度 (discriminant validity)。

表 2 教學者量表預試之因素分析

| 構面   | 題項 | 因素負荷量 | 特徵值   | KMO 值 | 解釋變異量 (%) |
|------|----|-------|-------|-------|-----------|
| 檔案製作 | 1  | 0.58  | 1.41  | 0.64  | 55.34     |
|      | 2  | 0.63  |       |       |           |
|      | 3  | 0.61  |       |       |           |
|      | 4  | 0.55  |       |       |           |
| 學習目標 | 1  | 0.79  | 2.32  | 0.83  | 57.63     |
|      | 2  | 0.73  |       |       |           |
|      | 3  | 0.71  |       |       |           |
|      | 4  | 0.81  |       |       |           |
| 作品   | 1  | 0.83  | 4.67  | 0.88  | 65.68     |
|      | 2  | 0.73  |       |       |           |
|      | 3  | 0.84  |       |       |           |
|      | 4  | 0.81  |       |       |           |
|      | 5  | 0.79  |       |       |           |
|      | 6  | 0.76  |       |       |           |
|      | 7  | 0.82  |       |       |           |
| 反思   | 1  | 0.71  | 3.039 | 0.81  | 51.32     |
|      | 2  | 0.73  |       |       |           |
|      | 3  | 0.78  |       |       |           |
|      | 4  | 0.69  |       |       |           |
|      | 5  | 0.64  |       |       |           |
|      | 6  | 0.72  |       |       |           |
| 態度   | 1  | 0.63  | 1.816 | 0.68  | 44.36     |
|      | 2  | 0.71  |       |       |           |
|      | 3  | 0.61  |       |       |           |
|      | 4  | 0.74  |       |       |           |

最後共形成五個特徵值大於1的因素構面。各構面的解釋變異量皆大於45%，顯示各構面皆具一定效度。作品構面的解釋變異量最大，態度構面最小。

### (七) 量表信度檢驗

教學者預試所得整體Cronbach's  $\alpha$  值為0.92，屬於高信度(表3)。各構面Cronbach's  $\alpha$  值皆大於0.7，因此保留各題項。各構面中以作品構面最高，態度構面最低。

表3 教學者量表預試Cronbach's  $\alpha$  值

| 構面   | Cronbach's $\alpha$ 值 |
|------|-----------------------|
| 檔案製作 | 0.75                  |
| 學習目標 | 0.86                  |
| 作品   | 0.94                  |
| 反思   | 0.84                  |
| 態度   | 0.70                  |
| 整體   | 0.92                  |

## 四、結果與討論

### (一) 學生自我評量與教學者評量結果是否一致？有否顯著差異？

這兩種評量方法結果的Pearson積差相關及t考驗結果顯示(表4)，在作品構面及整體上，兩者具有高度相關且達顯著，表示該兩種評量方式在作品構面及整體上皆具高度一致性。但在其它三個構面上，兩者皆未達顯著相關，顯示該兩種評量方式一致性低。在假設教學者評量具有效度而做為效標的前提下，整體而言，學生自評的評量方式是具有效度的。

表4 學生自評與教學者評分結果之Pearson積差相關及t考驗

| 構面   | 相關係數<br>(顯著性)     | 評分方式  | 平均數  | 標準差  | t值<br>(顯著性) |
|------|-------------------|-------|------|------|-------------|
| 檔案製作 | 0.19<br>(0.270)   | 自評    | 4.02 | 0.57 | -0.309      |
|      |                   | 教學者評分 | 4.05 | 0.40 | (0.759)     |
| 學習目標 | 0.28<br>(0.102)   | 自評    | 3.93 | 0.53 | 2.860       |
|      |                   | 教學者評分 | 3.69 | 0.24 | (0.007**)   |
| 作品   | 0.49<br>(0.002**) | 自評    | 3.70 | 0.44 | -0.775      |
|      |                   | 教學者評分 | 3.75 | 0.26 | (0.443)     |
| 反思   | 0.32<br>(0.061)   | 自評    | 3.82 | 0.52 | -3.289      |
|      |                   | 教學者評分 | 4.11 | 0.32 | (0.002**)   |
| 態度   | -0.07<br>(0.685)  | 自評    | 3.62 | 0.54 | 1.555       |
|      |                   | 教學者評分 | 3.44 | 0.40 | (0.129)     |
| 整體   | 0.83<br>(0.001**) | 自評    | 3.82 | 0.46 | 1.058       |
|      |                   | 教學者評分 | 3.74 | 0.26 | (0.297)     |

註：1.\*\* $p < 0.01$

2. 學生兩次自評平均後再與教學者評分的平均做Pearson積差相關檢定

自評與教學者評在檔案製作、作品、態度、整體構面上皆未達顯著差異，顯示兩種評量方法的結果在這四個構面的落差不大。但在學習目標、反思

兩個構面上，自評與教學者評的結果有顯著差異，顯示兩種評量結果在這兩個構面的落差很大。

就整體平均分數而言，學生自評略高於教學者評分，顯示學生自評較為寬鬆，教學者評分較為嚴格。以各構面平均數而言，除學習目標、態度之外，其餘三個構面皆為教學者評分高於自評分數。顯示在檔案製作、作品、反思這三個構面上，教學者評分反較為寬鬆，其原因需要進一步探討。

## (二)同儕互評與教學者評分結果是否一致？有否顯著差異？

這兩種評量方法結果的Person積差相關及t考驗結果顯示(表5)，在各構面及整體上，同儕互評結果與教學者評分結果的相關性都很低，且都未達顯著，表示兩者評分一致性非常低。在假設教學者評量具有效度而做為效標的前提下，同儕互評與教學者評分結果的相關度很低，顯示同儕互評的效度很低。兩種評量方法的結果在檔案製作、作品、整體構面上皆未達顯著差異，顯示兩種評量方法的結果在這三個構面的落差並不大。但在學習目標、反思、態度三構面上，兩種評量方法的結果達顯著差異，顯示兩種評量結果在這三個構面上落差很大。

就整體平均分數而言，同儕互評高於教學者評分，顯示同儕互評較為寬鬆，教學者評分較為嚴格。以各構面平均數而言，除檔案製作、反思之外，其餘三個構面皆為同儕互評評分高於教學者評分。顯示在學習目標、作品、態度構面上，同儕互評較為寬鬆。

表5 同儕互評與教學者評分結果之Person積差相關及t考驗

| 構面   | 相關係數<br>(顯著性)   | 評分方式  | 平均數  | 標準差  | t值<br>(顯著性) |
|------|-----------------|-------|------|------|-------------|
| 檔案製作 | 0.28<br>(0.097) | 互評    | 3.98 | 0.57 | -0.725      |
|      |                 | 教學者評分 | 4.05 | 0.40 | (0.474)     |
| 學習目標 | 0.29<br>(0.082) | 互評    | 3.92 | 0.25 | 3.636       |
|      |                 | 教學者評分 | 3.69 | 0.24 | (0.001**)   |
| 作品   | 0.18<br>(0.273) | 互評    | 3.77 | 0.35 | 0.244       |
|      |                 | 教學者評分 | 3.75 | 0.26 | (0.809)     |
| 反思   | 0.03<br>(0.854) | 互評    | 3.92 | 0.31 | -2.445      |
|      |                 | 教學者評分 | 4.11 | 0.32 | (0.020*)    |
| 態度   | 0.22<br>(0.190) | 互評    | 3.72 | 0.38 | 2.684       |
|      |                 | 教學者評分 | 3.44 | 0.40 | (0.011*)    |
| 整體   | 0.28<br>(0.097) | 互評    | 3.86 | 0.29 | 1.724       |
|      |                 | 教學者評分 | 3.74 | 0.26 | (0.094)     |

註：1.\* $p < 0.05$ , \*\* $p < 0.01$

2.學生互評平均後再與教學者評分之平均做Pearson積差相關的檢定

## (三)學生自評與同儕互評結果是否一致？有否顯著差異？

這兩種評量方法結果的Pearson積差相關及t考驗結果顯示(表6)，學生自評與同儕互評結果之間的相關性低且未達顯著，表示這兩種評量方法的結果一

致性低。就構面而言，作品構面的相關性最高且達顯著（僅此構面達顯著），反思構面最低且未達顯著，顯示反思構面的評審較不易達成一致。自評與互評結果在各構面、整體上，皆未達顯著差異，顯示兩種評量方法結果的落差並不太大。就整體平均分數而言，同儕互評略高於學生自評，顯示同儕評量較為寬鬆，學生自評較為嚴格。

以各構面平均數而言，除檔案製作、學習目標之外，其餘三個構面皆為同儕互評分數高於自評分數。顯示在作品、反思、態度構面上，同儕互評較自評為寬鬆。由於評量者間存在不同的主觀判斷與意見，可能導致評量結果的誤差而影響信度（余民寧，2003）。同儕之間因主觀意見不同，確實產生了評量結果不一致的情況。

表 6 學生自評與同儕評量結果之 Pearson 積差相關及 t 考驗

| 構面   | 相關係數<br>(顯著性)    | 評分方式 | 平均數  | 標準差  | t 值<br>(顯著性) |
|------|------------------|------|------|------|--------------|
| 檔案製作 | 0.24<br>(0.168)  | 自評   | 4.02 | 0.57 | 0.304        |
|      |                  | 同儕互評 | 3.98 | 0.30 | (0.763)      |
| 學習目標 | 0.13<br>(0.459)  | 自評   | 3.93 | 0.53 | 0.077        |
|      |                  | 同儕互評 | 3.92 | 0.25 | (0.939)      |
| 作品   | 0.34<br>(0.040*) | 自評   | 3.70 | 0.44 | -0.642       |
|      |                  | 同儕互評 | 3.77 | 0.35 | (0.525)      |
| 反思   | 0.11<br>(0.522)  | 自評   | 3.82 | 0.52 | -0.903       |
|      |                  | 同儕互評 | 3.92 | 0.31 | (0.373)      |
| 態度   | 0.23<br>(0.179)  | 自評   | 3.62 | 0.54 | -0.816       |
|      |                  | 同儕互評 | 3.72 | 0.38 | (0.420)      |
| 整體   | 0.30<br>(0.227)  | 自評   | 3.82 | 0.46 | -0.428       |
|      |                  | 同儕互評 | 3.86 | 0.28 | (0.672)      |

註：1.\* $p < 0.05$

2. 學生兩次自評平均後再與同儕互評的平均做 Pearson 積差相關檢定

#### (四) 三種評量是否足以測出學習成就(即檔案分數是否與測驗成績一致)？

根據 Gadbury-Amyot 等人(2003)，學生測驗分數可當做外在效標，其與檔案評量分數的相關性越高，表示檔案評量分數的效度越高。本研究檔案評量分數與學生測驗成績的 Pearson 積差相關結果顯示(表 7)，教學者(一位授課教師與 3 位線上助教的平均)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一致性，也顯示檔案評量表足以檢測出學習成就。此結果符合 Gelinias (1998) 提列的「檔案評量分數與被評者的考核成果表現有正相關」；其中以學習目標構面的相關性最高，其次為檔案製作構面，態度構面最低，顯示學習目標與學習成就最容易達成一致。學生的檔案自評分數與其測驗成績結果顯示具高度一致性且達顯著，個別構面亦具高度一致性且達顯著。同儕互評分數與測驗成績的一致性低，其中僅學習目標具一致性。

各構面整體觀之，以作品構面的相關性最高，檔案製作構面居次，態度構面最低且未達顯著。顯示作品最能檢測出學習成就，而態度最難評測出學習成就。

表7 檔案分數與測驗成績之 Pearson 積差相關

| 構面   | 相關係數(顯著性)      |                |              |                |
|------|----------------|----------------|--------------|----------------|
|      | 教學者評           | 學生自評           | 同儕互評         | 整體             |
| 檔案製作 | 0.78(0.000***) | 0.67(0.000***) | 0.26(0.134)  | 0.50(0.002**)  |
| 學習目標 | 0.79(0.000***) | 0.64(0.003**)  | 0.38(0.024*) | 0.48(0.003**)  |
| 作品   | 0.67(0.000***) | 0.68(0.000***) | 0.32(0.059)  | 0.59(0.000***) |
| 反思   | 0.63(0.000***) | 0.66(0.000***) | 0.24(0.161)  | 0.49(0.002**)  |
| 態度   | 0.59(0.000***) | 0.61(0.006***) | 0.17(0.322)  | 0.23(0.179)    |
| 整體   | 0.84(0.000***) | 0.71(0.023**)  | 0.15(0.403)  | 0.55(0.001***) |

註：\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$

### (五) 三種評量方法，何者信度較高何者較低？

正式教學實驗之三種評量方式 Cronbach  $\alpha$  值如表8，教師評量最高，同儕互評最低。各構面中以作品構面最高，態度構面最低。正式施測屬於高信度，代表各項指標之間具有高度的內部一致性。

表8 三種評量方式信度指標 Cronbach's  $\alpha$  值

| 構面   | Cronbach's $\alpha$ 值 |      |      |
|------|-----------------------|------|------|
|      | 教學者評                  | 學生自評 | 同儕互評 |
| 檔案製作 | 0.77                  | 0.75 | 0.70 |
| 學習目標 | 0.88                  | 0.82 | 0.75 |
| 作品   | 0.94                  | 0.92 | 0.90 |
| 反思   | 0.89                  | 0.88 | 0.87 |
| 態度   | 0.74                  | 0.72 | 0.71 |

## (六) 綜合討論

### 1. 三種評量方法之間的一致性與差異性

三種評量方法結果之間的一致性與差異性比較如表9。三種評量方式在檔案製作構面上不具一致性但亦無差異；在學習目標、反思構面上都不具一致性且有兩個「兩種評量結果具差異」的情況；在反思、態度構面上都不具一致性且有一個「兩種評量結果具差異」的情況。較特別的是，三種評量方式在作品構面上，都沒有差異，但有兩個「兩種評量結果具一致性」的情況。整體而言，僅有一個「兩種評量結果具一致性」的情況，即自評與教學者評兩者具一致性；自評與同儕互評之間、同儕互評與教學者評之間，都不具一致性。此結果與 Sadler & Good ((2006)、Sung 等人(2005)的研究結果「學生自評、同儕互評、教師評量的一致性高」不同。這可能是因評量對象、評量的作品、學生所受訓練、評量標準、是否網路評量環境等不同所致。

自評與教學者評具一致性的結果與這與 Sung 等人(2005)、Sadler & Good (2006)的研究結果「學生自評與教師評量的一致性高」吻合，但與 Knowles, Holton III 與 Swanson (2005) 的看法不一致。同儕互評與教學者評不具一致性的

結果與Chen(2010)的研究一致，但與大部分的研究「同儕互評是具有效度」的結果不同(卓宜青等，2001；楊國鑫，2002；劉勝鈺，2003；Bouzidi & Jaillet, 2009; Cho, Schunn, & Wilson, 2006; Liu, 2002; Sadler & Good, 2006; Sung et al., 2005; Topping, 1998; Tsai & Liang, 2009; Tseng & Tsai, 2007)，這可能是同儕互評的方式、學生的教育層級、互評的數量、學生對互評的認知不同所造成。自評與同儕互評不具一致性的結果與Sadler & Good(2006)、Sung等人(2005)的研究結果「學生自評與同儕互評的一致性高」不同。可能原因如前所述。

表9 不同評量方式評分結果之間相關與差異顯著性

| 構面   | 相關與差異之顯著性 |           |           |
|------|-----------|-----------|-----------|
|      | 學生自評—同儕互評 | 學生自評—教學者評 | 同儕互評—教學者評 |
| 檔案製作 |           |           |           |
| 學習目標 |           | D**       | D**       |
| 作品   | C*        | C**       |           |
| 反思   |           | D**       | D*        |
| 態度   |           |           | D*        |
| 整體   |           | C**       |           |

註：1.星號表示顯著性程度，\* $p<0.05$ ，\*\* $p<0.01$

2.空白表示未達顯著水準

## 2. 三種評量方法的分數比較

三種評量方法的分數與變異數分析如表10。在三種評量方式評分中，同儕互評最高(平均數=3.86)，其次為學生自評(平均數=3.82)，教學者評量最低(平均數=3.74)，且達顯著差異水準( $p<0.05$ )。顯示教學者評量最嚴格、其次為學生自評，而同儕互評最寬鬆。這與Lin、Liu與Yuan(2001a, 2001b)、劉旨峰、林珊如與袁賢銘(2000)、Sadler & Good(2006)研究結果「教師評分最嚴格，同儕互評次之，學生自評最寬鬆」不全一致。就評量構面而言，三種評量方式結果之間在三個構面上有顯著差異，在二個構面上無差異。反思構面的評分差異最大(效果量最大)且達顯著差異水準，其次為態度構面。作品構面的差異最小(效果量最小)。顯示三種評量方式在反思的評分結果上，最不容易相同，作品的評量結果最容易接近。

表10 教學者評、學生自評與同儕互評之分數與變異數分析

| 構面   | 教學者<br>評分 | 學生自<br>評分數 | 同儕互<br>評分數 | F值   | 顯著值     | 效果量<br>(順序) | 事後比較                       |
|------|-----------|------------|------------|------|---------|-------------|----------------------------|
| 檔案製作 | 4.05      | 4.02       | 3.98       | 0.22 | 0.058   | 3.18(5)     |                            |
| 學習目標 | 3.69      | 3.93       | 3.92       | 5.31 | 0.006** | 3.86(3)     | 學生自評>教學者評**<br>同儕互評>教學者評** |
| 作品   | 3.75      | 3.70       | 3.77       | 0.36 | 0.063   | 3.13(6)     |                            |
| 反思   | 4.11      | 3.82       | 3.92       | 4.78 | 0.010*  | 4.33(1)     | 同儕互評>學生自評*<br>教學者評>同儕互評*   |
| 態度   | 3.44      | 3.62       | 3.72       | 3.51 | 0.033*  | 4.14(2)     | 同儕互評>學生自評>教學者評*            |
| 整體   | 3.74      | 3.82       | 3.86       | 1.21 | 0.042*  | 3.54        | 同儕互評>學生自評*<br>學生自評>教學者評*   |

註：\* $p<0.05$ ，\*\* $p<0.01$ ，\*\*\* $p<0.001$

至目前為止，已有一些透過紙本式與網路化檔案進行教學者評、學生自評、同儕互評信、效度的研究。儘管這些研究的結果不一，有部分是信、效度足夠的，也有部分證明信、效度不足。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)。Falchikov與Goldfinch(2000)的研究顯示，使用設計良好的評量標準來判斷，整體上，大學生同儕互評與教師評量結果頗為接近；但從各個評量構面來看，大學生同儕互評與教師評量結果不一定在每一個構面都接近。這與本研究結果有些許不同，本研究除在整體上，同儕互評與教師評量結果不一致；在各個評量構面上，也不一致。然而Bouzidi與Jaillet(2009)的研究顯示，在各個評量構面上，大學生同儕互評與教師評量結果具一致性是可能的，這與本研究結果有部分吻合。但無論如何，要讓網路化檔案教學者評、學生自評、同儕互評具有足夠信、效度是可以做到的。不管網路化學生自評檔案或同儕互評檔案的信、效度如何，都不宜讓它來取代教師評量或網路化教師評檔案。

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；而增加評分者人數是可增加評分信度。Sung等人(2003)及Sung等人(2005)研究發現，若同儕互評的機會或次數增多，則學生評量結果會越接近教師評量結果。這個發現顯示，經過較多次同儕互評的機會，學生評量的能力提升，評量信、效度也跟著增加。但許多研究提到的，同儕互評耗時費力；尤其當超過負荷時，容易影響評量的結果(Ballantyne et al., 2002; Davies, 2000; Lin et al., 2001a; Miller, 2003; Sung et al., 2003; Sung et al., 2005; Topping et al., 2000; Tsai et al., 2002; Tsai & Liang, 2009; Wen, & Tsai, 2006; Yang & Tsai, 2010)。因此，學生參與自評與同儕互評時，教師須兼顧學生評量的練習機會與負荷，不因其中一項而影響到評量結果。

## 五、結論與建議

在三種評量方式結果之間差異上，教學者評、學生自評、同儕互評三種評量結果有顯著差異，其中教學者評量最為嚴格，其次為學生自評，同儕互評最為寬鬆。Sadler & Good(2006)以高中生為研究對象，結果顯示學生對於同儕的作品打分數會較低(嚴)，但對於自己的作品打分數會較高(鬆)。此結果與本研究相反。在三種評量結果之間一致性上，僅自評與教師評之間具一致性；自評與同儕互評之間、同儕互評與教學者評之間，都不具一致性。就評量構面而言，反思構面的評分差異最大，作品構面的評分差異最小。劉旨峰、林珊如與袁賢銘(2000)及Liu(2002)等人的研究亦指出，教學者分數較同儕互評與學生自評嚴格。若三種評量方式皆欲納入學生的期末成績，所佔比例由大至小應

為教學者評、學習自評、同儕互評的分數，以符合評分信度與嚴謹性的順序。以各構面來看，作品構面的信、效度最高；自評與教學者評之一致性亦以作品構面最高，三種評量方式的結果差異以作品最小。態度構面的信、效度最低，自評與教學者評之一致性亦以態度構面最低。此結果顯示，作品比較容易打分數，而態度則較難。

在三種評量方式結果與測驗成績一致性上，教學者評分結果與測驗成績在作品構面及整體上具高度一致性，學生自評與其測驗成績亦具高度一致性；而同儕互評結果與其測驗成績不具一致性。因此，教學者若欲將學生自評與同儕互評的分數納入最後的學期成績，前者所佔比例可較後者為高，以減少同儕互評所造成的負面影響。但自評所牽涉到的是較基本的認知能力，而同儕互評比自評能發展更高層次認知的學習 (Anderson & Krathwohl, 2001; Bouzidi & Jaillet, 2009)。因此，若不過度強調評量的信、效度，則同儕互評可以是較自評優先考量的學習活動。若能同時有自評與同儕互評活動更好，且評量信、效度能增加 (Bouzidi & Jaillet, 2009)，因為學習者有較多且不同層次的評量活動。

檔案評量的工作包括評估檔案內容與評量學生學習成就，目的主要是希望透過評估檔案內容來了解學生的學習成就。學生的檔案本身製作得好 (內容好、品質佳) 並不一定代表學生對學科知識的了解足夠 (學習成就高、學業成績佳)。反之亦然，學業成績好的學生，檔案不一定就製作得好。因此，兩者代表的意涵不同，但有些人會將兩者混淆，以為兩者是同一件事。如果將製作出的檔案亦當作是學生的一項學習成就 (譬如把檔案製作當作專題作業之一)，評估檔案優劣的同時也等於是評量學生學習成就高低的其中一項。相反地，如果要把評估檔案內容的結果就視為學生學習成就表現，則應再加入其它能代表學習成就的事項，譬如考試成績表現。本研究為使檔案評量工作不過於複雜，檔案內容並未含各項考試成績表現，因此量表也未含各項考試的成績。因此，學生的學習成就應由檔案評量分數再加上考試成績。

評量工具本身、評量程序與情境、受測者當時身心反應、外在效標良窳、學生能力差異程度等因素均可能影響評量結果的信度 (余民寧, 2003)。學生對檔案有充分的了解、學生的檔案內容與格式一致可提升評分的信度 (Oskay, Schallies, & Morgil, 2008)。本研究努力對評量工具 (網路化檔案評量標準)、評量程序與情境、外在效標的良窳、學生對檔案的了解、對學生檔案內容的要求等可控制的部分盡量力求嚴謹。但面對受試者受測當時的身心反應在掌控上仍有不足，有可能影響評量結果，此為本研究限制之一。未來進行更進一步的相關研究時，對受試者受測當時的身心反應更加注意，在檔案評量前給受測者足夠的評量標準解說及評量範例展示。

大樣本會有更好的評量信度 (Gadbury-Amyot, 2003)。本研究屬小樣本，信度難免受到影響。Johnson 等人 (2000) 提到，當增加第三個評分者之後，評

分的信度自然會提高。但評分的一致性會因特定評分者而產生相當大的差異，所以給予評分者足夠的訓練使其對評分標準有相同的認知是有必須的。Derham 與 Diperna (2007)、Sulzen, Young 與 Hannifin (2008) 的研究就建議，評分者接受足夠的訓練可提升評分者之間的一致性。訓練有素的評分者、具體明確的評量標準可提升評分的信度 (Gadbury-Amyot, 2003; Oskay, Schallies, & Morgil, 2008)。有鑑於此，以提高評分者人數的方式來降低誤差變異的方式可能遠不及加強評分者訓練對評分者一致性的影響。

## 參考文獻

- 王保進 (2006)。中文視窗版 SPSS 與行為科學研究。台北市：心理。
- 余民寧 (2003)。教育測驗與評量：成就測驗與教學評量 (二版)。台北市：心理。
- 卓宜青 (2001)。網路化學習歷程檔案系統及同儕評量。未出版之碩士論文，國立交通大學資訊科學系，新竹市。
- 吳明隆、涂金堂 (2008)。SPSS 與統計應用分析。台北市：五南。
- 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。台北市：五南。
- 張振成 (1997)。教學評量的新趨勢：實作評量與檔案評量。中等教育，48(6)，90-94。
- 曾吉雄 (1998)。新式評量：檔案評量。教育資料文摘，48(1)，133-137。
- 楊國鑫 (2002)。推廣網路同儕互評系統於高級工業職業學校教學課程之研究。未出版之碩士論文，國立交通大學資訊科學系，新竹市。
- 鄒慧英 (2000)。檔案於國語科教學評量之應用。上網日期：2009年8月7日，檢自：  
<http://www.nknu.edu.tw/~edu/item/item4-article.file/item4-article19.htm>
- 劉旨峰、林珊如、袁賢銘 (2002)。網路化學習歷程檔案之學習成效分析及未來展望。在大葉大學主辦，第三屆電子化企業經營管理理論暨實務研討會，彰化縣。
- 劉勝鈺 (2003)。使用資訊科技學習數學：以網路同儕互評為例。未出版之碩士論文，國立交通大學網路學習學程碩士班，新竹市。
- Anderson, L., & Krathwohl, D. (2001). *A taxonomy for learning teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Wesley Longman.
- 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.
- Barrett, H. (2006). Using electronic portfolios for classroom assessment. *Connected Newsletter*, 13(2), 4-6.
- 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.
- Burch, C. B. (1999). Inside the portfolio experience: The student's perspective. *English Education*, 32(1), 34-49.
- Chang, C. C. (2008). Enhancing self-perceived effects using web-based portfolio assessment. *Computers in Human Behavior*, 24(3), 1753-1771.

- 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), 25-37.
- Chang, C. C., & Tseng, K. H. (2009b). Use and performances of web-based portfolio assessment. *British Journal of Educational Technology, 40*(2), 358-370.
- Chang, C. C., Tseng, K. H., Yueh, H. P., & Lin, W. C. (2011). Consideration factors and adoption of type, tabulation and framework for creating e-portfolios. *Computers & Education, 56*(2), 452-465.
- Chen, C. H. (2010). The implementation and evaluation of a mobile self- and peer-assessment system. *Computers & Education, 55*(1), 229-236.
- Cho, K., Schunn, C., & Wilson, R. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology, 98* (4), 891-901.
- Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Training International, 37*(4), 346-355.
- Derham, C., & Diperna, J. (2007). Digital professional portfolios of preservice teaching: An initial study of score reliability and validity. *International Journal of Technology and Teacher Education, 15*(3), 363-381.
- 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.
- Gadbury-Amyot, C., Kim, J., Palm R. L., Mills, E., Noble, E., & Overman, P. (2003). Validity and reliability of portfolio assessment of competency in a baccalaureate dental hygiene program. *Journal of Dental Education, 67*(9), 991-1002.
- Gadbury-Amyot, C. (2003). *Validity and reliability of portfolio assessment of competency in a baccalaureate dental hygiene program*. Unpublished doctoral dissertation, University of Missouri-Kansas City.
- 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, Ohio.
- High Plains Regional Technology in Education Consortium (2005). *Educational technology MA online program rubric: EPortfolio*. Retrieved February 22, 2010, from [http://www.msarreola.com/Master%20Canvas/Assessment%20&%20Evaluation\\_files/eport\\_rubric.pdf](http://www.msarreola.com/Master%20Canvas/Assessment%20&%20Evaluation_files/eport_rubric.pdf)
- Jenkins, M. (2004). Unfulfilled promise: Formative assessment using computer-aided assessment. *Learning and Teaching in Higher Education, 1*, 67-80.
- Johnson, R. L., McDaniel, F., & Willeke, M. J. (2000). Using portfolios in program evaluation: An investigation of interrater reliability. *American Journal of Evaluation, 21*(1), 65-81.
- Knowles, M. S., Holton III, E. F., & Swanson, R. A. (2005). *The adult learner* (6th ed.). Burlington, MA: Elsevier.
- Lin, S. J., Liu, Z. F., & Yuan, S. M. (2001a). Web-based peer assessment: Attitude and achievement. *IEEE Transactions on Education, 44*(2), 211-223.

- Lin, S. J., Liu, Z. F., & Yuan, S. M. (2001b). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, 17(4), 420-432.
- Liu, Z. F. (2002). *Studies of networked peer assessment*. Unpublished doctoral dissertation. National Chiao Tung University, Hsinchu.
- Liu, Z. F., Lin, 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.
- Milford School District, (2009). *Information, communication, technology (ICT) portfolio assessment*. Retrieved February 22, 2010, from <http://www.milfordschools.net/mms/documents/ePortfolio%20Rubric%202009.pdf>
- Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment & Evaluation in Higher Education*, 28(4), 383-394.
- Oskay, O., Schallies, M., & Morgil, I. (2008). A closer look at findings from recent publication. *H. U. Journal of Education*, 35, 263-272
- 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, UK: Routledge.
- 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.
- Reckase, M. A. (2002). *Portfolio define*. Paper presented at the Workshop of Portfolio Assessment, Taipei, Taiwan.
- Russell, J. D., & Butcher, C. (1999). Using portfolios in educational technology courses. *Journal of Technology and Teacher Education*, 7(4), 279-289.
- Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11 (1), 1-31.
- Sulzen, J., Young, M., & Hannifin, R. (2008). Reliability and validity of an ecologically-grounded student teacher electronic portfolio rubric. In K. McFerrin et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2008* (pp. 153-159). Chesapeake, VA: AACE.
- 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, 187-202.
- 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(4), 331-334.
- Sweat-Guy, R., & Buzzetto-More, N. (2007). A comparative analysis of common e-portfolio platforms and available features. *Issues in Informing Science and Information Technology Education*, 5(1). 327-342.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.
- 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.

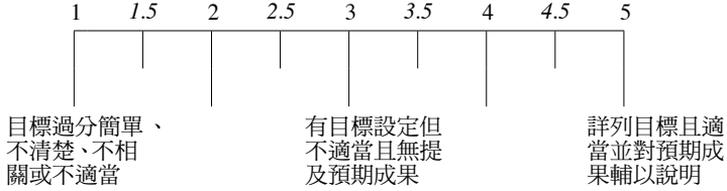
- Tsai, C. C., & Liang, J. C. (2009). The development of science activities via on-line peer assessment: The role of scientific epistemological views. *Instructional Science*, 37, 293-310.
- Tsai, C. C., Lin, S. J., & Yuan, S. M. (2002). Developing science activities through a networked peer assessment system. *Computers & Education*, 38, 241-252.
- Tseng, S. C., & Tsai, C. C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education*, 49 (4), 1161-1174.
- Wen, L. M. C., & Tsai, C. C. (2006). University students' perceptions of and attitudes toward (Online) peer assessment. *Higher Education*, 51, 27-44.
- Yang, Y. F., & Tsai, C. C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction*, 20, 72-83.



2.學習目標 (25/160)

2-1 個人學習目標設定：

學生針對自身情況(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

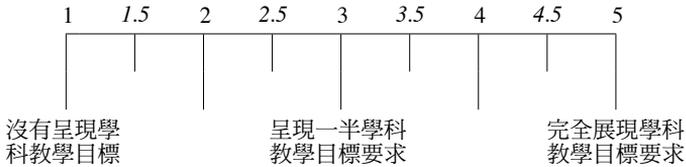


評論： \_\_\_\_\_

2-2 學科教學目標：

學生達到學科所訂立的教學目標

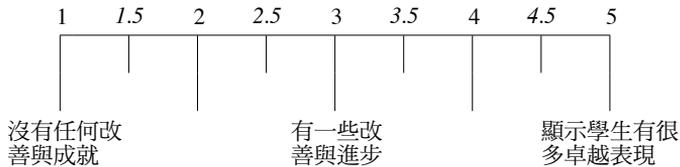
如：以「高職計算機應用」學科為例，學科教學目標有展現操作套裝軟體的基本知識與能力、了解資訊網路之建構與應用、科技新知的發展在電子計算機應用上的效益、展現團體合作精神、正確使用電子計算機的態度等相關證據



評論： \_\_\_\_\_

2-3 成長與成就：

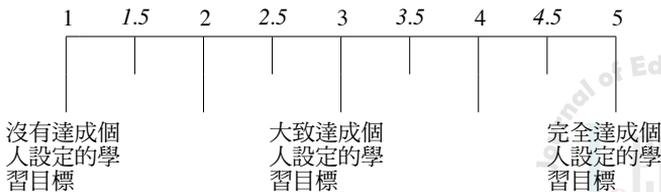
學生顯示在學習內容、成果、態度等方面的改善與進步



評論： \_\_\_\_\_

2-4 學習目標達成程度：

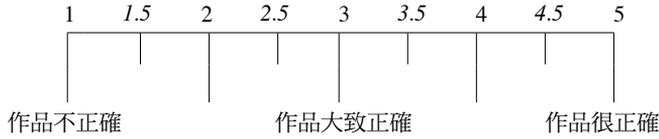
學生達成個人學習目標的程度



評論： \_\_\_\_\_

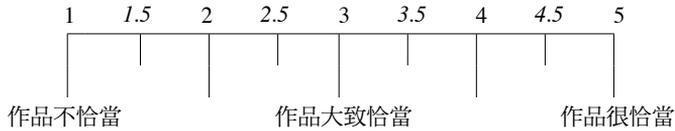
3.作品 (40/160)

3-1 作品正確性：  
學生作品的正確性



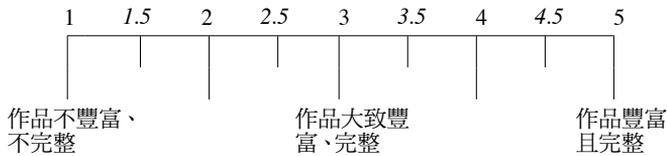
評論： \_\_\_\_\_

3-2 作品恰當性：  
學生作品的恰當性



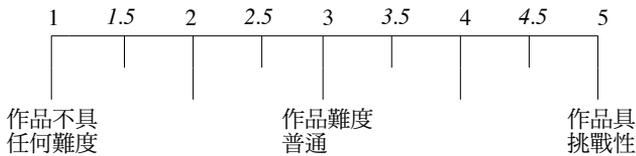
評論： \_\_\_\_\_

3-3 作品豐富性與完整性：  
學生作品的豐富性與完整性



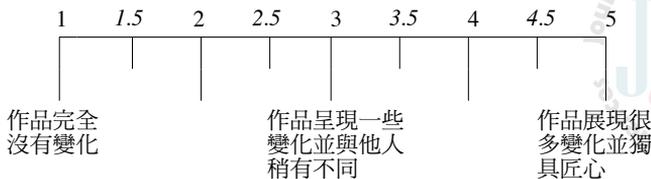
評論： \_\_\_\_\_

3-4 作品難度：  
學生作品的難度

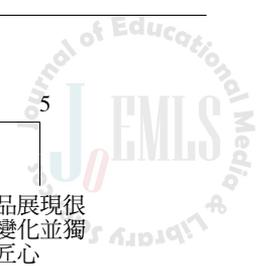


評論： \_\_\_\_\_

3-5 作品創意與創新：  
作品展現創意與創新

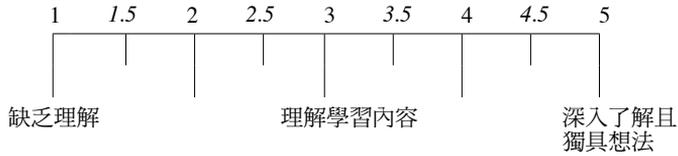


評論： \_\_\_\_\_



3-6 對學習內容的理解：

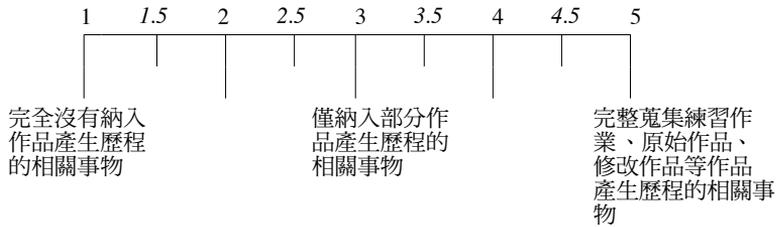
學生展現對學習內容的理解程度



評論：

3-7 作品產生的歷程：

學生上傳其它可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

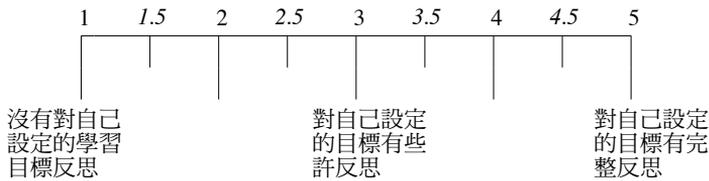


評論：

4. 反思 (35/160)

4-1 對學習目標的反思：

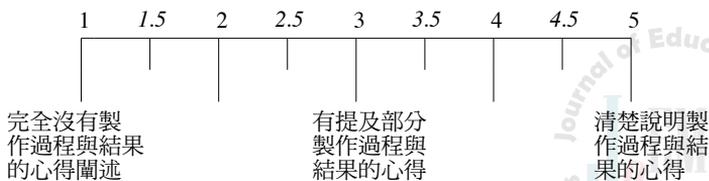
對自己設定的學習目標進行反思



評論：

4-2 對作品的反思：

學生對作品產生過程與結果的心得闡述

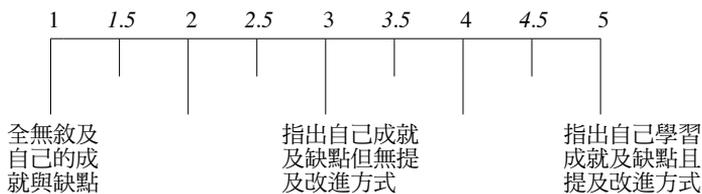


評論：



4-3 學習成就的反思：

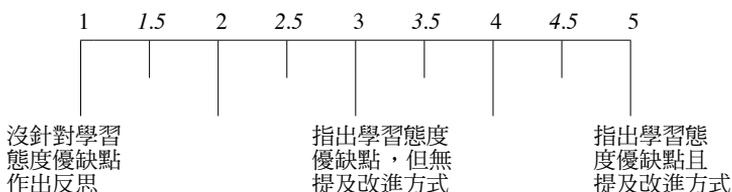
學生在反思中舉出自己的學習成就、缺點、進步情形



評論： \_\_\_\_\_

4-4 學習態度的反思：

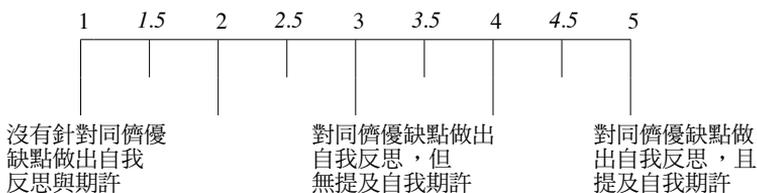
學生對自己的學習態度優缺點、進步情形進行反省思考



評論： \_\_\_\_\_

4-5 觀摩同儕表現的反思：

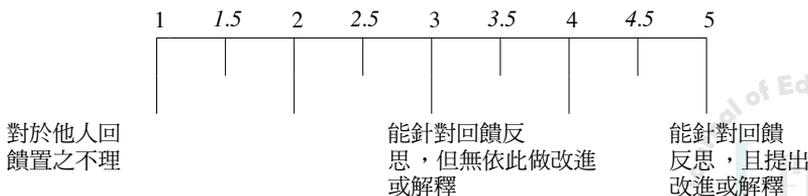
觀察同儕表現(含檔案製作、學習目標、作品、反思、學習成就、學習態度等)所做自我期許(註：網路化檔案評量系統提供同儕觀摩與互評)



評論： \_\_\_\_\_

4-6 針對回饋的反思：

對於他人(教師、同儕、家長等)所給予回饋進行反思，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)



評論： \_\_\_\_\_







# Comparisons of Reliability and Validity between Teacher-Assessment, Student Self-Assessment, and Peer-Assessment in a Web-based Portfolio Assessment Environment

**Chi-Cheng Chang\***

Professor  
Department of Technology Application & Human Resource Development  
National Taiwan Normal University  
Taipei, Taiwan  
E-mail: samchang@ntnu.edu.tw

**Ming-Fun Wu**

Teacher  
Department of Data Processing  
Taipei City Sung-Shan Junior High School  
Taipei, Taiwan  
E-mail: wumingfang@ms26.url.com.tw

## **Abstract**

*This study compares the reliability and validity among instructor assessment, student self-assessment and peer-assessment under the Web-based portfolio assessment environment. The research samples consist of 72 students who take "Computer Application" course of the second grade in some vocational high school. The students use the Web-based portfolio assessment system for creating their portfolio, emulating peers' portfolio, and performing self-assessment and peer-assessment. The teacher uses the system for reviewing and assessing students' learning performances. In regard to the difference among the three assessment approaches, there exists a significant difference among the three approaches. Instructor assessment is most rigorous, while peer-assessment is most slack. In regard to the correlation among the three assessment approaches, there exists a significant consistency between self-assessment and instructor assessment. There exist inconsistencies between self-assessment and peer-assessment as well as peer-assessment and instructor assessment. In regard to the correlation among the three assessment results and test score, there exist a high consistency between instructor assessment and test score. As well, there exist a high consistency between self-assessment and test score. However, the consistency between peer-assessment and test score is inadequate.*

**Keywords:** Web-based portfolio; Web-based portfolio assessment; Self-assessment; Peer-assessment

---

\* Principal author for all correspondence.

## **SUMMARY**

### **Introduction**

There are three common methods (i.e. teacher-assessment, student self-assessment and peer-assessment) used in Web-based portfolio assessment, which teachers may adopt in order to serve pedagogical needs. Although teacher-assessment plays an indispensable role, it has been shown that the authenticity of portfolio assessment is readily enhanced when self- and peer-assessment are included.

Some researchers have called into question the reliability of self- and peer-assessment, while others have been concerned with whether or not the assessment results are able to thoroughly reflect learning achievements. In that case, the issues regarding teacher-, self-, and peer-assessment validity and reliability are of great importance.

This study explored the reliability and validity of the three assessment methods (teacher-, peer- and self-assessment) on Web-based portfolio assessment. The research questions dealt with the consistency as well as the reliability and validity of the three methods. In addition, we made a further attempt to examine the extent to which each method was able to reflect learning achievements and what grading standards were adopted.

### **Method**

The participants were 72 students, including 34 males and 38 females, in a computer course at a senior high school. The duration of the study was 12-weeks with 3 hours of class time each week. A Web-based portfolio assessment system was developed, which allowed the students to perform portfolio creation (goal setting, reflection, artifact, etc.), inspection, self-assessment and peer-assessment. The teachers (one instructor and three teacher assistants), employed this system to assess students' portfolio work. The computer course had two units, "Animation Creation" and "Website Creation", covering design skills and web page creation abilities. Furthermore, in the digital-based setting, students were required to electronically save and present their work, which was ideal for the practice of Web-based portfolio assessment.

The scale was comprised of six aspects (Portfolio Creation, Learning Goal, Artifact, Reflection, Attitude, Other), with a total of 27 items. Scores were assigned on a scale of 1 to 5, with 0.5-increments to allow greater differentiation of assessment results. The higher the score, the better the portfolio. The pre-test assessment scale was measured using item analysis. The *t*-values for both high-scoring (27%) and low-scoring groups (27%) achieved significance level, which implied that each rubric had good discrimination capability and should be

reserved. The Pearson's correlation between each rubric mean and overall mean was significant, indicating an acceptably high level of consistency among rubrics.

The Kaiser-Meyer-Olkin (KMO) values for each rubric aspect were greater than 0.5, implying that factor analysis could be applied. One approach in factor analysis—Principal Component Analysis—was further used to construct validity. The results of factor analysis indicated that factor loadings for the rubric aspect Attitude were lower than 0.3; thus the rubric was deleted. However, the results of another factor analysis revealed that all rubrics yielded factor loadings greater than 0.3. Therefore, all rubrics were kept.

Finally, five constructs of the Web-based portfolio assessment scale with eigenvalues higher than 1 were refined. The explained variances of each rubric aspect were greater than 45%, which meant that all rubrics were acceptably valid. The overall Cronbach's was 0.92 for the scale, suggesting a high reliability. All rubric aspects were higher than 0.7.

## **Results and Discussion**

### **1. Are self-assessment results consistent with teacher-assessment results? Are there significant differences between the two assessment results?**

Overall, the two assessment methods were shown to be correlated and significant, which suggested a high degree of consistency. Assuming that teacher-assessment is a valid exterior criterion, self-assessment in general can be considered a valid method.

In evaluating the difference between the two methods, the overall difference was insignificant, which means that self- and teacher-based scores were consistent. Overall, self-grading was more lenient compared to teacher-grading, referring to the estimation of mean scores.

### **2. Are peer-assessment results consistent with teacher-assessment results? Are there significant differences between the two assessment results?**

Results of the analysis showed that peer scoring and teacher scoring were neither correlated or significant, indicating inconsistency between the two assessment methods. In this case, peer-assessment failed to produce acceptable validity, if teacher-assessment is a valid exterior criterion. Overall, the difference between the two assessment methods was estimated as insignificant, which suggested an acceptable degree of consistency. With regard to the estimation of mean scores, peer grading was generally more lenient compared to teacher grading.

### **3. Are self-assessment results consistent with peer-assessment results? Are there significant differences between the two assessment results?**

This analysis indicated that the results of self-assessment and peer-assessment were not correlated or significant. As for the mean scores of each

assessment method, peer-grading was slightly higher than self-grading, showing that peer-raters tended to adopt a more lenient grading standard than self-raters did. Rater's bias in evaluation could be an influential factor affecting assessment results and reliability. In this study, grading discrepancies among peer-raters was observed due to raters' subjectivity in scoring.

#### **4. Are the three assessment methods able to evaluate learning achievements (i.e. the consistency between assessment scores and end-of-course examination scores)?**

The assessment scores and end-of-course exam scores were not only correlated but also meaningful. This implied that the three methods were consistent and were therefore able to assess students' learning achievements.

#### **5. Discussion**

##### **(1) Consistency and Difference among Assessment Methods**

The results showed the consistency and differences among the three assessment methods. Overall, consistency was shown only once: between self- and teacher-assessment. Self- and peer-assessment and peer- and teacher-assessment were never found to be consistent.

##### **(2) Comparison of Assessment Results**

Peer-assessment resulted in the highest mean scores (mean=3.68), followed by self-assessment (mean=3.82), and teacher-assessment tended to assign the lowest scores (mean=3.74). These differences were significant ( $p < 0.05$ ). According to the analysis, teacher-grading was the strictest, while peer-grading was the most lenient.

To date, validity and reliability studies related to portfolio assessment have been conducted in both Paper-based and Web-based settings. However, researchers have proposed divergent views on this issue based on their own evidence. Some research has reported sufficient reliability and validity, while other research has shown the opposite results. If the studies showing reliability and validity are to be conclusive, related investigations must cover large enough samples, a variety of course units, and participants from diverse educational backgrounds. Therefore, in self- and peer-assessment, teachers should recognize the issues regarding grading practices in order to avoid factors that might adversely influence assessment results.

#### **Conclusion**

In conclusion, there were significant differences in the evaluation of portfolios using the three assessment methods, among which teacher-raters generally had the strictest grading standards, and peer-raters were the most lenient. The results of teacher- and self-assessment were consistent with end-of-course examination scores, but peer-assessment showed inconsistency with

exam scores. Therefore, when self-grading and peer-grading are considered in determining semester grades, the former should be given greater weighting than the latter.

However, the involvement of peer-assessment uses students' meta-cognitive skills, which is not possible in self-assessment. In this regard, peer-assessment is still regarded as a valuable learning activity, especially in collaboration with self-assessment. The practice of multi-dimensional assessment is beneficial for increasing assessment reliability and validity.

### **ROMANIZED & TRANSLATED REFERENCES FOR ORIGINAL TEXT**

- 王保進 [Wang, Bao-Jinn] (2006)。中文視窗版 SPSS 與行為科學研究 [*Zhongwen shichuangban SPSS yu xingwei kexue yanjiu*]。台北市：心理出版社 [Taipei: Psychological Publishing Co., Ltd.]。
- 余民寧 [Yu, Min-Ning] (2003)。教育測驗與評量：成就測驗與教學評量 [*Jiaoyu ceyan yu pingliang: Chengjiu ceyan yu jiaoxue pingliang*] (二版) [(2nd ed.)]。台北市：心理出版社 [Taipei: Psychological Publishing Co., Ltd.]。
- 卓宜青 [Cho, Yi-Ching] (2001)。網路化學習歷程檔案系統及同儕評量 [*Networked portfolio system with peer assessment*]。未出版之碩士論文 [Unpublished master's thesis]，國立交通大學資訊科學系 [Department of Computer and Information Science, National Chiao Tung University]，新竹市 [Hsinchu]。
- 吳明隆 [Wu, Ming-Lung]、涂金堂 [Tu, Chin-Tang] (2008)。SPSS 與統計應用分析 [*SPSS yu tongji yingyong fenxi*]。台北市：五南 [Taipei: Wu-Nan]。
- 邱皓政 [Chiou, Haw-Jeng] (2008)。量化研究與統計分析：SPSS 中文視窗版資料分析範例解說 [*Lianghua yanjiu yu tongji fenxi: SPSS zhongwen shichuangban ziliao fenxi fanli jieshuo*]。台北市：五南 [Taipei: Wu-Nan]。
- 張振成 [Chang, Chen-Cheng] (1997)。教學評量的新趨勢：實作評量與檔案評量 [*Jiaoxue pingliang de xinqushi: Shizuo pingliang yu dangan pingliang*]。中等教育 [*Secondary Education*]，48(6)，90-94。
- 曾吉雄 [Tzeng, Jir-Shung] (1998)。新式評量：檔案評量 [*Xinshi pingliang: Dangan pingliang*]。教育資料文摘 [*Education Information Digest*]，48(1)，133-137。
- 楊國鑫 [Yang, Kuo-Hsin] (2002)。推廣網路同儕互評系統於高級工業職業學校教學課程之研究 [*Popularizing networked peer assessment system in courses of vocational industrial high school*]。未出版之碩士論文 [Unpublished master's thesis]，國立交通大學資訊科學系 [Department of Computer and Information Science, National Chiao Tung University]，新竹市 [Hsinchu]。
- 鄒慧英 [Tsu, Hui-Ying] (2000)。檔案於國語科教學評量之應用 [*Dangan yu guoyuke jiaoxue pingliang zhi yingyong*]。上網日期：2009年8月7日 [Retrieved August 7, 2009]，檢自 [from]：<http://www.nknu.edu.tw/~edu/item/item4-article.file/item4-article19.htm>
- 劉旨峰 [Liu, Zhi-Feng]、林珊如 [Chang, Shan-Ju L.]、袁賢銘 [Yuan, Shyan-Ming] (2002)。網路化學習歷程檔案之學習成效分析及未來展望 [*Wangluhua xuexi lichen dangan zhi*

- xuexi chengxiao fenxi ji weilai zhanwang]。在大葉大學主辦[In Da-Yeh University]，第三屆電子化企業經營管理理論暨實務研討會[Disanjie dianzihua qiye jingying guanli lilun ji shiwu yantaohui]，彰化縣[Changhua]。
- 劉勝鈺[Liu, Sheng-Yu] (2003)。使用資訊科技學習數學：以網路同儕互評為例[*Learning mathematics with computer technology : Using networked peer assessment in junior high school*]。未出版之碩士論文[Unpublished master's thesis]，國立交通大學網路學習學程碩士班[Degree Program of CS College, National Chiao Tung University]，新竹市[Hsinchu]。
- Anderson, L., & Krathwohl, D. (2001). *A taxonomy for learning teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Wesley Longman.
- 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.
- Barrett, H. (2006). Using electronic portfolios for classroom assessment. *Connected Newsletter*, 13(2), 4-6.
- 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.
- Burch, C. B. (1999). Inside the portfolio experience: The student's perspective. *English Education*, 32(1), 34-49.
- Chang, C. C. (2008). Enhancing self-perceived effects using web-based portfolio assessment. *Computers in Human Behavior*, 24(3), 1753-1771.
- 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), 25-37.
- Chang, C. C., & Tseng, K. H. (2009b). Use and performances of web-based portfolio assessment. *British Journal of Educational Technology*, 40(2), 358-370.
- Chang, C. C., Tseng, K. H., Yueh, H. P., & Lin, W. C. (2011). Consideration factors and adoption of type, tabulation and framework for creating e-portfolios. *Computers & Education*, 56(2), 452-465.
- Chen, C. H. (2010). The implementation and evaluation of a mobile self- and peer-assessment system. *Computers & Education*, 55(1), 229-236.
- Cho, K., Schunn, C., & Wilson, R. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98 (4), 891-901.
- Davies, P. (2000). Computerized peer assessment. *Innovations in Education and Training International*, 37(4), 346-355.
- Derham, C., & Diperna, J. (2007). Digital professional portfolios of preservice teaching: An initial study of score reliability and validity. *International Journal of Technology and Teacher Education*, 15(3), 363-381.
- 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.

- Gadbury-Amyot, C., Kim, J., Palm R. L., Mills, E., Noble, E., & Overman, P. (2003). Validity and reliability of portfolio assessment of competency in a baccalaureate dental hygiene program. *Journal of Dental Education*, 67(9), 991-1002.
- Gadbury-Amyot, C. (2003). *Validity and reliability of portfolio assessment of competency in a baccalaureate dental hygiene program*. Unpublished doctoral dissertation, University of Missouri-Kansas City.
- 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, Ohio.
- High Plains Regional Technology in Education Consortium (2005). *Educational technology MA online program rubric: EPortfolio*. Retrieved February 22, 2010, from [http://www.msarreola.com/Master%20Canvas/Assessment%20&%20Evaluation\\_files/eport\\_rubric.pdf](http://www.msarreola.com/Master%20Canvas/Assessment%20&%20Evaluation_files/eport_rubric.pdf)
- Jenkins, M. (2004). Unfulfilled promise: Formative assessment using computer-aided assessment. *Learning and Teaching in Higher Education*, 1, 67-80.
- Johnson, R. L., McDaniel, F., & Willeke, M. J. (2000). Using portfolios in program evaluation: An investigation of interrater reliability. *American Journal of Evaluation*, 21(1), 65-81.
- Knowles, M. S., Holton III, E. F., & Swanson, R. A. (2005). *The adult learner* (6th ed.). Burlington, MA: Elsevier.
- Lin, S. J., Liu, Z. F., & Yuan, S. M. (2001a). Web-based peer assessment: Attitude and achievement. *IEEE Transactions on Education*, 44(2), 211-223.
- Lin, S. J., Liu, Z. F., & Yuan, S. M. (2001b). Web-based peer assessment: Feedback for students with various thinking-styles. *Journal of Computer Assisted Learning*, 17(4), 420-432.
- Liu, Z. F. (2002). *Studies of networked peer assessment*. Unpublished doctoral dissertation. National Chiao Tung University, Hsinchu.
- Liu, Z. F., Lin, 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.
- Milford School District, (2009). *Information, communication, technology (ICT) portfolio assessment*. Retrieved February 22, 2010, from <http://www.milfordschools.net/mms/documents/ePortfolio%20Rubric%202009.pdf>
- Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment & Evaluation in Higher Education*, 28(4), 383-394.
- Oskay, O., Schallies, M., & Morgil, I. (2008). A closer look at findings from recent publication. *H. U. Journal of Education*, 35, 263-272.
- 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, UK: Routledge.
- 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.
- Reckase, M. A. (2002). *Portfolio define*. Paper presented at the Workshop of Portfolio

- Assessment, Taipei, Taiwan.
- Russell, J. D., & Butcher, C. (1999). Using portfolios in educational technology courses. *Journal of Technology and Teacher Education*, 7(4), 279-289.
- Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1-31.
- Sulzen, J., Young, M., & Hannifin, R. (2008). Reliability and validity of an ecologically-grounded student teacher electronic portfolio rubric. In K. McFerrin et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2008* (pp. 153-159). Chesapeake, VA: AACE.
- 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, 187-202.
- 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(4), 331-334.
- Sweat-Guy, R., & Buzzetto-More, N. (2007). A comparative analysis of common e-portfolio platforms and available features. *Issues in Informing Science and Information Technology Education*, 5(1), 327-342.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.
- 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.
- 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, 293-310.
- Tsai, C. C., Lin, S. J., & Yuan, S. M. (2002). Developing science activities through a networked peer assessment system. *Computers & Education*, 38, 241-252.
- Tseng, S. C., & Tsai, C. C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education*, 49(4), 1161-1174.
- Wen, L. M. C., & Tsai, C. C. (2006). University students' perceptions of and attitudes toward (Online) peer assessment. *Higher Education*, 51, 27-44.
- Yang, Y. F., & Tsai, C. C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction*, 20, 72-83.