

國立臺灣師範大學教育心理與輔導學系
教育心理學報，民 97，39 卷，測驗與評量專刊，43—60 頁

國小學童資訊素養檔案評量之信度研究^{*}

張 郁 雯

國立台北教育大學
國民教育學系

本研究的目的是在探討國小資訊素養檔案的分數信度，透過三個面向檢視檔案分數的信度，分別是整體得分的類推性係數與可靠性係數、運用資訊能力歷程向度評分之變異數成分估計和類推性係數以及內外評分者的一致性信度。參與研究的為國小三年級和五年級各兩個班的學生，檔案設計以大六教學法為參考架構，結合語文、自然與生活科技、社會三個學習領域，每位學生在每個領域需完成 2-3 份作品，研究為期 3 個學期。研究結果顯示不分年級與學習領域，在僅有一位評分員的條件下，整體評分類推性係數多半可達 .75 以上，而可靠性係數僅略低於類推性係數，信度良好。歷程向度評分的變異數成分分析顯示，變異的主要來源有四，分別為學生差異、歷程向度、學生與歷程向度的交互作用，以及學生、評分者、和歷程向度的交互作用。歷程向度的類推性係數最佳值在 0.5 和 0.6 左右，仍有改善的空間。校內教師間的評分相關普遍高於教師與外部研究助理的相關，校內教師的給分多半較外部評分者寬鬆，不過兩者的差異量並不大。本研究顯示透過適當的檔案設計，任課教師能一致地評量資訊素養檔案。

關鍵詞：可靠性係數、信度、資訊素養、檔案評量、類推性係數

緒 論

九年一貫課程的目標在發展國民所需的十大基本能力—了解自我與發展潛能，欣賞、表現與創新，生涯規劃與終身學習，表達、溝通與分享，尊重、關懷與團隊合作，文化學習與國際瞭解，規劃組織與實踐，運用科技與資訊，主動探索與研究，獨立思考與解決問題（教育部，民 92；陳伯璋，民 88）。依據綱要這十項基本能力不僅是各個學習領域課程綱要的編輯基準，同時也是未來設計基本能力評量的重要依據。換言之，學生十項基本能力的發展水準是教學成效檢測的基準。如何結合各學習領域的分段學習指標與十大基本能力以設計適合的能力檢測評量成為現職教師面臨的新挑戰。

研究者以 Eisenberg 和 Berkowitz(1990) 的資訊能力大六教學法 (the Big Six Skills approach) 為參考架構，結合國語、社會及自然與生活科技學習領域的教材，發展出評量國小學童運用資訊能力的檔案評量系統。採用檔案評量形式是基於九年一貫課程設計是將資訊教育融入各學習領域，而美

^{*} 本研究承行政院國家科學委員會專案計畫 (NSC 92-2413-H-152-007) 以及 (NSC 93-2413-152-004) 補助，謹此致謝。

國學校圖書館學會主張資訊素養教育的教學原則之一是融入至學校課程的內容與目標中 (American Association of School Librarians, 1998)。檔案評量形式最適合將不同學習領域中涉及運用資訊能力的教學活動所產生的學生作品加以蒐集，進行評量。

檔案評量雖然由來已久，但從 1980 年代才受到教學評量學者的重視，研究發現檔案評量所能提供的訊息不同於傳統選擇題型，檔案評量具有下述的優點：對教學與學習的正向影響，能描述學生思考歷程，整合評量與教學，捕捉成長歷程，讓學生主動學習，以及提供父母與教育決策者有價值的訊息 (Arter & Spandel, 1992; Bridgeman, Chittenden, & Cline, 1995; Frechtling, 1991; Glazer & Brown, 1993; Herman & Winters, 1994; Myers & Pearson, 1996; Tierney, Carter, & Desai, 1991)。由於運用資訊能力強調完成作品的過程，因此本研究選用檔案評量方能看出學童的學習狀況。雖然在有許多教育文獻探討檔案評量，然而在運用資訊能力方面的研究仍屬有限 (Fourie, 1999)

檔案評量的技術品質經常為人所質疑，其中又以信、效度最為核心 (Herman & Winters, 1994; Linn, Baker, & Dunbar, 1991)。而信度是良好效度的先備條件，檔案評量研究雖然不少，然而有關檔案信度的實徵研究仍屬稀少，而研究的結果並不一致。有些研究得到令人滿意的信度水準 (Herman, Gearhart, & Baker, 1993; Johnson, McDaniel, & Willeke, 2000; LeMahieu, Gitomer, & Eresh, 1995; Supovitz, MacGowan, & Slattery, 1997)，有些則得到低的評分者信度 (張麗麗，民 91; 鄒慧英，民 93; Koretz, Stecher, & Deibert, 1993)。本研究探討國小學童之資訊素養檔案的信度議題。這個研究的重要性為，第一，這是一個整合不同學習領域的檔案評量之信度研究，可探討不同學習領域的信度。第二，目前有關小學生檔案分數的信度研究，無論是國內或國外多半集中於寫作與數學領域 (張麗麗，民 91; 鄒慧英，民 93; Shapley & Bush, 1999)，這個研究則是以基本能力為評量標的。第三，這個檔案設計由現場教師執行，進行評分，並加入一位外部評分者，能夠探討運用於教學現場的信度以及內外評分者之信度。第四，這個檔案評量蒐集國小三到六年級學童的作品，能夠評估信度跨年級的穩定程度。

本研究主要探討的問題如下：國小資訊素養檔案的信度如何？影響信度的主要原因為何？是評分者評分不一致？不同向度表現差異太大？還是因素間的交互作用造成的？不同學習領域的信度是否一致？不同學期檔案的信度都近似嗎？原任課教師與其他教師的評分者 (內部評分者) 一致性如何？任課教師與研究助理 (外部評分者) 的信度又如何？

一、資訊素養檔案

資訊素養這個名詞是由 Zurkouski 在 1974 年由首先提出 (吳美美，民 89)。根據美國圖書館學會 (American Library Association, 1989) 的定義，資訊素養就是個人為了應付瞬息萬變的環境，需要更多的知識基礎以及技術性的能力來發掘這些知識，並且能夠將所得的知識進行整合，靈活地使用它們。國內李德竹 (民 89) 則認為資訊素養在於培養國民具備瞭解資訊的價值，使國民在需要資訊時可以有效地查詢資訊、評估資訊、組織資訊與利用資訊。資訊素養教育的目標在培養個體分辨、蒐集、評鑑及利用資訊之能力 (賴麗珍，民 88)。

資訊素養教育模式由過去著重資源中心的觀點的資源取向與探索者取向逐漸轉為以學生為中心的歷程取向 (Kuhlthau, 1987)。歷程取向的教育模式強調從思考與問題解決的角度教導運用資訊能力。著名的歷程模式有 Irving(1985) 的九步驟模式 (Nine-step model)、Eisenberg 和 Berkowitz (1990) 的大六教學法、Stripling 與 Pitts(1988) 的研究歷程模式、Joyce 和 Tallman (1997) 的 I-Search 模式、以及 Pappas 和 Tepe(1997) 的 Pathways to Knowledge。這些模式均提供解決資訊問題所需經歷的歷程，理論雖多，但基本上這些理論高度相似 (Eisenberg & Berkowitz, 1999)，其中 Irving 和大六模式主張他們的模式適用於解決各種生活決策的問題，是基本的生活能力。因此，此模式所培養的能力不僅可用

於解決資訊問題，也能用於解決日常生活問題，更具應用價值。在教學現場，實用的理論應盡可能符合簡潔的原則，本研究選擇以大六教學法作為評量設計的參考架構。

大六教學法認為學生作業中若需用到多元資訊資源，基本上就可視為一個資訊「問題」，而問題的解決可以透過系統性的六個階段思考歷程而完成，此六個歷程分別為作業定義、搜尋策略、尋找資訊、使用資訊、整合資訊、評鑑（吳美美，民 89）。Eisenberg 和 Berkowitz（1999）認為在大六技能的教學與學習上，與課程結合的學習成效最佳，而且各種電腦與資訊科技能力也都可視為大六技能的部分。例如，學生利用 Word 文書軟體寫信，為整合資訊的歷程，而用網路搜尋資料，則為尋找資訊的階段。

成功的大六教學法必須同時兼顧兩個基本的學習脈絡：（1）學生必須清楚大六歷程。解決資訊問題不必然要全部經歷六個歷程，某個階段可能反覆數次；有時毫不費力就能完成某個階段，但有時卻很困難，需要花費許多時間才能完成。因此，對學生而言，他們必須很清楚自己所在的階段，以便知道之後該如何做。（2）學生必須有真正的資訊需求，無論是基於個人的需要或是課程的要求。在學校，最適合的情境即是課程。Eisenberg 和 Berkowitz（1999）不主張為了教大六技能而增加新的課程，而是在現有的課程找出適當教導與學習大六技能的時機。

基於上述的原則，研究者綜合資訊素養教育文獻回顧；國語、社會及自然與生活科技學習領域的教學目標、教材內容、習作之分析；以及教學現場教師如何培養學童運用資訊能力之訪談結果，定義出三個學習領域涉及蒐集資料的相關作業類型，供教師能迅速發展適合資訊能力運用的相關作業設計。作業設計的架構為概分二大類作業：迷你作業是結構較清楚，涉及六大資訊歷程中某些步驟，而非全部歷程，適合短時間完成，所須資訊較少；深度作業則相對而言是屬於弱結構、涉及六大資訊歷程中五個歷程以上，完成作業需要較多的時間，需要豐富的資訊。依領域，可再細分二到三類。表 1 呈現資訊素養檔案的作業類型。

表 1 資訊素養檔案各學習領域作業類型

	迷你作業	深度作業
國語	課文主題、內容之延伸學習	蒐集並閱讀主題相關的讀物
	認識文學家	寫作練習
	蒐集相關詩作	
社會	故事、傳說、由來、節慶與習俗	過去、現在、未來 -- 事實現象之變遷
	圖表、照片的蒐集與解讀	現象分析與問題解決
		實際參觀與資料蒐集
自然	特定概念的延伸學習	原理與生活經驗之連結
	自然觀察	專題
	實驗觀察	科學史

研究者針對每個領域，每個學期分別設計兩個到三個運用資訊能力的作業，作業數量兩個或三個取決於兩個因素，一個是該學期學習領域內容適合與運用資訊能力結合的單元數以及作業類型是迷你或深度作業，由於學生完成深度作業所需時間較多，因此，該學期某領域若有一個深度作業，通常會再配合一個迷你作業。每個作業分別評量 2 到 6 個資訊歷程向度能力。在四下和六下，國語科教師希望以一個深度的專題報告檢驗學生運用資訊能力的學習成果，因此，該學期的國語學習領域，只有一個作業。表 2 呈現小三下學期各學習領域教學單元與資訊歷程雙向細目表，表中 * 表示該項作業評量到的資訊歷程。為了增進讀者對此檔案內作業的瞭解，附錄一呈現三下社會領域的作業「家鄉的地

名一小小搜查員」之內容、評分表，以及學生在各歷程不同水準的表現範例。有關本檔案系統的發展與設計詳細內容請參考張郁雯（民 93）之論文。

表 2 各學習領域教學單元與資訊歷程雙向細目表

領域	單元名稱	大六資訊歷程					
		作業定義	搜尋策略	尋找資訊	使用資訊	整合資訊	評鑑
國語	心情的河流				*	*	*
	語文麻雀	*	*	*	*		
自然	我的秘密花園	*	*	*	*		
	生活中的力	*			*	*	*
社會	社區居民活動中心			*	*		
	家鄉的地名—小小搜查員	*		*	*		*

方 法

一、參與學生

本研究參與對象為台北市某所小學三年級及五年級各兩個班級，每班約有 35 名學生，進行為期一年半的研究。該所學校是屬於額滿學校，研究初期透過教師訪談獲知，該校每學期因課程的需要，每個領域教師大約要求學生進行一到二次資料蒐集之相關作業，因此，課外蒐集資料的作業對該校學生應不陌生。教師通常較重視資料蒐集結果的內容，較少進行運用資訊歷程的教學，教師認為多數學生此類作業的品質並不理想。建立資訊素養檔案對參與研究的師生，均屬初次經驗。

該校的選取是方便取樣，因為在小學現場進行研究，必須獲得教師合作之首肯。該校教師過去參與過一些研究案，故能協助本研究之進行。參與本次研究之教師半數會讓學生建立學習檔案，但並未對檔案進行評量。

三年級的檔案是結合國語、自然與生活科技和社會三個領域。其中自然與生活科技領域由科任老師擔任，其餘兩個領域由導師擔任。高年級的自然與生活科技與社會領域都是由科任老師擔任，兩個班的國語科都由同一位教師擔任。高年級的社會科任教師因故無法參與研究，因此高年級的檔案只涵蓋兩個領域。校內教師教學年資平均為 14.8 年。由於部分學生缺交作業，或者作業中某些項目缺漏，因此，每個年級每學期實際可供分析的完整資料份數因領域不同而不同，三下時，語文、自然、社會領域的有效資料份數分別為 48、57、50 份；四上時，有效份數依序為 64、57、65 份；四下時，則為 66、59、65 份；五上的語文與自然領域有效資料份數為 41 份和 53 份，六上則為 52 份與 57 份；六下則為 58 份與 66 份。

二、評分過程

檔案中的每份作品都經過班級任課教師、同領域的另一位教師，以及一位研究助理的評分。因此，校內教師共有六位參與評分工作，研究助理共有三位參與評分，每一位助理分別負責一個學習領域，這三位助理皆為師範教育學院四年級的在學生，經研究者施以評分訓練。教師在研究開始與結束均接受蒐集資料教學的晤談。然後由研究者解說大六教學法的六個重要運用資訊的歷程以及資訊檔案作業的設計原則，接著請教師從該學期任課的領域，挑出適合整合運用資訊能力的單元，由研究者初步規劃作業，在每學期初建構各學習領域教學單元與資訊歷程雙向細目表如表 2，最後由研究者與教

師設計作業與評分表，在學期中與學期末進行檔案實施的分享、檢討與修正。

評分表是根據運用資訊各個歷程的表現標準為基礎，考慮作業之特性，以及學生所在年級，設計出每項作業特定的評分表。表 3 呈現自然與生活科技領域的一個作業—用根莖葉繁殖的評分表，附錄一也有作業評分表可供參考。從這兩個評分表可發現評分量尺有時是兩點量尺，有時為三點量尺，在其他的作業還用到四點量尺。之所以採取變動點數的評分表是在評分試作的過程中，教師反映劃一的評分表使用不易，因為有些技能只須檢核學生是否做到，有些技能之表現則有等級的差異。採用劃一的評分表無法因應此種作業設計的特性。

進行作業評分前，再次與教師回顧六個重要運用資訊的歷程，並說明評分表與六項歷程能力的關係。評分初期先進行試評，然後對評分的一致性進行檢視，如果評分出現不一致，則進一步瞭解評分不一致的原因，若是由於評分者對評分標準的解釋不一，則修正或澄清評分表定義與表現水準。

評分時，檔案內個別作品逐一評分，由於評分的標準是以運用資訊的大六歷程為架構，因此，將個別作品的在各向度的得分加總，並將同一學習領域內的作品得分相加，可以得到該領域該學期的整體評分分數。另外，以運用資訊歷程向度為軸，將領域內不同作品在同一向度中的得分加總，可以得到各歷程向度之得分。

表 3 評分表範例

運用資訊歷程	3 分	2 分	1 分
一、定義問題			
能透過資料蒐集的結果，選擇想要種植的植物並說明原因。	<input type="checkbox"/>	<input type="checkbox"/> 是	<input type="checkbox"/> 否
二、搜尋策略			
蒐集資料的方式多元。	<input type="checkbox"/>	<input type="checkbox"/> 資料蒐集方式多元。	<input type="checkbox"/> 單一蒐集方式。
四、使用資訊			
能夠將植物生長的情形記錄在表中。	<input type="checkbox"/> 完成記錄且內容清楚完整。	<input type="checkbox"/> 完成紀錄，但記錄內容不夠清楚完整。	<input type="checkbox"/> 有部分項目未完成記錄。
五、整合			
能夠依據植物生長紀錄歸納出實驗的結果。	<input type="checkbox"/> 回答內容清楚扼要，能切合重點。	<input type="checkbox"/> 回答內容有提到重點但過於冗長	<input type="checkbox"/> 回答內容未能切合重點。
六、評鑑			
能夠欣賞其他人的觀察紀錄之優點。	<input type="checkbox"/>	<input type="checkbox"/> 能具體提出優點。	<input type="checkbox"/> 未能具體提出優點。
能夠自我評量觀察紀錄結果。	<input type="checkbox"/> 均能具體說明優點與待加強之處。	<input type="checkbox"/> 只具體說明優點或待加強之處其中一項。	<input type="checkbox"/> 均未能具體說明優點與待加強之處。

三、分析單位

本研究的資料分析係以檔案內的各學習領域為單位。由於檔案作品必須在教學現場蒐集，因此，研究設計特別顧慮現實的可行性。評分者只負責特定學習領域作品的評分，信度分析時，無法切割評

分者與學習領域所造成的分數變異，因而，本研究採取各領域分開分析。這樣的設計方式雖然不是最理想，但符合小學現場的狀況，因為目前在學期成績單的呈現上，各領域是分別給分的。

檔案不僅在評量學童運用資訊的整體能力還希望評估學生在解決資訊問題時，在各個歷程的表現情形，所以，本研究同時探討檔案整體評分以各大六技能歷程評分的信度。

根據類化理論（generalizability theory, Shavelson & Webb, 1991），總分的評分，變異的來源有學生能力差異（p）、評分者（r）和學生差異與評分者的交互作用（pr）。學生能力是測量對象，當著重相對分數解釋時，類推性係數（generalizability coefficient）計算如下：

$$\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{pr}^2}{n_r}}$$

其中 σ_p^2 為學生能力變異， σ_{pr}^2 為學生與評分者的交互作用， n_r 為決策研究打算使用的評分者人數。當決策是依賴絕對分數，則需估計的是可靠性係數（dependability coefficient），此時，除了學生差異與評分者的交互作用，評分者的變異也視為誤差來源，可靠性係數的估計公式為：

$$\Phi = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_r^2}{n_r} + \frac{\sigma_{pr}^2}{n_r}}$$

其中 σ_p^2 ， σ_{pr}^2 ， n_r 與上式的定義相同， σ_r^2 為評分者的變異。

歷程向度評分變異的來源有學生能力差異（p）、評分者（r）、向度（d）、學生差異與評分者的交互作用（pr）、學生差異和向度的交互作用（pd）、向度與評分者的交互作用（rd）以及三個因子的交互作用。類推性係數的計算，在分子的部分仍是 σ_p^2 ，分母的部分除了 σ_p^2 ， σ_{pr}^2 ，另加上 σ_{pd}^2 及 σ_{prd}^2 。歷程向度的類推性係數之計算公式如下：

$$\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \frac{\sigma_{pr}^2}{n_r} + \frac{\sigma_{pd}^2}{n_d} + \frac{\sigma_{prd}^2}{n_r n_d}}$$

n_r 為決策研究打算使用的評分者人數， n_d 為決策研究打算評量的歷程向度數。

結果與討論

一、整體評分之信度

整體評分時，各領域在運用資訊能力上得到一個總分，信度分析探討學生能力差異、評分者評分一致程度對總分變異的影響。表 4 的數值顯示不同學習領域、不同年級，在做相對與絕對決策時，不

同評分者人數能夠達到的評分信度。當只用一位評分員時，這個檔案系統在每個學期各領域的作業數只有 2 到 3 個的情形下，類推性係數多能達到 .75 以上，信度品質算是不錯。在教學現場通常只能有一位教師進行評分，在這個條件限制下，本檔案系統仍有相當不錯的信度品質。當評分者增為兩人時，不論年級、領域，類推性係數都能達到 .80 以上。

語文領域的類推性係數在四下和六上不盡理想，其可能原因為四下的作業設計是一個涵蓋五個歷程向度的專題報告，教師擔心學生負荷過重，因而採取小組合作方式，評分時，小組成員除了在評鑑歷程向度外，其餘向度的得分均相同，可能因而使得學生差異無法顯現出來。學生能力變異小，類推性係數自然不理想，表 5 歷程向度評分之變異數成分估計結果支持這項推測。六上語文領域類推性係數不佳的可能原因是該學期設計兩項作業，其中一項作業過於簡單，學生幾乎全會做，導致類推性係數不佳。未來研究可進一步探討，作為作業設計注意事項。

類推性係數之大小並未受到年級及領域因素系統性的影響，顯示在不同年級及不同領域，運用資訊能力的評分信度皆能達到高的一致性。可靠性係數皆略低於類推性係數但結果近似，唯在自然與生活科技領域在四上、六上和六下和類推性係數的差距較大，顯示評分者給分的寬鬆不一。與過去的研究相較，本檔案的信度相當良好。過去的研究多半是寫作檔案，以鄒慧英（民 93）為例，其檔案分析式總分以學生為測量對象，決策研究評分者為一、二、三人時，其類推性係數分別為 .58、.73、.81，而可靠性係數則為 .44、.61、.70。這個差異除了來自檔案內容的性質不同，還有可能是由於鄒慧英（民 93）的研究是針對整個檔案做評分，而且檔案結構採取非標準化，而本研究是針對檔案內各個作品做評分，檔案內容相當標準化，在 Herman, Gearhart 及 Baker（1993）的研究，檔案的結構也較為統一。除此之外，評分表標準清楚、能適度反應學生表現都是可能原因。

Koretz 等人（1993）主張對實作評量而言，評分者的相關在 .80 以上算是相當好的，而 Gentile（1992）認為檔案評量的係數在 .80 以上算是很好，在 .65 以上算是好的（p.24）。以此標準，本檔案評量的信度已達相當好程度。

表 4 不同學習領域歷程總分之類推性係數與可靠性係數

		學習領域								
		語文			自然與生活科技			社會		
年級學期	評分者人數	1	2	3	1	2	3	1	2	3
三下	相對	.86	.93	.95	.88	.93	.96	.78	.88	.91
	絕對	.86	.93	.95	.86	.92	.95	.74	.85	.90
四上	相對	.87	.93	.95	.77	.87	.91	.86	.92	.95
	絕對	.86	.92	.95	.62	.77	.83	.85	.92	.94
四下	相對	.70	.82	.87	.87	.93	.95	.84	.91	.94
	絕對	.70	.82	.87	.85	.91	.94	.79	.88	.92
五下	相對	.79	.89	.92	.90	.95	.96			
	絕對	.79	.88	.92	.90	.95	.96			
六上	相對	.70	.82	.87	.76	.86	.90			
	絕對	.68	.81	.86	.64	.78	.84			
六下	相對	.77	.87	.91	.86	.92	.95			
	絕對	.74	.85	.90	.62	.76	.83			

二、歷程向度評分之信度

將各歷程向度的分數加總，適用於總結性評量。上節顯示本檔案在對學生運用資訊能力之整個學期的整體評分其信度達可接受程度。但是檔案評量除了瞭解學生運用資訊能力的整體表現之外，也想知道學童對六個歷程向度的發展情形。因為學生在六個歷程向度的表現情形，能幫助教師適時調整教學內容，設計新的作業，以強化學生表現較差的歷程向度。因此，有必要瞭解歷程向度得分的信度，表 5、6 分別呈現中年級與高年級歷程向度評分各變異數成分估計分析。

表 5 歷程向度評分之變異數成分估計（中年級組）

語文	三下		四上		四下	
變異來源	變異量	百分比	變異量	百分比	變異量	百分比
學生 p	.0077	11.26%	.0057	16.00%	.0004	1.93%
評分者 r	.0000	.00%	.0000	.00%	.00001	.04%
向度 d	.0157	23.05%	.0049	13.95%	.0007	3.30%
學生 × 評分者 pr	.0010	1.47%	.0000	.00%	.0006	2.90%
學生 × 向度 pd	.0257	37.70%	.0150	42.35%	.0150	72.50%
評分者 × 向度 rd	.0061	8.90%	.0008	2.29%	.0000	.0000
學生 × 評分者 × 向度	.0120	17.62%	.0090	25.41%	.0040	19.33%
自然與生活科技	三下		四上		四下	
學生 p	.0109	18.44%	.0079	11.38%	.0056	8.68%
評分者 r	.0002	.30%	.0024	3.45%	.0002	.31%
向度 d	.0064	1.78%	.0189	27.13%	.0285	44.53%
學生 × 評分者 pr	.0003	.56%	.0008	1.19%	.0003	.52%
學生 × 向度 pd	.0250	42.13%	.0233	33.41%	.0197	3.74%
評分者 × 向度 rd	.0005	.83%	.0024	3.39%	.0017	2.70%
學生 × 評分者 × 向度	.0160	26.96%	.0140	2.05%	.0080	12.51%
社會	三下		四上		四下	
學生 p	.0045	9.07%	.0075	16.31%	.0149	2.66%
評分者 r	.0004	.85%	.0001	.26%	.0007	1.00%
向度 d	.0220	44.71%	.0111	24.15%	.0169	23.35%
學生 × 評分者 pr	.0006	1.21%	.0003	.54%	.0015	2.08%
學生 × 向度 pd	.0120	24.20%	.0190	41.33%	.0260	36.01%
評分者 × 向度 rd	.0010	2.02%	.0000	.00%	.0002	.28%
學生 × 評分者 × 向度	.0090	18.15%	.0080	17.40%	.0120	16.62%

中年級組的歷程向度評分的主要變異來源為學生、向度、學生與向度的交互作用以及學生、向度與評分者三者的交互作用。但是這四者所佔變異百分比的多寡，因學習領域，不同學期而有所不同。在自然與生活科技和社會領域，最大的分數變異成分來自向度或向度與學生的交互作用。向度佔分數變異來源的兩成到四成左右，顯示學生在不同向度的表現很不一樣，歷程向度評分有其必要性。此外，在各個領域學生和向度的交互作用約佔變異來源的兩成到四成間，意味不同學生擅長的運用資訊能力向度是不同的。從表 5 看來，評分者、評分者與學生、評分者與向度的交互作用所能解釋分數變

異的百分比相對而言都是小的。一般人面對檔案評量通常對評分者的信度抱持懷疑的態度，這個研究顯示若評分者瞭解所評的能力，適當設計作業，評分者一致性可達接近 .80 的水準。

高年級組的歷程向度評分之變異成分分析結果大抵和中年級相似，都是以學生、向度、學生與向度的交互作用以及學生、向度與評分者三者的交互作用四者為主要變異來源，四者所佔變異百分比的大小，也因學習領域，不同學期而有所不同。高年級組比較特別的發現是六年級上學期的語文與自然與生活科技兩個領域，評分者與向度的交互作用可以解釋 28.53% 和 13.54% 的分數變異，顯示不同評分者的給分因向度不同有所不同。六上的語文領域由於其中一份作業過於容易，很多學生得到滿分，可能因而使得向度分數不穩定。

評分者、評分者與學生以及評分者與向度對分數變異影響小，這個結果和鄒慧英（民 93）研究發現雷同，也與張麗麗（民 91）四個寫作檔案向度中的兩個向度之評分變異成分分析結果相似。

表 6 歷程向度評分之變異數成分估計（高年級組）

語文	五下		六上		六下	
	變異量	百分比	變異量	百分比	變異量	百分比
學生 p	.0146	28.10%	.0040	11.92%	.0104	17.03%
評分者 r	.000	.00%	.0000	.00%	.0000	.00%
向度 d	.0066	12.64%	.0027	7.92%	.0199	32.64%
學生 × 評分者 pr	.0038	7.23%	.0000	.00%	.0016	2.62%
學生 × 向度 pd	.015	28.90%	.0023	6.95%	.0130	21.29%
評分者 × 向度 rd	.0010	1.93%	.0096	28.53%	.0021	3.50%
學生 × 評分者 × 向度	.0110	21.20%	.0150	44.68%	.0140	22.92%
自然與生活科技	五下		六上		六下	
學生 p	.0165	23.01%	.0031	9.75%	.0083	22.11%
評分者 r	.0000	.00%	.0011	3.57%	.0035	9.24%
向度 d	.0163	22.85%	.0034	1.95%	.0127	33.66%
學生 × 評分者 pr	.0003	.35%	.0005	1.59%	.0008	1.99%
學生 × 向度 pd	.0253	35.36%	.0100	31.89%	.0063	16.80%
評分者 × 向度 rd	.0012	1.71%	.0042	13.54%	.0021	5.59%
學生 × 評分者 × 向度	.0120	16.74%	.0090	28.70%	.0040	10.61%

由上述兩個表發現學生與向度的交互作用、以及學生、評分者與向度的三因子交互作用對分數變異影響頗大，因此，可推論歷程向度分數的類推性係數必然不理想。表 7 呈現歷程度數為 6，只有一位評分者評分時的類推性係數。由於，在作業設計時，最多只能要求學生執行六個運用資訊歷程，因此，表 7 的結果可視為每個科目只用一位教師評分，歷程向度評分的最佳情形。從表 7 可知，歷程向度評分一致性太低，四下的語文領域之歷程向度類推性係數甚至只有 .10，表 5 歷程向度評分之變異數成分估計顯示學生與向度的交互作用達 72.5%，學生差異佔總變異數的 1.93%。造成此項結果如前文所述可能是採取小組分派作業的緣故。評分時，小組成員僅有在評鑑歷程向度有不同得分，其餘向度得分相同，由於評鑑歷程向度因學生不同而變化，使得學生與向度的交互作用變異特別大。大致而言，歷程向度評分之類推性係數在 .5、.6 左右，因此，要從檔案評分去推論個別學生的那個歷程向度表現較為弱勢，必須注意歷程向度分數的不一致性。

表 7 歷程向度評分之類推性係數（向度數為 6，評分者為 1）

學習領域	三下	四上	四下	五下	六上	六下
語文	.51	.59	.10	.64	.58	.63
自然與生活科技	.60	.53	.53	.72	.45	.77
社會	.53	.61	.66			

三、內外評分者的一致性分析

表 4 的結果顯示評分者一致性尚可接受，由於過去的研究顯示評分者對表現標準的理解，以及評分者的經驗和是否受過訓練會影響信度的高低 (Shapley & Bush, 1999)，學校現場教師有較多評分的經驗，對學童一般的表現較有概念，其評分的一致性可能較高，不過，過去研究發現外部評分者與內部評分者間的一致性反而高於內部評分者間的一致性，可能是任課教師因認識學生反而受到月暈效應的影響（鄒慧英，民 93）。本研究以 Pearson 的積差相關（Pearson product-moment correlation）分析內部評分者間以及內部評分者與外部評分者間的信度。因為各學期各領域分數總分的全距不同，差異頗大，因此不做完全一致百分比（percent of exact agreement）的分析而是比較兩組評分者總分的差異。

由於每個作業都有三位評分員，其中兩位是校內教師，一位是研究助理，因此，內部評分員與外部評分員會有兩個相關值。檢視表 8 可以發現除了六上的語文、三下的社會、四上的自然與生活科技，教師間的評分相關都能達到 .8 和 .9 以上，此外，學校教師間的相關高於教師與研究助理間的相關，有些相關值僅有些微的差異，有些相關值則差距頗大。從此結果可以推論在表 4 的類推係數若只用校內教師，評分者的信度會更高些。值得注意的是在自然與生活科技領域，內部評分者之相關與內外評分者的相關都較為接近，甚至四上和六下校內教師與研究助理的評分相關高於校內教師間的相關。在這個領域校內的兩位評分員是擔任中高年級的自然科任教師，中年級的教師並不認識高年級的學生，高年級的教師也不認識中年級的學生，因此，前述的月暈效應無法解釋這個結果。一個可能的原因是，在其他領域校內教師是擔任同一年級的課程，而自然與生活科技領域的老師一為中年級，一為高年級，同一年級的教師面對相同年齡學童，在評分時能夠達到較一致的結果。但並不是所有的作業都受此項因素的影響。

表 8 內外評分者之積差相關

	語文		自然與生活科技		社會	
	內部	內外	內部	內外	內部	內外
三下	.95	.86/.78	.90	.89/.86	.77	.82/.78
四上	.94	.84/.84	.71	.81/.79	.98	.80/.79
四下	.96	.60/.55	.87	.90/.86	.99	.90/.90
五下	.85	.81/.79	.91	.94/.87		
六上	.66	.63/.84	.81	.78/.82		
六下	.90	.70/.70	.84	.88/.92		

教師的評分與研究助理的評分絕對分數是否存在著差異呢？表 9 呈現內外評分者給分之差異分析。校內教師評分的相關多半在 .8、.9；校內教師與研究助理的評分相關則較為分散，從最低的 .55 到最高的 .94。大致而言，教師給分顯著的高於外部評分者，尤其在自然與生活科技領域最為明顯。這也可以解釋何以表 4 中自然與生活科技領域的可靠性係數偏低。校內教師給分高於研究助

理，可能是因為研究助理缺乏教學經驗，較不能掌握該年齡層學童一般的表現水準。不過，Supovitz, MacGowan III 和 Slattery (1997) 的研究也發現在閱讀的評量上，任課教師的評分高於外部評分者，在該研究外部評分者同樣是有經驗的教師。另一個可能的原因是任課教師在判斷學生表現時，會受到學生在其他方面的表現影響，在研究後的訪談中，教師們常提及學生們很用心，進步多，也許教師在評分時，將學生進步與努力的因素考慮進去。

表 9 內外評分者分數之差異分析

		語文			自然與生活科技			社會		
		平均值	標準差	t 值	平均值	標準差	t 值	平均值	標準差	t 值
三下	內	19.23	3.06	.460	19.38	3.52	2.97**	20.43	2.08	.522
	外	19.10	3.40		18.75	3.33		20.34	2.17	
四上	內	50.44	4.84	-.831	31.82	3.81	7.74**	33.49	4.01	-2.20*
	外	50.70	4.45		29.65	4.15		34.15	3.53	
四下	內	26.83	2.15	-1.30	39.14	4.45	4.45*	22.85	3.30	5.99**
	外	27.11	1.34		37.98	4.69		21.51	4.12	
五下	內	21.24	4.54	-2.28*	13.99	2.85	-.190			
	外	22.15	3.41		14.02	2.55				
六上	內	26.41	2.28	3.354*	33.88	2.67	5.60**			
	外	25.67	2.59		32.44	3.55				
六下	內	19.26	3.68	3.213*	44.56	5.28	13.86**			
	外	18.16	3.19		41.21	5.66				

* $p < .05$, ** $p < .01$

結論與建議

透過各學習領域培養基本能力，是九年一貫課程非常重要的一環。過去的檔案評量研究皆是以學習領域為主，主要集中在讀寫檔案。國小學童的資訊素養檔案結合語文、自然與生活科技以及社會學習領域設計相關作業，評量課程所欲培養的運用資訊基本能力。檔案信度研究顯示不論年級、不論領域，在一位評分者，且領域內的作業量 2-3 個的情況下，類推性係數能達到 .70-.91 的水準，可靠性係數則在 .60-.90 之間。這項研究結果有助於減低一般人對檔案評量信度的疑慮。

向度歷程評分的分析顯示向度和向度與學生能力的交互作用各自佔分數變異的兩成到四成左右，顯示學生在解決資訊問題的各歷程能力程度不同，透過適當作業設計能知道學生在各歷程向度的優劣能力，但是目前檔案的設計在歷程向度評分的類推性上仍有改善的空間，不適合用於總結性評量，但在教學決策具有相當的參考價值。

研究顯示透過適當檔案設計，任課教師能一致地評量學生作品。各學習領域教師間評分相關達 .80 以上，變異成分的分析顯示，評分者，評分者與學生、與向度的交互作用所佔分數變異的比例相對是微小的。多半的情形下，校內教師間的評分相關高於教師與外部的研究助理間的相關，校內教師評分也高於外部評分者。從本研究無從得知校內或校外的評分效度較高，不過，內外評分員的實質差異並不大。

雖然參與本研究的教師人數僅有六位，但在估計類推係數和可靠性係數時，本研究皆將三位研究助理的評分納入計算，研究助理是師範學院在學的大學生，較缺乏教學經驗，也較不能掌握該年齡層

學童的一般表現，因此，預期其評分會與校內教師有較大落差。由於對每份作業都只有三個評分，若其中一個分數偏離其他分數應會使評分者的變異擴大，換言之，如果完全是由校內教師評分，應會得到更佳的信度。據此可推論若是以一位校內教師評分，本檔案可為學生的運用資訊能力提供可信的測量結果。至於本研究結果能否推論到其他學校教師使用此一檔案信度情形，則有待進一步的探究。任課教師與外部評分員在評分上的差異來源以及這些差異對評分效度的影響亦值得未來研究加以探討釐清。

信度的分析結果也提供未來設計檔案系統作業時值得參考的要項。第一，以理論為架構（在本研究是大六教學法）結合學習領域內容的檔案系統，可在有限的作業數量下，達到不錯的信度。第二，由於每個學習領域的作業數少，因而作業難度的掌握更形重要，過於容易的作業（如六上的語文領域的作業），會使得信度偏低。第三，分組作業若缺乏個別計分的項目，也容易造成信度偏低的狀況（如四下語文領域的作業指派）。第四，評分者最好是熟知該年級學童的教師。

本研究設計時，考慮教學現場的現實，各領域的評分是以該學習領域的任課老師為主。由於自然與科技學習領域是由科任教師擔任，沒有教師同時擔任三個領域的課程，評分者與學習領域非屬交叉設計，是在分析檔案分數的變異成分時，無法直接估計學習領域因素對分數變異的影響大小，只能從各學習領域個別的類化係數頗為接近，推論學習領域不是影響分數變異的重要因素。這是本研究的主要限制，沒有請教師同時評所有領域的作業，主要是檔案評量需要長時間蒐集作品，從作業的設計、蒐集到評分若由校內任課教師負責較為可行，教師們因為時間因素，無法配合研究需要進行跨領域的作業評分。

其次，本研究採取方便取樣，雖然參與研究的師生都未曾接觸過資訊素養檔案評量，然此一結果是否可類推到其他學校，仍待後續研究的探討。此外，每個年級參與研究的學生總數為 70 名，研究之初，五年級學生缺繳作業情形嚴重，在語文領域甚至只收到 41 份完整可分析作業，經與教師溝通，之後的兩個學期，可分析的作業數大都能維持在 60 份以上。作業未能完全回收是此一研究另一個限制，雖然從結果並未發現三下和五下的結果和其他兩個學期有明顯的差異。教師們表示學生平日就有作業缺交問題，因此，未來進行檔案評量研究，要特別注意此一問題。

對於打算使用此一檔案評量系統之教師，根據信度研究之結果，提出六點注意事項供使用時的參考。

第一，以學生在各領域的總分做決策。若做為總結性評量的一部份分數，一位經過訓練的現場教師做為評分員即可，但若為重大決策，則須增加到三位評分員，才能使絕對與相對決策的信度都能維持在 .80 以上。

第二，檔案評量系統採融入課程的設計。隨著課程內容的變化，檔案內作業的設計必須隨之調整，由於每個領域的作業數目只有 2 到 3 份，因此，作業難度的掌握會嚴重影響到檔案的信度。這是使用此一檔案評量系統應特別注意之處。

第三，教師設計作業時，應同時發展評分標準，提供評分範例，如附錄一的評量表所示，以確保評量系統的信度。

第四，分組合作是教學常用的課程設計，當指派分組作業時，也須兼顧讓組內成員得以呈現個別的努力與成果，給予個別的評分，如此評分信度才能穩定。

第五，評分員的選擇上應考慮評分員對評分對象的熟悉度以及是否為校內人員。外部評分員可能會給予較低的評分，擔任不同年級教師間的評分一致性可能會較差。

第六，本檔案系統個別學生在各個資訊歷程向度的得分，信度仍不夠理想，因此，不宜以歷程向度分數診斷學生之優弱勢，僅可視為參考資訊。

參 考 文 獻

- 李德竹 (民 89)：資訊素養的意義、內涵與演變。圖書與資訊學刊，35 卷，1-25 頁。
- 吳美美 (民 89)：資訊素養教育與 e 世代教學。國家圖書館非同步遠距教學網站：<http://cu.ncl.edu.tw/learn/index.php>。檢索日期：民 92.01.27。
- 陳伯璋 (民 88)：九年一貫課程的理念、內涵與評析。「國民中小學課程教學研討會」宣讀之論文 (台北)。
- 教育部 (民 92)：國民中小學九年一貫課程綱要。國民教育司：<http://140.122.120.230/ejedata/kying/20031241215/index.htm>。檢索日期：民 92.02.08。
- 張郁雯 (民 93)：運用資訊能力之檔案評量系統之發展研究。國立台南教育大學主辦「科技化測驗與能力指標評量國際學術研討會」宣讀論文 (台南)。
- 張麗麗 (民 91)：檔案評量信度與效度分析—以國小寫作檔案為例。教育與心理研究，25 卷，1 期，1-24 頁。
- 鄒慧英 (民 93)：讀寫檔案的信度與評分者一致性。國立台南教育大學主辦「科技化測驗與能力指標評量國際學術研討會」宣讀論文 (台南)。
- 賴麗珍 (民 88)：國高中課程與資訊素養教育。國立台灣師範大學主辦「資訊素養與終身學習社會國際研討會」宣讀論文 (台北)。
- American Association of School Librarians (1998). *Information power: Learning and teaching principles of school library media programs*. Retrieved February 8, 2003, from <http://www.ala.org/aasl/ip/principles1.html>
- American Library Association (1989). American Library Association Presidential Committee on Information Literacy: Final report. Retrieved February 8, 2003, from <http://www.infolit.org/documents/89Report>.
- Arter, J. A., & Spandel, V. (1992). Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, 11(1), 36-44.
- Bridgeman, B., Chittenden, E., & Cline, F. (1995). *Characteristics of a portfolio scale for rating early literacy*. Princeton, NJ: Educational Testing Service.
- Eisenberg, M. B., & Berkowitz, R. E. (1990). *Information problem-solving: The Big Six Sills approach to library and information skills instruction*. Norwood, NJ: Ablex.
- Eisenberg, M. B., & Berkowitz, R. E. (1999). *Teaching information & Technology skills: The Big6 in Elementary Schools*. Worthington, OH: Linworth.
- Fourie, I. (1999). Using portfolio assessment in a module in research information skills. *Education for Information*, 17(4), 333-352.
- Frechtling, J. A. (1991). Performance assessment: Moonstruck or the real thing? *Educational Measurement: Issues and Practice*, 10(4), 23-25.
- Galzer, S. M., & Brown, C. S. (1993). *Portfolios and Beyond: Collaborative assessment in reading and writing*. Norwood, MA: Christopher-Gordon.
- Gentile, C. (1992). Exploring new ways for collecting school-based writing: NAEP's 1990 portfolio study. Washington, DC: Office of Educational Research and Improvement.
- Herman, J. L., Gearhart, M., & Baker, E. L. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment*, 1, 201-224.
- Herman, J.L., & Winters, L.(1994).Portfolio research: A slim collection. *Educational Leadership*, 52(2),

48-55.

- 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), 65-80.
- Koretz, D., Stecher, B., & Deibert, E. (1993). *The reliability of scores from the 1992 Vermont portfolio Assessment program (Tech. Rep. No. 355)*. Los Angeles: University of California. CRESST: Center for the Study of Evaluation.
- Kuhlthau, C. C. (1987). An emerging theory of library instruction. *School Library Media Quality, 16*, 23-27.
- Irving, A. (1985). *Study and information skills across the curriculum*. London: Heineman.
- Joyce, M. Z., & Tallman, J. I. (1997). *Making the writing and research connection with the I-search process*. New York: Neal-Schuman.
- LeMahieu, P. G., Gitomer, D. H., & Eresh, J. T. (1995). Portfolios in Large-scale assessment: Difficult but not impossible. *Educational Measurement: Issues and Practice, 14*(3), 11-28.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessments be psychometrically sound? In J. B. Baron & D. P. Wolf (Eds.), *Performance-based student assessment: Toward access, capacity and coherence* (pp. 126-143). Chicago: National Society for the Study of Education.
- Myers, M., & Pearson, P. D. (1996). Performance assessment and the literacy unit of the New Standards Project. *Assessing Writing, 3*(1), 5-29.
- Pappas, M. & Tepe, A. (1997). *Introduction to the pathways to knowledge*. Follett's information skills model. McHenry, IL: Follett Software.
- Shapley, K. S., & Bush, M. J. (1999). Developing a valid and reliable portfolio assessment in the primary grades: building on practical experience. *Applied Measurement in Education, 12*(2), 111-132.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A Primer*. Newbury Park, CA: Sage Publication.
- Stripling, B. K., & Pitts, J. M. (1988) *Brainstorm and blueprints: Library research as a thinking process*. Englewood, Co: Libraries Unlimited.
- Supovitz, J. A., MacGowan III, A., & Slattery, J. (1997). Assessing agreement: An examination of the interrater reliability of portfolio assessment in Rochester, New York. *Educational Assessment, 4*(3), 237-259.
- Tierney, R. J., Carter, M. A., & Desai, L. E. (1991). *Portfolio assessment in reading-writing classroom*. Norwood, MA: Christopher-Gordon Publishers.

收 稿 日 期：2007 年 02 月 15 日

一稿修訂日期：2007 年 08 月 20 日

接受刊登日期：2007 年 08 月 21 日

附錄一 資訊素養檔案作業範例

家鄉的地名—小小搜查員

班級：____ 姓名：____ 座號：__ 日期：93 年__ 月__ 日

小朋友，我們的學校位在台北市，如果我們想要了解台北市裡的地名有哪些，那麼我們就必須動手查詢有關地名的書籍或資料囉！你知道如何利用圖書館的網路查詢系統找到你要的書籍或資料嗎？請小朋友依照下列的步驟，試著查詢各圖書館中有關**台北市地名**的書籍。

一、我想要查詢哪一個圖書館的書籍？（請選擇一個打勾）

- 台北市立圖書館
 國立台北師範學院圖書館
 本校圖書館

二、我們要找的是有關台北市地名的書籍，在**中國圖書分類法**中是屬於哪一類呢？

- 000 總類 100 哲學類 200 宗教類
 300 自然科學類 400 應用科學類 500 社會科學類
 600 中國史地類 700 世界史地類 800 語文類
 900 美術類

三、我進入了_____圖書館的網路圖書查詢系統，我使用的查詢關鍵字是_____。

四、我所找到的有關台北市地名的書籍名稱：

_____。（列出一本書即可）

這本書的書籍條碼或者索書號：_____。

小朋友，你已經學會如何使用圖書館的網路查詢系統查詢書目了！現在要請你查詢**台北市裡兩個地名的由來**，只要簡單的說明就可以了，你可以使用前面學過的查詢方法，也可以使用其他你覺得更方便的搜尋方式喔！

五、我使用的查詢方法：（請勾選）

- 利用圖書館網路查詢系統，再到圖書館查閱書籍。
 利用搜尋引擎進行查詢，我使用的關鍵字是：_____。
 其他方式：_____。

六、我的查詢結果：

◎地名一：_____。

由來：_____。

◎地名二：_____。

由來：_____。

◎資料來源（書籍名稱或者網址）：

七、完成了這份作業，回想一下，蒐集資料的步驟有：

1. _____

八、在蒐集資料的過程中我遇到的困難以及解決的方法：

九、學習心得

評量表

六大資訊能力	3 分	2 分	1 分
一、定義問題			
1. 能列出位於『台北市』裡的地名。	<input type="checkbox"/> 2 個	<input type="checkbox"/> 1 個	<input type="checkbox"/> 0 個
三、尋找資訊			
1. 能透過網際網路搜尋到需要的資料。(針對步驟三&四)		<input type="checkbox"/> 是	<input type="checkbox"/> 否
2. 能夠利用圖書館查詢系統或搜尋引擎查詢方式(或其他方法)取得資料。(針對步驟五)		<input type="checkbox"/> 是	<input type="checkbox"/> 否
四、使用資訊			
1. 能正確並清楚的寫出兩個地名的由來(請見下面例子)	<input type="checkbox"/> 兩個皆清楚、正確	<input type="checkbox"/> 只有一個地名的由來清楚、正確	<input type="checkbox"/> 兩個皆不清楚、不正確
2. 能列出資料來源：書籍名稱或網址。		<input type="checkbox"/> 是	<input type="checkbox"/> 否
六、評鑑			
1. 能列出搜尋資料的步驟。(針對步驟七)(請見下面例子)	<input type="checkbox"/>		<input type="checkbox"/>
2. 能寫出搜尋資料解決遇到的困難與解決方法。(針對步驟八)(請見下面例子)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

四、使用資訊

1. 能正確並清楚的寫出兩個地名的由來

3 分 --- 兩個地名由來「皆」清楚、正確。

※由來一：萬華古稱艋舺，其名源自於莽葛或文甲，為平埔凱達格蘭族之音譯。1709 年陳賴章率領福建人民移居於此，見舟木會集，所以取名為艋舺。

由來二：士林舊名為八芝連林，又稱芝蘭。八芝連林是藩語，是溫泉的意思，至清末改稱士林，是「士子如林」的意思。

2 分 只有一個地名的由來清楚、正確

1 分 --- 兩個地名由來「皆不」清楚、正確。

※由來一：天母教侍奉神明為「天」上聖「母」，「天母」之名因而傳開，成為本地的代稱。(不正確)

由來二：他們豎「木」為「柵」，防禦侵襲，因此得又由於「柵」的音有殺氣。(不清楚)

六、評鑑

1. 能列出搜尋資料的步驟。

3 分 --- 能具體清楚寫出蒐集資料的歷程。

※ (1) 我先上圖書館的網站，利用關鍵字查詢書籍的檢索號 (2) 再到圖書室借閱或直接參考

※ (1) 搜尋 Google 網站 (2) 輸入關鍵字地名 (3) 台北地名的由來 (4) 搜尋

※ (1) 打開奇摩網 (2) 打台北市地名 (3) 再點搜尋就會到台北市地名的網站 (4) 點北市地名 (5) 再點選北市地名 (6) 就可以看到台北 12 個區的由來

1 分 --- 沒有具體清楚寫出蒐集資料的歷程。

※ (1) 上圖書館找 (2) 上網找

※ (1) 先去圖書館 (2) 找出要用的書及資料 (3) 再選出有用的資料寫上去 (4) 完成了！

※ (1) 要先知道如何找到你要的那一類 (2) 要先知道你要的書 (3) 先把找好的書放著 (4) 再慢慢看，不是你要找的書就放旁邊

2. 能寫出搜尋資料解決遇到的困難與解決方法。

3 分 --- 問題與解決方法具體而清晰。

※我們用台北市立圖書館的館藏目錄查詢系統的館藏目錄查詢，裡面的題名我們打了「台北市」，後來因為這個題名的範圍太大，所以我們才回目錄查詢打上關鍵詞和題名，才查到我們想要的東西。

※困難：如何選擇對的關鍵詞。解決：一直試不同的關鍵詞。

※上網查不到臺北的故事這本書，後來我把臺北的故事打成臺北市的故事，我就查到了。

※因地名由來範圍太大，所以輸入關鍵字錯，所以再加入台北才找到。

2 分 --- 困難與解決方法寫得不夠具體。

※困難：我跟媽媽查資料的時候，雖然查了圖書館的網路系統還是找不到。解決：然後我們就去問圖書館的阿姨。

※我在蒐集資料的時候，遇到網路上查不到台北市地名的由來，後來我慢慢的終於找到了。

我在市立圖書館時一直找不到有關台北市地名的書，後來我問了圖書館的阿姨，他告訴我在哪裡，我就找到了。

1 分 --- 問題的解決與資料蒐集歷程關聯性不強。

※我找不到資料的時候開始覺得很煩，不想寫了，姊姊對我加油，我開始又有了精神，後來我才找到了我要的資料。

※我不知道地名由來的大意，我就問阿姨怎麼寫，後來阿姨就教我怎麼寫大意。

※在找資料時常常找不到適合的資料。請大家幫忙找。

Bulletin of Educational Psychology, 2008, 39, Special Issue on Test and Measurement, 43-60
National Taiwan Normal University, Taipei, Taiwan, R.O.C.

The Reliability of Information Literacy Portfolio Assessment in the Primary Grades

Yuwen Chang

National Taipei University of Education
Department of Education

This article reports on the reliability of information literacy portfolio assessment in the primary grades. Three central questions were examined: (1) how reliable are total scores of information portfolio? (2) how reliable are scores of big six skills? and (3) what are probable sources of unreliability? Do teachers who rate their own students' work systematically score differently than do outside raters? The framework of information portfolio was based on the big six skills approach. This assessment was based on a sample of approximately 70 portfolios integrated three learning areas for students in Grades 3 and 5: language arts, mathematics, and science and technology. The teachers were required to collect each student's portfolio for 3 semesters. The products of portfolios in each learning areas were rated by 3 raters: the student's classroom teacher, a teacher of the same primary school, and an external reviewer. Results indicated that most of the generalizability coefficients for total scores were above 0.75 across learning areas and semesters with a single rater. The dependability coefficients revealed a similar pattern with slightly lower size. Variance in scores of each processing dimension was attributed to individual difference, processing dimension, interaction of individual difference and dimension, and interaction of individual difference, dimension, and rater. The generalizability coefficients for the six-dimensional task were around 0.5 to 0.6. The results indicated that instructional planning based on scores of specific information skill should be questioned. In general, the correlations between classroom teachers were higher than those between classroom teachers and external raters. Classroom teachers tended to rate student portfolios higher than did external raters; however, these differences were small. This study showed that given a solidly structured portfolio, teachers can reliably rate their students' work.

KEY WORDS: dependability coefficient, generalizability coefficient, information literacy, portfolio assessment, reliability.