

網路化學習歷程檔案同儕互評之信效度研究

張基成* 吳炳宏**

本研究探討網路化檔案同儕互評的信、效度。以某高職二年級修習「計算機應用」課程的七十二位學生為研究樣本。學習者藉由網路化檔案評量系統進行個人檔案的製作、觀摩與同儕互評。結果顯示：(1) 不同學生彼此間評分結果(對同一份檔案)不具一致性(評分者間信度)。(2) 三分之二學生的評分結果(每位學生對不同份檔案)不具一致性(評分者內部信度)。(3) 同儕互評與教學者評分量結果不具一致性(外在效標關聯效度)。(4) 同儕互評分數與測驗成績一致性低, 顯示網路化檔案同儕互評不易反應學習成就(外在效標關聯效度)。簡言之, 網路化檔案同儕互評不具信、效度。

關鍵字：學習歷程檔案、網路化學習歷程檔案、網路化檔案評量、同儕互評

* 作者現職：國立台灣師範大學科技應用與人力資源發展學系教授

** 作者現職：台北市立松山家商資料處理科教師

通訊作者：張基成，e-mail: samchang@ntnu.edu.tw

壹、研究背景

Zimmerman, Bonner 與 Kovach(1996)表示學習歷程檔案(learning portfolio)已受到教育專家推崇為一個值得使用的評量工具(引自林心如譯,2000)。Burch(1999)強調學習歷程檔案記錄學生的學習過程及成果,符合多元評量的精神。面對如此具有特色的評量工具,Lankes(1998)表示將會有越來越多學校以歷程檔案作為評選學習者的工具。而這一項預測同時也從美國 Kalamazoo College 自 1998 年起便將學習歷程檔案列為學生畢業條件之一(Chang, 2001)的做法上得到印證。然而紙本式學習檔案及評量有一些不方便處,加上網路的興起與方便性,網路化學習檔案及評量已越來越普遍。網路不但有助於紀錄、編輯、搜尋和分析學習者的學習歷程檔案資料外,也方便學習者之間分享彼此的檔案資料。另外,Rice(2001)也肯定網路能讓教學活動更具動態性與彈性,表示若將學習歷程檔案與網路兩者互相結合會是個很好的方式。這的確是傳統紙本式學習歷程檔案所不能望其項背的地方!

同儕評量、自我評量、與教師評量是三個重要的學生學習紀錄的評量方式(林珊如、卓宜青、劉旨峰與袁賢銘,2000;Lin, Liu, & Yuan, 2001a; 2001b),但大多數的研究較著重探討同儕互評議題。以網路系統作為同儕互評時的媒介與控制中心,能化解地域阻隔或者無法找到共同時間等的溝通障礙(林珊如等,2000;張基成,2001;Chang, 2001);教師、家長與同儕能夠即時給予回饋,以檢視學生學習歷程。另外,藉由同儕的肯定更可增強學習興趣與信心(李同立,2002)。同儕互評可以提升學生動機,且同儕意見對學生影響高過教師意見(Jenkins, 2004)。評量者評分時,會將他人表現與自己做比較,亦可從被評者作品獲得想法,進而改善自己的作品。而被評者亦可從不同同儕評量結果獲得比來自於教師更多的靈感,因為同儕比教師更瞭解自己(Chen, 2010)。Oskay、Schallies 與 Morgil(2008)綜合幾個檔案評量信、效度的研究結果,認為檔案評量是一個恰當且可靠的評量方法,但可惜的是並未說明是否也包括同儕互評。盧貞穎(2001)指出學生互評的能力不足及其信、效度問題是檔案評量過程中同儕評量常見的問題。但相反地,劉旨峰、林珊如與袁賢銘(2000)表示在網路同儕互評活動中,同儕評分的效度高於自我評量的效度,並進一步建議網路評量可採同儕互評方式(Liu, Lin, Chiu & Yuan, 2001)。一些研究顯示同儕互評與教師評量結果一致(Bouzidi & Jaillet, 2009; Cho, Schunn, & Wilson, 2006; Liu, 2002; Sadler & Good, 2006; Sung et al., 2005; Topping, 1998; Tsai & Liang, 2009; Tseng & Tsai, 2007),但有研究顯示同儕互評與教師評量結果並不一致(Chen, 2010)。上述文獻顯示,不同研究因研究樣本的教育層級、樣本大小、評量準規、評量環境等的不同,而在同儕互評信、效度的結果上並不一致。

上述研究大多是針對紙本式的檔案評量或同儕互評，但網路化評量的信、效度較一般紙本式評量的信、效度更難檢測，也更難（或更容易）達到，因為有更多因素會影響（提高或減低）其信、效度的檢測與達成，譬如科技的支援、科技帶給人們的焦慮、人在科技環境下的心智等（Bouzidi & Jaillet, 2009）。在網路化檔案評量環境下，同儕互評的信、效度是否也有相同的結果？還是不同？因此本研究之目的為探討網路化檔案之同儕互評的信、效度。

貳、文獻探討

一、網路化檔案評量

Chang(2001)、Chang 與 Tseng(2009a)表示學習歷程檔案(learning portfolio)能掌握學習過程的複雜性，幫助學習者或教學者反省、檢討整個學習與教學的過程，並給予回饋與改進，對學習提供真正的助益。透過學習歷程案評量的方式，不但可反應學生的學習成果，更可透過學生的自省紀錄來檢視其真實的學習過程。這些紀錄除了讓學生發覺自己成長軌跡、獲得學習成就感外，更可藉由同儕的檢視與回饋讓學生可以修正自己的觀點（卓宜青、劉旨峰、林珊如與袁賢銘，2001）。

網路化學習歷程檔案(web-based learning portfolio)的特性之一是易於瀏覽、展示與彼此分享資料，增加觀摩學習機會，這使得同儕線上互評與回饋變得更為方便。透過網路化學習檔案內容資料的分享，不但能幫助教師了解學生的學習狀況且有助於增進師生間互動；同時對學生而言，提供了解其他人學習成果及過程機會，增加同儕互動與觀摩學習，促進反省與思考（Chang & Tseng, 2009b；Chang, 2008）。網頁超連結的功能可在社交活動上營造出更大的互動性空間，例如網路上同儕互評活動多少增添了傳統學習歷程檔案所欠缺的互動性（Yancey, 2001）。

檔案評量(portfolio assessment)是學生經由教師指導，有系統地蒐集紀錄其學習的歷程、心得、作品及個人的反省紀錄等證據，將它編輯成個人的學習檔案，然後再依據此檔案來進行評量；不但可檢視學生的學習成果，更可透過學生的自省紀錄來了解其真實的學習過程（Chang, 2001）。所謂的網路化檔案評量指的是透過網路來輔助檔案評量的進行，評量者可於線上進行學習者自評、同儕觀摩與互評、教學者評或其它評量活動。Chang 與 Tseng(2009b)的研究就證實網路化檔案評量的使用對學生的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，學習者線上互評與觀摩扮演重要角色。

二、檔案評量準規

Burch (1999) 的檔案評量準規包括檔案呈現、內容組織、反省思考(深入思考且察覺自己的表現情況)、內容紀錄、內文鋪陳技巧。Reckase (2002) 在檢閱過美國幾個州內對於課程的相關文件規定之後, 由教學者與學習者共同討論選出一份檔案所需包含的內容項目至少五個, 可以做為檔案評量的主要依據, 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。上述兩個檔案評量準規只有反省思考是學習歷程檔案的重要內容項目, 其餘皆屬於檔案製作上的表現。馬里蘭大學 Eastern Shore 校區商業、管理與會計系的評量準規包括學習目標、反思、檔案品質。(Sweat-Guy & Buzzetto-More, 2007)。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、反思、作品 (artifact) 四個面向 (University of Wisconsin - Stout, 2008)。上述兩個檔案評量準規中的學習目標、反思、作品是學習歷程檔案的重要內容項目, 其餘則是有關於檔案製作的表現。

ePortfolio Portal Team 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)。Worcester (2000) 發展的評估準規, 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。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。上述評估準規只包括學習歷程檔案的一個重要項目內容 -- 反思, 其餘皆屬於檔案製作的表現, 尤其是考慮了多媒體因素。

綜合上述所有準規, 主要包括檔案製作與呈現、學習目標的設定、作品、反思面向, 但較欠缺學習態度的檢視。歐滄和 (2002) 表示, 檔案評量的準規除了包括達成學習目標的程度之外, 尚須考慮與他人合作的程度、展現個人特色的程度。因此, 可再增加學習態度面向; 展現個人特色較為獨立, 可放置於「其它」面向中。另外, 與網路有關的特性也應納入並分散在各面向內, 譬如網路多媒體的應用可加入檔案製作面向內, 線上觀摩及同儕互動可加入學習態度面向內等。

三、同儕互評的信、效度

(一) 同儕互評的信度

同儕互評的信度通常是指不同同儕評分者之間的一致性 (Bouzidi & Jailliet, 2009; Cho, Schunn, & Wilson, 2006; Falchikov & Goldfinch, 2000; Sluijsmans, Dochy, & Moerkerke, 1999)。余民寧 (2003) 指出信度是多次評量結果間的一致性。陳嘉鴻 (2002) 亦指出信度是相同的人在不同的時間, 以相同的評量所

得結果的一致性。張麗麗(2002)在國小學童寫作檔案的信度分析研究中建議，檔案評量的信度除了檢視評分者間一致性外，可進一步再檢視評分者內部一致性。評分者內部一致性是指某位評分者不同時間對同一個檔案評量結果的一致性或同一時間對數個同儕檔案評量結果的一致性。但是，檔案評量項目、計分規則、評分者對項目的熟悉度、評分者背景差異等因素會對評分的一致性有所影響，所以在檔案評量之前應針對這些影響因素做好評量者訓練的工作(張麗麗，2002)。

對於信度的檢驗，通常是以前面所述的外部一致性(即外部信度)及內部一致性(即內部信度)來表示評量的信度。Derham 與 Diperna(2007)、Gadbury-Amyot 等人(2003)、Lenze(2003)的研究是檢視不同檔案評分者之間的一致性。此作法可適用於同儕互評，亦即同儕互評的外部信度。同儕互評外部信度指不同學習者對同一個檔案評量結果的一致性，而同儕互評內部信度指某位學習者對數個同儕檔案評量結果的一致性。這兩種同儕互評信度的檢驗方式也是本研究採用的。

(二) 同儕互評的效度

Topping(1998)在分析過一百餘篇同儕互評的研究報告後指出，整體而言，高等教育各學科實施同儕互評具有一定的信度與效度，可以培養學生的高階批判思考與提升學習動機，對成年學習者而言是很好的評量方式。此外，陳信汎(2002)針對國小高年級學生的研究結果顯示，同儕互評具有不錯的效度。通常若欲檢驗同儕互評的效度，需要授課教師的評分或教學助理的評分作為參照的效標(Bouzidi & Jaillet, 2009; Cho, Schunn, & Wilson, 2006; Falchikov & Goldfinch, 2000; Sluijsmans, Dochy, & Moerkerke, 1999)。劉旨峰、林珊如、袁賢銘(2002)表示在網路同儕互評活動中，同儕評分的效度高於自我評量的效度；並指出自我評量的分數最高、同儕互評次之，而以老師所給的分數最為嚴格；並進一步建議網路評量可採同儕互評方式。

Liu 等人(2001)的研究便以教師給的分數作為同儕互評的外在效標，進行效度檢驗。Chen(2010)以師培生為研究對象，結果顯示線上同儕互評與教師評的分數不具一致性。但 Sadler 與 Good(2006)以四門課的高中生為研究對象，結果顯示同儕互評與教師評量結果有高度一致性。Tseng 與 Tsai(2007)以修習電腦課高中學生為研究對象，發現線上同儕互評與教師評的分數具一致性。因此，他們宣稱高中生同儕互評是有效度的一種評量方式。Cho、Schunn 與 Wilson(2006)等人的研究顯示，寫作課程同儕互評的信、效度與教師評量的信、效度一致，並建議同儕互評是可行的一種評量方式。卓宜青等(2001)、

主題文章

楊國鑫（2002）、劉勝鈺（2003）、Bouzidi 與 aillet（2009）、Tsai 與 Liang（2009）的研究亦有相同結果。上述研究顯示，大家對同儕互評效度的看法不同，但大多趨向於同儕互評與教師評量的分數具一致性。

研究者認為，同儕互評的效度可以運用外在效標的方式予以檢驗。可採用的外在效標的來源可以是教學者與學生測驗分數。Gadbury-Amyot 等人（2003）以學生測驗分數為外在效標，檢驗它與檔案評量分數的相關性。相關性達顯著，表示檔案評量具足夠的效度。張麗麗（2002）表示為求外在效標有更高的效度，教學者數量越多越好，或者接受過檔案評量的評分訓練。研究者認為如果可以的話，再加上對擔任外在效標的諸位教學者評分結果之間一致性的檢驗，如此一來此一外在效標的參考程度會更高。反觀，由於教學者較具備公正評量的能力，因此以教學者做為外在效標再輔以對不同教學者間評量結果一致性作出確認，應足以作為同儕互評的外在效標。

參、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實驗對象，共七十九人。刪除檔案內容不全者七人，剩七十二人之檔案作為統計分析用，其中男生三十四名，女生三十八名。大部份的學生之前製作過紙面式的學習歷程檔案，但未進行過檔案評量。他們未製作過網路學習歷程檔案，也未聽過或使用過網路學習歷程檔案系統。他們之間雖然相互認識，但是進行網路同儕互評時是採匿名方式，並不知道對方是誰。教學實驗共十二週，每週三小時。

學習者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、觀摩與同儕互評。教學內容為動畫製作與網頁製作。兩個單元的作業是數位化檔案內主要的項目。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。本研究從開始至結束，均未告知學生正在進行實驗，以避免產生所謂的霍桑效應（Hawthorne effect）與強亨利效應（John Henry effect），減少對實驗結果的干擾。

二、資料分析

資料分析與統計方法詳如表 1。

表 1 資料分析與統計方法

| 研究目的 | 研究問題 | 分析方式 | 統計方法 |
|---------|----------------------------------|----------------------|---------------------|
| 同儕互評的信度 | 不同學習者彼此間(同儕互評)對同一份檔案評量結果是否一致? | 同儕之間評分一致性 | Kendall 和諧係數 |
| | 每一位學習對不同份檔案評量結果是否一致? | 評分者內部一致性 | 同質性係數分析 |
| 同儕互評的效度 | 同儕互評與教學者(授課教師)評量結果是否一致? 是否有顯著差異? | 外在效標關聯效度 (效標為教師所評分數) | Pearson 積差相關 / t 檢定 |
| | 同儕互評分數是否與測驗成績一致? | 外在效標關聯效度 (效標為測驗成績) | Pearson 積差相關 |

三、量表發展

首先透過文獻歸納出檔案評量準規的雛型，與授課教師討論並修正數次後形成評量準規的量表。之後委請三位網路化檔案評量的學者專家進行修正，依據其看法決定各評量準規是否保留、修正、或刪除，並提供文字上的建議，以協助內容效度的建立。依據專家所給意見修正評量準規後形成預試用的量表，之後再對修課的學生說明及經過討論之後做一些調整，最後以網頁形式置入網路化檔案評量系統中進行預試。評量準規包括六個面向，二十二項準規重點如下：

檔案製作：資料完整性、內容恰當程度、內容豐富程度、組織與呈現。

學習目標：個人學習目標設定、學科教學目標、成長與成就、學習目標的達成程度。

作品：作品正確性、恰當性、豐富性及完整性、難度、創意與創新、對學習內容的理解、作品產生的歷程。

反思：學習目標的反思、作品的反思、學習成就的反思、學習態度的反思、觀摩同儕表現的反思，對他人回饋的反思。

主題文章

態度與互動（簡稱態度）：線上觀摩、欣賞、互評與回饋；網路資源分享；線上溝通討論、知識分享、意見交換、問題解決；線上互動的數量；線上多元互動品質。

其它：其它事蹟或證據。

量尺以李克特五等第為理念，每項準規的滿分分數為五分。但考慮實際評分上的需要，擴充為 1、1.5、2、2.5、3、3.5、4、4.5 與 5 分。點數間以 0.5 為單位可增加評分時的選項，方便評分及減少評量結果的差異性過大情況。每項準規的點數間皆定義不同的績效表現水準，以利同儕評分。

四、量表項目分析

在預試量表的鑑別度（discrimination）部分，研究者採用極端組檢驗法，將每一準規平均數分為高(27%)、低(27%)分二組並進行兩組 t 檢定。每一題的 t 值(即決斷值)皆達顯著水準，顯示每一準規的鑑別力足夠，應予保留。在一致性(consistency)部分，研究者將每一準規平均數與所有準規平均數進行 Pearson 積差相關檢定。每一題的 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，應予保留。

五、量表效度檢驗

預試量表各面向的 Bartlett 球形檢定之近似卡方分配皆達顯著，表示各面向的準規之間有共同因素存在，適合進行因素分析。各面向的取樣適切性量數值（Kaiser-Meyer-Olkin, KMO）值皆大於 0.5（表 2），且每一準規的效度係數（共同性估計值）皆大於 0.1，適合投入因素分析。效度係數表示某準規與其它準規之間共同性，若越高表示表示該準規適合投入因素分析；若太小（小於 0.1），則該準規不適合存在，應刪除（王保進，2006）。利用主成份分析法（Principal Component Analysis, PCA）建構效度，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）。第一次因素分析確定準規特徵值大於 1 的五個因素面向，其中「態度」面向內有一準規的因素負荷量小於 0.3，將此準規刪除。第二次因素分析結果顯示，因素負荷量皆大於 0.3，故保留所有準規。

最後形成五個因素面向（特徵值皆大於 1）。各面向的解釋變異量皆大於 45%，顯示各面向皆具一定之效度。作品面向的解釋變異量最大，態度面向最小。

表 2 預試量表之因素分析

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

註：「其它」面向僅有一題，不做因素分析

六、量表信度檢驗

預試與正式評量各面向的 Cronbach's α 值皆大於 0.7，屬於高信度。顯示各項準規之間具有高度的內部一致性，因此預試後保留各項準規。兩次施測都以作品面向最高，態度與互動面向最低（表 3）。

表 3 兩次施測 Cronbach' s α 值

| 面向 | 預試 | 正式施測 |
|------|------|------|
| 檔案製作 | 0.77 | 0.70 |
| 學習目標 | 0.83 | 0.75 |
| 作品 | 0.86 | 0.90 |
| 反思 | 0.80 | 0.87 |
| 態度 | 0.71 | 0.71 |
| 其它 | 無 | 無 |
| 整體 | 0.93 | 0.93 |

註：「其它」面向僅有單獨一項，無法計算 Cronbach' s α 值

七、實驗程序

實驗分兩階段進行，第一階段為預試，第二階段為正式實施。預試為第一單元「動畫製作」，而正式實施為第二單元「網頁製作」。正式實施時會針對預試所遭遇困難做修正，譬如增加評分者對評量準規的熟悉度與認知上一致性、嚴謹的評分者訓練等，以提升正式施測的信、效度。Chang 與 Tseng (2009a, 2009b) 設計的檔案評量活動完整而豐富，且已經過數次的教學實驗，故為本研究採用。實驗過程如下：

(一) 第一單元 (預試)

1. 第一週

在進行檔案製作之前，教學者在課堂上對學習者說明評量準規及計分方式，並呈現評分範例供學習者參考，以增進學習者對準規的了解。一些研究提到(Bloxham & West, 2004; Tsai et al., 2002; Price & O'Donovan, 2006; Prins et al., 2005)，評量準規的訂定要與學生討論，以提升學生對評量的參與程度。因此評量準規及計分方式經過討論之後，根據學生意見做一些調整，以符合學生需要及並提升其參與動機。另外，由助教教導學習者試用網路化檔案評量系統，使學習者能熟悉系統的操作。

2. 第二至第五週

學習者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。教學者進入該系統查看學習者的學習狀況。

3. 第六週

學習者於課後對他人的檔案做同儕互評。在考量學習者負荷量之後，由教師隨機將學習者分為六人一組，組中的每一位成員對另一組所有成員的檔案作同儕互評，亦即每位學習者評六位同儕。這與 Cho、Schunn 與 Wilson (2006) 的研究“每位學習者評四位同儕”很接近。為避免學習者在非匿名機制下可能產生評分不公現象，評量過程以匿名方式進行。同樣地，教學者(一位授課教師與三位線上助教)亦在此單元結束後數日內，對每人的檔案做評量。

學習者點選系統左邊的「檔案評量區」，會出現隨機產生的同儕編號供挑選。學習者點選某位同儕編號者後，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，會出現另外一個評分表視窗網頁。此時，評分者可點選每一項評量指標旁的分數，及指標下面的文字方塊填寫回饋意見。評分時，評分者可同時瀏覽學習者的檔案內容與線上參與情形，以做為評分的參考。



圖 1 檔案評量區內之學習者編號名單

(二) 第二單元 (正式實施)

1. 第七週

教學者在課堂上對學習者說明第一單元實施需要改進的事項，尤其加強同儕互評的說明。另外，教學者整理同儕互評數據及進行信、效度檢驗，並修正評量準規。最後，教學者針對修正的評量準規對學習者做說明。

2. 第八至第十二週

重覆第二至第六週的活動。

肆、結果與討論

一、不同學習者彼此間對同一份檔案評量結果不一致 (評分者間信度)

同儕互評方式是將實驗對象分為十二組 (每組六人)，進行兩兩跨組相互評分；每人評另一組的六人，因此每人的檔案都會被六位同儕評分，如圖 2 所示。每一組六人對另一組六人的檔案之評分結果，經 Kendall 和諧係數的分析如表

主題文章

4。結果均未達顯著水準，表示六位學習者對同一份檔案的評分結果之間並無一致性，意即同儕互評不具有評分者間信度(inter-rater reliability)。此結果與 Topping (1998) 所提出「同儕互評具有一定信度」的看法不同。造成這個現象的可能原因為不同學習者對評量準規的看法寬鬆不一。上述結果符合陳信汎 (2002) 提出的「同儕互評情況下，同學無法信任同儕之間的評分能力」及楊國鑫 (2002) 擔心「同儕在評分上的專業能力不足」的看法。

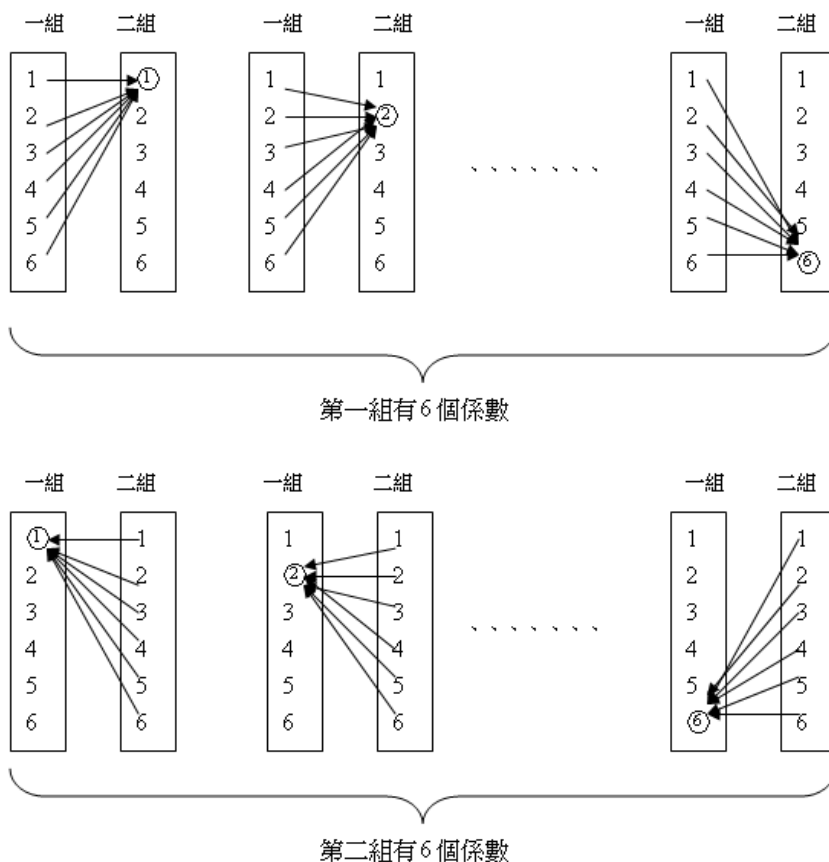


圖 2 不同學習者對同一份檔案的評量 (第一與第二組跨組互評)

表 4 不同學習者對同一份檔案評量結果 Kendall 和諧係數

| 人員 (組別) | 個人和諧係數 (顯著性) | 組整體和諧係數 (顯著性) |
|---------------|--------------|---------------|
| 學習者 NO.11 (一) | 0.26 (0.89) | 0.05 (0.97) |
| 學習者 NO.12 (一) | 0.08 (0.19) | |
| 學習者 NO.13 (一) | 0.37 (0.17) | |
| 學習者 NO.14 (一) | 0.19 (0.73) | |
| 學習者 NO.15 (一) | 0.34 (0.48) | |
| 學習者 NO.16 (一) | 0.07 (0.88) | |
| 學習者 NO.21 (二) | 0.28 (0.44) | 0.17 (0.61) |
| 學習者 NO.22 (二) | 0.13 (0.67) | |
| 學習者 NO.23 (二) | 0.06 (0.35) | |
| 學習者 NO.24 (二) | 0.27 (0.34) | |
| 學習者 NO.25 (二) | 0.36 (0.25) | |
| 學習者 NO.26 (二) | 0.15 (0.13) | |
| 學習者 NO.31 (三) | 0.35 (0.58) | 0.33 (0.37) |
| 學習者 NO.32 (三) | 0.33 (0.38) | |
| 學習者 NO.33 (三) | 0.34 (0.45) | |
| 學習者 NO.34 (三) | 0.18 (0.86) | |
| 學習者 NO.35 (三) | 0.36 (0.64) | |
| 學習者 NO.36 (三) | 0.58 (0.18) | |
| 學習者 NO.41 (四) | 0.71 (0.76) | 0.94 (0.06) |
| 學習者 NO.42 (四) | 0.64 (0.87) | |
| 學習者 NO.43 (四) | 0.86 (0.67) | |
| 學習者 NO.44 (四) | 0.81 (0.55) | |
| 學習者 NO.45 (四) | 0.69 (0.44) | |
| 學習者 NO.46 (四) | 0.73 (0.56) | |
| 學習者 NO.51 (五) | 0.74 (0.78) | 0.67 (0.14) |
| 學習者 NO.52 (五) | 0.58 (0.66) | |
| 學習者 NO.53 (五) | 0.79 (0.76) | |
| 學習者 NO.54 (五) | 0.28 (0.71) | |
| 學習者 NO.55 (五) | 0.78 (0.66) | |
| 學習者 NO.56 (五) | 0.44 (0.73) | |
| 學習者 NO.61 (六) | 0.73 (0.74) | 0.94 (0.06) |
| 學習者 NO.62 (六) | 0.85 (0.59) | |
| 學習者 NO.63 (六) | 0.73 (0.78) | |
| 學習者 NO.64 (六) | 0.65 (0.63) | |
| 學習者 NO.65 (六) | 0.67 (0.43) | |
| 學習者 NO.66 (六) | 0.78 (0.67) | |

註：1.NO.mn 表示第 m 組第 n 個學習者；2.此表僅顯示六組的資料。

二、每一位學習者對不同份檔案評量的結果不完全一致 (評分者內部信度)

評量方式如圖 3，同儕評分的同質性係數分析如表 5。就個人而言，只有三分之一學生的同質性係數達顯著性。就組而言，僅有兩組的成員檔案評分結果的同質性係數達顯著性，表示這四組員對不同份檔案評分結果皆分別具一致性 (評分者內部一致性)。就整體而言，整體同質性係數未達顯著，表示絕大部份評量者對不同份檔案的評量結果並不一致，亦即同儕互評不具有評分者內部信度(inner-rater reliability)。推測可能原因為學習者本身人格成熟度問題，即同儕評分可能不公正，對不同對象的評分有個人的好惡。結果符合 Gelinas (1998) 的研究「檔案評量分數會因為評量者的個人喜好而有所不同」。另一原因可能為檔案評量準規使用上的問題，即準規稍多或被評者證據資料不足而不易評斷；加上評分者負擔增加，致使評分一致性減低。正如同許多研究提到的，同儕互評耗時費力，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)。

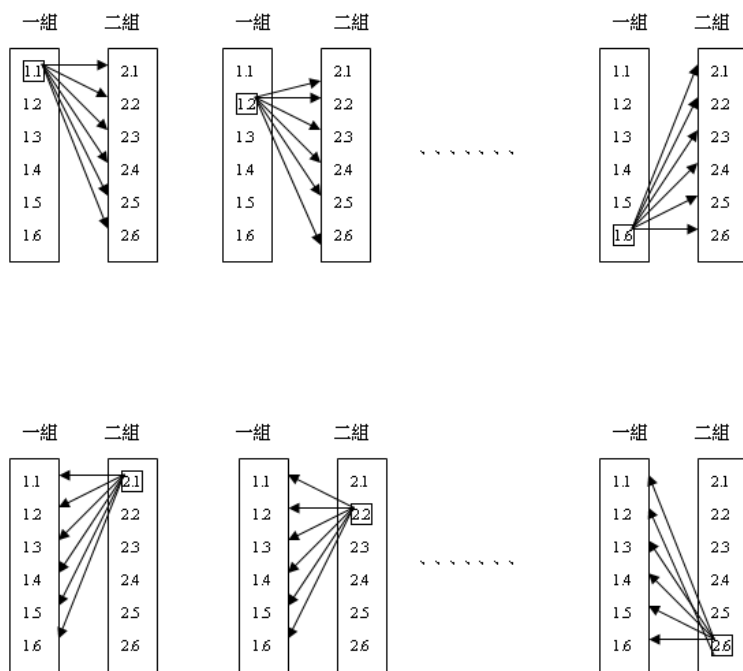


圖 3 每一位學習者對不同份檔案評量(第一與第二組跨組互評)

表 5 每一位學習者對不同份檔案評量結果同質性係數分析

| 人員(組別) | 個人同質性係數 | 個人 Z 值 | 組整體同質性係數 | 組整體 Z 值 | 整體同質性係數 | 整體 Z 值 |
|--------------|---------|--------|----------|---------|---------|--------|
| 學習者 NO.11(一) | 0.96 | 1.96* | | | | |
| 學習者 NO.12(一) | 0.99 | 2.13** | | | | |
| 學習者 NO.13(一) | 0.99 | 2.13** | 0.98 | 2.07** | | |
| 學習者 NO.14(一) | 0.98 | 2.06** | | | | |
| 學習者 NO.15(一) | 0.99 | 2.13** | | | | |
| 學習者 NO.16(一) | 0.97 | 2.02** | | | | |
| 學習者 NO.21(二) | 0.94 | 1.85* | | | | |
| 學習者 NO.22(二) | 0.77 | 1.06 | | | | |
| 學習者 NO.23(二) | 0.80 | 1.18 | 0.80 | 1.21 | | |
| 學習者 NO.24(二) | 0.80 | 1.18 | | | | |
| 學習者 NO.25(二) | 0.73 | 0.90 | | | | |
| 學習者 NO.26(二) | 0.70 | 0.79 | | | 0.79 | 1.16 |
| 學習者 NO.31(三) | 0.93 | 1.80* | | | | |
| 學習者 NO.32(三) | 0.92 | 1.75* | | | | |
| 學習者 NO.33(三) | 0.86 | 1.45 | 0.92 | 1.69* | | |
| 學習者 NO.34(三) | 0.95 | 1.91* | | | | |
| 學習者 NO.35(三) | 0.96 | 1.96* | | | | |
| 學習者 NO.36(三) | 0.88 | 1.55 | | | | |
| 學習者 NO.41(四) | 0.57 | 0.38 | | | | |
| 學習者 NO.42(四) | 0.85 | 1.40 | | | | |
| 學習者 NO.43(四) | 0.61 | 0.49 | 0.58 | 0.38 | | |
| 學習者 NO.44(四) | 0.43 | 0.08 | | | | |
| 學習者 NO.45(四) | 0.65 | 0.62 | | | | |
| 學習者 NO.46(四) | 0.39 | 0.02 | | | | |
| 學習者 NO.51(五) | 0.76 | 1.01 | | | | |
| 學習者 NO.52(五) | 0.98 | 2.07** | 0.82 | 1.25 | | |
| 學習者 NO.53(五) | 0.68 | 0.72 | | | | |
| 學習者 NO.54(五) | 0.74 | 0.94 | | | | |

主題文章

| | | | | |
|--------------|------|-------|------|------|
| 學習者 NO.55(五) | 0.82 | 1.27 | | |
| 學習者 NO.56(五) | 0.94 | 1.85* | | |
| 學習者 NO.61(六) | 0.75 | 0.98 | | |
| 學習者 NO.62(六) | 0.81 | 1.22 | | |
| 學習者 NO.63(六) | 0.51 | 0.24 | 0.67 | 0.63 |
| 學習者 NO.64(六) | 0.75 | 0.98 | | |
| 學習者 NO.65(六) | 0.69 | 0.75 | | |
| 學習者 NO.66(六) | 0.49 | 0.19 | | |

註：1.* $p < .05$, ** $p < .01$ ；2.NO.mn 表示第 m 組第 n 個學習者；3.每位學習者評另一組的六位學習者；4.此表僅顯示六組的資料。

三、同儕互評與教學者評分結果之間的一致性非常低且有顯著差異

學習者同儕相互評分與教學者評分結果的 Person 積差相關及 t 考驗如表 6。結果顯示在各面向及整體上，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。在假設教學者評量是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，此結果與 Chen (2010) 的研究一致，但與大部份的研究結果不同(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)。這可能是同儕互評對象之教育層級、評量方式、評量環境等不同所造成。

兩種評量方法的結果在檔案製作、作品、整體面向上皆未達顯著差異，顯示兩種評量方法的結果在這三個面向上的落差並不大。但在學習目標、反思、態度三面向上，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面向上落差很大。以差異大小(效果量)來看，以學習目標差異最大(效果量)，其次為態度面向，作品面向差異最小。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。以各面向平均分數而言，除了檔案製作、反思之外，其餘三個面向皆為同儕互評評分高於教學者評分。顯示在學習目標、作品、態度面向上，同儕互評較為寬鬆。

表 6 同儕互評與教學者評分結果之 Person 積差相關及 t 考驗

| 面向 | 相關係數 (顯著性) | 評分方式 | 平均數 | 標準差 | t 值 (顯著性) | 效果量 |
|------|-----------------|-------|------|------|--------------------|-------|
| 檔案製作 | 0.28 (0.097) | 互評 | 3.98 | 0.57 | -0.725 (0.474) | 0.142 |
| | | 教學者評分 | 4.05 | 0.40 | | |
| 學習目標 | 0.29 (0.082) | 互評 | 3.92 | 0.25 | 3.636 (0.001**) | 0.939 |
| | | 教學者評分 | 3.69 | 0.24 | | |
| 作品 | 0.18 (0.273) | 互評 | 3.77 | 0.35 | 0.244 (0.809) | 0.065 |
| | | 教學者評分 | 3.75 | 0.26 | | |
| 反思 | 0.03 (0.854) | 互評 | 3.92 | 0.31 | -2.445 (0.020*) | 0.603 |
| | | 教學者評分 | 4.11 | 0.32 | | |
| 態度 | 0.22 (0.190) | 互評 | 3.72 | 0.38 | 2.684 (0.011*) | 0.718 |
| | | 教學者評分 | 3.44 | 0.40 | | |
| 其它 | 0.21 (0.339) | 互評 | 3.51 | 0.34 | 1.524 (0.405) | 0.209 |
| | | 教學者評分 | 3.59 | 0.42 | | |
| 整體 | 0.28 (0.097) | 互評 | 3.86 | 0.29 | 1.724 (0.094) | 0.436 |
| | | 教學者評分 | 3.74 | 0.26 | | |

註：1.*p<.05, **p<.01 2.「其它」面向僅有單獨一項，無法計算 Pearson 積差相關係數 3.學習者互評之平均後再與教學者評分之平均做 Pearson 積差相關的檢定

四、同儕互評檔案分數無法完全檢測出學習成就（即檔案分數是否與測驗成績不完全一致）

同儕互評分數與學生測驗成績的 Pearson 積差相關如表 7。測驗內容主要為學科知識，與檔案評量標準無重覆地方，以提高統計檢驗的可信度。如果兩者有統計上的顯著相關，表示同儕互評分數可以反應學習成就。結果顯示，同儕互評的分數與測驗成績的一致性低。就面向來看，僅學習目標具一致性。這顯示同儕互評結果儘在學習目標面向上可反應學習成就，但其餘面向皆具無法反應學習成就。整體來看，同儕互評的分數無法反應學習成就。

表 7 檔案分數與測驗成績之 Pearson 積差相關

| 面向 | 相關係數 | 顯著性 |
|------|------|--------|
| 檔案製作 | 0.26 | 0.134 |
| 學習目標 | 0.38 | 0.024* |
| 作品 | 0.32 | 0.059 |
| 反思 | 0.24 | 0.161 |
| 態度 | 0.17 | 0.322 |
| 其它 | 0.18 | 0.290 |
| 整體 | 0.15 | 0.403 |

註：* $p < .05$

五、綜合討論

由表 4 學習者檔案互評結果的 Kendall 和諧係數發現，不同學習者對同一份檔案的評量結果並不一致（評分者間信度）；表 5 每一位學習者對不同份檔案評量結果同質性係數分析發現，有三分之二的學習者對不同份檔案的評量結果不一致（評分者內部信度）。以上兩項結果顯示，高職學生同儕互評的信度不足。表 6 學習者互評與教學者評分結果之 Person 積差相關顯示，學習者同儕互評結果與教學者評分結果的一致性非常低。表 7 檔案分數與測驗成績之 Pearson 積差相關顯示，同儕互評的檔案分數與學生測驗成績結果的一致性很低。上述兩項結果顯示，高職學生同儕互評的效度很低。綜上所述，高職學生同儕互評檔案並不具信度與效度。

至目前為止，已有一些透過紙本式與網路化檔案進行同儕互評信、效度的研究。這些研究的結果不一，有部份是信、效度足夠的，也有部份證明信、效度不足。無論同儕互評信、效度是足夠或不足夠，都各有其支持的理論（Cho, Schunn, & Wilson, 2006）。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。

Chen (2010) 的研究顯示，同儕互評的信度不足，而教師對評量活動的監督是可以提高同儕互評的信度。Sung 等人 (2003) 及 Sung 等人 (2005) 研究發現，同儕互評的次數多一些，則學生評量結果與教師評量結果的差異會漸漸減少。這個發現顯示，學生經過多次評量的歷練，對其評分能力有幫助，使得評量信度也增加。同儕互評應同時考慮評量次數與負擔，若互評次數增多，固然可增加評量練習機會，但也可能增加評量負擔；若互評次數太少，固然減少評量負擔，但可能缺乏評量練習機會，進而影響評量結果。因此，如何兼顧互評次數與負擔，是很重要的議題。

Falchikov 與 Goldfinch (2000) 的研究顯示，在容易瞭解的評量準規下，整體上來判斷，大學生同儕互評與教師評量結果頗為接近；但從各個評量面向來檢視，大學生同儕互評與教師評量結果不一定在每一個面向都接近。這與本研究結果有些許不同，本研究除了在整體上，同儕互評與教師評量結果不一致；在各個評量面向上，也不一致。然而 Bouzidi 與 Jaillet (2009) 的研究提出，在各個評量面向上，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份吻合。但無論如何，要讓網路化檔案同儕互評具有足夠信、信度是可以做到的。

伍、結論與建議

本研究受限於教學者授課班級數的限制，僅以一個班級學生為實驗對象。同樣地，受限於學校課程進度，實驗時間非整個學期且僅實施兩個課程單元。若能持續一學期且更多課程單元，更能增加檔案評量的真實性。

一、結論

(一) 同儕互評有其限制所在

本研究對同儕互評的信度問題探討，除了檢視評量者彼此間的一致性外，也以同質性係數來檢視同儕互評參與者的評分者內部一致性。為了取得同質性係數的資料，必須將參與評分的學習者予以分組，並且以組對組的方式兩兩進行互評。在評分過程中，教學者必須掌握學習者互評狀況。一但在評分上發生漏評或者評分對象弄錯的現象，將會影響到評分者信效度。

網路化檔案評量指標多，為減少同儕互評的負擔，讓每位學習者只評六位同儕而非全部的同儕，致使同儕互評信、信度檢定的樣本人數較少，因此恐影響信、信度的正確。但每一位學習者要評六份同儕的檔案，負擔仍不小，可能影響評分意願與表現。

(二) 網路化檔案同儕互評信、效度皆不足

實驗結果顯示高職學生同儕互評不具評分者間信度、評分者內部信度。在各面向及整體上，學習者同儕互評結果與教學者評分結果、與學習者測驗成績的一致性都非常低。因此在假設教學者評量或學習者測驗成績是具有效度而做為效標的前提下，高職學生同儕互評效度很低。有研究成果顯示同儕互評具有一定水準的效度(卓宜青等, 2001; Bouzidi & Jaillet, 2009; Cho, Schunn, & Wilson, 2006; Liu, 2002; Sadler & Good, 2006; Sung et al., 2005; Topping, 1998; Tsai & Liang, 2009; Tseng & Tsai, 2007)，因此本研究認為仍可嘗試採行同儕互評的方式來進行檔案評量。尤其是年齡層越低的學生，越需要足夠的評量訓練。另外，應設法提高同儕互評的效度，或僅作教師評分參考而不計入最後的學期成績。若欲將同儕互評的分數納入最後的學期成績，所佔比例不可高，以減少其效度較低所造成的負面影響。固然，同儕互評的信、效度十分重要，但誠如 Gibbs (2006) 所言「過度強調同儕互評的信、效度反而沖淡了它有助於同儕學習與互動的意義」。亦提到，教師固然要注意同儕互評的信、效度問題，但不應讓它阻礙同儕互評活動的進行。因此，當同儕互評的信、效度足夠時，納入學期成績實屬恰當；但若其信、效度不足時，只能當做提升學生之間互動的學習活動，而不必納入學期成績。無論網路化同儕互評檔案的信、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。

(三) 檔案評量與實驗的掌控勞力、費心且耗時

由於教學實驗與網路化檔案評量系統結合實施，從系統的預試、正式使用、學生的檔案製作、作品的繳交、檔案評量，每個步驟均環環相扣，缺一不可。一位教學者與三位線上助教在實驗過程中，需經常進入系統檢視學習者線上參與狀況。除此之外，也需仔細核對每一位學習者是否有依規定完成同儕互評。在實驗過程中，最令研究者操心的是學習者是否有準時完成每一項工作任務。一開始曾發生一兩位學習者因未能按時繳交作品，導致無法進入評量階段；及學生為急於完成檔案製作，而導致學習目標、反思等填寫的資料有些許不完全。教師發現上述問題後，為求實驗過程的嚴謹，對作品採取積極催繳措施及請學生再檢查與詳填相關資料後，才得以順利進入評量階段。

二、建議

根據上述結論，提出以下建議：

(一) 加強同儕互評訓練與能力，以提升其信效度

本研究雖在實驗進行過程中介紹評量準規並解說評量要點，之後更在準規預試之後針對評量上常發生的問題進行討論，以提高同儕互評能力。但研究發現，同儕互評的信、效度很低，顯示同儕互評的訓練與心理建設等仍需加強，以提升互評的意願與能力。根據 Gadbury-Amyot (2003)、Oskay, Schallies 與

Morgil (2008)、Sulzen, Young 與 Hannifin (2008)，經過訓練的評分者、對檔案有充分的瞭解可以提升評分的信度。研究者也認為，對參與評量者提供更嚴謹的評分訓練並加強評分者對檔案評量的瞭解會是個提升信、效度可行的方式。

(二) 採用配合套措施，預防評分不公正的現象

同儕互評的結果發現部分學生有評分不公的現象，譬如對不同人、作品的評分有個人的好惡，對自己較熟悉或較要好的給分較高等。本研究進行同儕互評之初有提醒學習者評分時需秉持公正態度，勿流於鄉愿或個人好惡。但研究結果發現約 2/3 的學習者未能善盡其事。同儕間評分不公的現象對同儕互評信度、效度有舉足輕重的影響，同時也是影響教學者能否放心運用同儕評量方式來進行檔案評量的因素。根據 Oskay、Schallies 與 Morgil (2008)，評分者對評量目的與準規有共同的認知，可以提高評量的信度。因此，除了本研究所運用的提醒方式外，或可採較積極的手段，如加強評分者對對評量目的與準規的認知、正面的獎賞或負面的懲罰等，來防止評分不公現象的發生。

(三) 推廣應用同儕互評在不同學科與對象上

Topping (1998) 表示，同儕互評的評量方式在寫作、工程、理科、社會科學、...等不同領域上均有人運用，並且具有一定水準的信度。根據 Oskay、Schallies 與 Morgil (2008)，檔案評量是適合的且具有信度的評量方法，可適用於不同教育階段的學生及不同的學科上。評量信、效度的檢驗最好是經過不同教育層級研究對象、大樣本、多課程的實證研究，所獲結果才完整且可靠 (Bouzidi & Jailet, 2009)。建議後續研究可將本研究所建立的準規運用於不同學科與對象，進行驗證及修改。

(四) 擴大研究樣本並增加同儕互評數量

本研究的參與對象共有七十二位學習者，學習者六人一組進行兩組人員之間的互評（即學習者每人評 6 份同儕的檔案）。Johnson 等人 (2000) 提到，評分者數目增加，評分的信度可提高。小樣本在統計結果的推估上易有誤差情況發生，建議未來研究能增加樣本人數，以提昇研究結果的信、效度。另外，同儕互評的檔案份數亦可酌予增加，但亦須考量同儕的負擔，在兩相平衡下考慮適合的互評數量。通常大樣本的評量信度會較高 (Gadbury-Amyot, 2003)，但同儕互評樣本太多也會影響學生互評的意願。要如何在兩者之間取得平衡，確實是一項挑戰。

(五) 比較不同互評方式之間的信、效度差異

本研究為簡化評量過程，採組間小組成員匿名互評方式來進行檔案評量。建議未來可嘗試採用其它互評方式（譬如隨機評數人、評全部等），並比較不同互評方式的信、效度，以確認何種方式較佳。

參考文獻

- 王保進（2006）。**教中文視窗版 SPSS 與行為科學**。臺北：心理出版社。
- 余民寧（2003）。**教育測驗與評量 - 成就測驗與教學評量**（2版）。臺北：心理出版社。
- 李同立（2002）。**網路化學習歷程檔案在國小主題教學之研究**。國立臺南大學自然科學教育研究所碩士論文，未出版，臺南。
- 卓宜青、劉旨峰、林珊如、袁賢銘（2001）。網路化學習歷程檔案系統之設計及同儕評量。載於中央大學舉辦之「**第五屆全球華人學習科技研討會暨第十屆國際電腦輔助教學研討會**」論文集（p.956-963）。桃園：中央大學。
- 林心茹譯（B. J. Zimmerman, S. Bonner, & R. Kovach 原著，1996）（2000）。**自律學習**。臺北：遠流出版公司。
- 林珊如、卓宜青、劉旨峰、袁賢銘（2000）。網路化學習歷程與同儕互評系統之建構經驗。**資訊與教育雜誌**，80，57-64。
- 張基成（2001）。網路化學習檔案之系統化建構經驗及相關問題探討。**視聽教育**，43(2)，9-22。
- 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。
- 陳信汎（2002）。**運用群集技術支援國小學童網路同儕互評之研究**。國立臺南大學資訊教育研究所碩士論文，未出版，臺南。
- 陳嘉鴻（2002）。**高職資訊性向量表信、效度與結果運用之研究**。國立高雄師範大學資訊教育研究所碩士論文，未出版，高雄。
- 歐滄和（2002）。**教育測驗與評量**。臺北：心理。
- 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**。國立交通大學資訊科學研究所碩士論文，未出版，新竹。
- 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。論文發表於**第三屆電子化企業經營管理理論暨實務研討會**，彰化，大葉大學。
- 劉勝鈺（2003）。**使用資訊科技學習數學：以網路同儕互評為例**。國立交通大學網路學習學程碩士班論文，未出版，新竹。

- 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。
- 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.
- 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.
- Bouzidi, L., & Jaillet, A. (2009). Can online peer assessment be trusted?. *Educational Technology & Society*, 12 (4), 257-268.
- Burch, C. B. (1999). Inside the portfolio experience: The student's perspective. *English Education*, 32(1), 34-49.
- 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.
- 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.
- 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*

主題文章

of Technology and Teacher Education, 15(3), 363-381.

ePortfolio Portal (2009). *A generic rubric for evaluating ePortfolios*. Retrieved February 22, 2011, from <http://www.danwilton.com/eportfolios/rubric.php>

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. 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.

Gibbs, G. (2006). How assessment frames student learning. In C.Bryan & K.Clegg, (Eds.), *Innovative Assessment in Higher Education* (pp. 23-36). London: Routledge.

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.

Lankes, A. M. D. (1998). Portfolios: A new wave in assessment. *T H E Journal*, 25(9), 18-19.

Lenze, J. (2004). Inter-rater reliability in the evaluation of electronic portfolios: A survey of empirical research results. In R. Ferdig et al. (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2004* (pp. 164-169). Chesapeake, VA: AACE.

Lin, S.-J., Liu, Z.-F., & Yuan, S.-M. (2001a). Web-based peer assessment: Attitude

and achievement. *IEEE Transactions on Education*, 44(2), 13-13.

- 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.
- Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment & Evaluation in Higher Education*, 28(4), 383-394.
- Morris, J. L. (2007). *Rubric for Assessing Electronic Portfolios*. <http://www.uvm.edu/~jmorris/>.
- 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.
- Rcampus (2010). *iRubric: ePortfolio rubric - University of New England*. Retrieved February 22, 2011, from <http://www.rcampus.com/rubricshowc.cfm?code=W38X23&sp=true&>
- Reckase, M. A. (2002). *Portfolio define*. Paper presented at the Workshop of Portfolio Assessment, Taipei, Taiwan.
- Rice, R. (2001). Composing the Internet-based electronic portfolio using “common” tools. 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

主題文章

Education Press.

- Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11* (1), 1-31.
- Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer-, and co-assessment. *Learning Environments Research, 1*(3), 293-319.
- 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.
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276.
- 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.

- University of Wisconsin - Stout (2010). *Evaluating your portfolio*. Retrieved February 22, 2011, from <http://www.uwstout.edu/art/artedportfolios/evaluating/portevaluationrubric.html>
- Wen, L. M. C., & Tsai, C.-C. (2006). University students' perceptions of and attitudes toward (Online) peer assessment. *Higher Education, 51*, 27-44.
- Worcester, T. (2000). *Electronic portfolio scoring rubric*. Retrieved February 22, 2011, from <http://www.sv400.k12.ks.us/port/rubric.html>
- Yancey, K. B. (2001). General patterns and the future. In B. L. Cambridge, S. Kahn, D. P. Tompkins, & K. B. Yancey (Eds.). *Electronic Portfolios - Emerging Practice in Student, Faculty, and Institutional Learning* (pp.83-87), Washington, DC: American Association for Higher Education Press.
- Yang, Y.-F., & Tsai, C.-C. (2010). Conceptions of and approaches to learning through online peer assessment. *Learning and Instruction, 20*, 72-83.
- Zimmerman, B. J., & Bonner, S., & Kovach, R. (1996). *Developing self-regulated learners: Beyond achievement to self-efficacy*. Washington, DC: American Psychological Association.

A Study on Reliability and Validity of Web-Based Portfolio Peer-Assessment

Chang, Chi-Cheng * Wu, Bing-Hong **

This study explores the reliability and validity of web-based portfolio peer-assessments. The research samples consist of seventy-two students who take “Computer Application” course of the second grade in a vocational high school. The students use the web-based portfolio assessment system for creating their portfolio, emulating peers’ portfolio, and performing peer-assessments. The results show that the assessment scores between different students (on the same one portfolio) have no adequate consistence (reliability between assessors). The assessment scores of two-third students (each student on different portfolios) have no adequate consistence (reliability within an assessor). The scores of peer-assessments and teacher assessment have no adequate consistence (external object association validity). The consistence between peer-assessments and test scores is low, meaning the web-based portfolio peer-assessments do not demonstrate learning achievement (external object association validity). In short, web-based portfolio peer-assessment have no adequate reliability and validity.

Keywords: Portfolio, Web-Based Portfolio, Web-Based Portfolio Assessment, Peer-Assessment

* Chang, Chi-Cheng, Professor, Department of Technology Application and Human Resource Development, National Taiwan Normal University

** Wu, Bing-Hong, Teacher, Department of Data Processing, Taipei Municipal Sung-Shan Senior High School