

第二章 文獻探討

為探討文書處理實作試題的難度值，本章蒐集相關文獻資料，分為三部分，即實作試題、技能檢定、及試題難度，茲分別說明於后。

第一節 實作評量

由於教學的範圍非常廣泛，評量的方式必須配合學科特性及教材內容擁有多樣的面貌。目前資訊化的時代，教學評量更注重彈性、變通與多元化，並且強調動態過程的評量。

「實作評量」，係指根據學生實際完成一項特定任務或工作表現所作的評量。這些任務或工作，可能是實際操作、口頭報告、科學實驗、數學解題、寫作等。實作評量係透過直接的觀察學生表現或間接的從學生作品去評量。此種評量方式異於傳統的紙筆測驗，它重視實作的能力，從實際的行為表現來評量，而非依賴筆試來決定。近年來評量的方式愈來愈多元，不僅重視學習的成果，更重視學習過程的評量，使實作評量的應用愈來愈廣。

實作評量崛起於一九九〇年代早期，其受重視的原因有三：

1. 認為選擇式反應測驗(selected-response tests)只能測出學生所學再認知的部份，無法測出學生較高層次的思考能力。實作評量是任務導向的評量，較接近學生實際應用所學的情境，較能測出學生的應用或實際能力。
2. 受到認知心理學的影響，認為學生應該兼顧內容知識和過程知識，而過程知識的獲得，需要經過實作，才能夠呈現。

3. 考試領導教學普遍存在，傳統的紙筆測驗，會影響到教師教學偏重於學生記憶的學習，對教學有負面影響；若是透過實作評量的方式，即可以轉移教師教學活動的方式。

實施實作評量的情境愈接近真實情境，愈能顯現出學生真正學會的技能。然而，有時礙於發生學習時的真實情境無法再複製一次，因此，只好使用模擬的情境。模擬的情境愈接近真實情境，則實作評量的結果將愈具有教學期望結果的代表性，愈能符合教師期望學生真實學會的重要教學目標和內容。因此，根據施測情境的真實性程度來分，在教學情境下常用的實作評量可以分成下列五種類型：

1. 紙筆表現(paper-and-pencil performance)：有別於傳統所使用的紙筆測驗，它是一種比較強調在模擬情境中應用知識和技能的評量方式。
2. 辨認測驗(identification test)：是指由各種不同真實性程度的測驗情境所組合成的一種評量方式。
3. 結構化表現測驗(structured performance test)：是指在標準且有控制的情境下進行評量的工具。
4. 模擬表現(simulated performance)：配合或替代真實情境中的表現，局部或全部模擬真實情境而設立的一種評量方式。
5. 樣品範本(work sample)：在實作評量的各種類型中，樣品範本是真實性程度最高的一種評量方式，它需要學生在實際作業上，表現出所要測量的全部真實技能。

實作評量的評量內容，可以歸納為兩方面：過程的評量及成果的評量。而其計分的工具有下列五種：(1) 系統的觀察和軼事記錄(systematic observation and

anecdotal records), (2) 檢核表(checklists), (3) 評定量表(rating scales), (4) 成果量表(product scales)及(5) 檔案(portfolios)。「檢核表」是一組列出表現或成果的測量項目, 並且提供簡單記錄「是」或「否」判斷的資料表。觀察者只要逐一核對每個被觀察的表現是否發生或出現並作紀錄即可。丙級檢定的文書處理實作試題所採用的即是檢核表的計分方式。提供評分者一份評分標準, 讓評分者依照評分標準中所列出的評分項目逐項評分。然而, 檢核表有其使用上的兩個限制:

1. 它只提供教師「是」和「否」兩種選擇。而實際上, 學生的表現, 可能是介於這兩個極端之間, 並非兩種絕然的狀態。檢核表強迫作二分法的選擇時, 所蒐集到的資料可能有失精確。
2. 評量結果很難以一個總分來表示。檢核表評量的結果是一種質性的資料, 而非量化的資料。

目前的文獻中大部分都是採用下列三種量化的方式來轉換檢核表的非量化評量結果。

1. 取測量項目中正確表現的百分比作為量化資料

若要進行量化, 即以一個客觀的總分來代表評量結果。教師必須先計算檢核表中打勾的記號數目, 再將其除以全部的評核項目數, 得到一個百分比值, 此值即被當成是代表檢核表評量結果的量化指標。此種量化方式的特色是——(1) 檢核表中每一個測量項目的分數權重均相同; (2) 分數介於 0 到 100 之間, 滿分是 100 分, 此與我國目前各級學校的評分習慣相同。

2. 依測量項目中表現正確的數量作分級

由教師事先設定評定學生表現行為的標準, 以下表(表 2-1)為例:

表 2-1 學生行為表現的分級標準表

等第	表 現
卓越	如果評定結果達到 10 至 12 項表現標準
優良	如果評定結果達到 7 至 9 項表現標準
普通	如果評定結果達到 4 至 6 項表現標準
低劣	如果評定結果達到 1 至 3 項表現標準

藉此方法教師便可以將學生的表現轉換為上述的等第，用來解釋和描述學生表現行為的好壞。此種量化方式的特色是——(1) 測驗項目的權重較有變化；(2) 學生得到的測驗結果單純且明確。

3. 合格與不合格

以電腦丙檢為例，檢定的評分標準中明列各測驗項目的扣分方式，事先設定的評定標準為：扣分未達 40 者「合格」；扣分達 40 以上則「不合格」。每個測驗項目的扣分標準不同，且測驗結果只有合格與否，非常明確。

事實上，實作評量除了評分方式外，仍有一些技術性的問題亟待解決，例如，所花的時間較多，評分標準的客觀性，教師本身的評量能力，以及所選定的工作或任務是否足以能代表學生整體的表現等問題，都可能受到質疑。所以，實作評量的應用及推廣，仍有一些限制(吳清山、林天祐，2000)。

實作評量對於改進傳統的評量方式，的確有其實質的意義和價值，尤其重視學生的實際工作的表現，轉移過去所強調的紙筆測驗，就評量角度而言，實為一大突破。

第二節 技能檢定

本研究所採用的試題，是以電腦丙級檢定術科第二站的文書處理實作試題作為試題的基本模式，因此，在此節針對技能檢定與文書編輯操作技能試題兩部分加以說明，並以此界定本研究的試題模式。

一、技能檢定

技能檢定(skill test)是指根據公訂的工作規範(job specification)對專業技術人員所須具備的知識及技能，依一定的程序予以測驗，並由據權威及公信力的機關核發證書，以茲證明其專業能力的一種制度(施能仁，1994；陳 階，1980；蕭錫錡，1993)。

職業證照(vocational certificate)是指根據技能檢定核發證書，以作為從業憑證，並依法給予某種程度的保障(職訓局，1994)。證照代表對一個人在某個專業領域中的認證。透過證照可以讓別人對持有證照者所提出的建議或意見是值得尊重與採納的。職業證照的推行，是技職教育培養國家人力資源的工作中很重要的一部分。其他國家的文獻中亦顯示，職業證照制度是世界各國提高生產力，保障國民就業的重要措施之一。先進國家均已建立完善的證照制度，其成功要素有六：(1) 證照種類及範圍廣、核發證照數量多、(2) 企業積極參與、(3) 相關法令完備、(4) 技能檢定具公信力及權威性、(5) 職業教育與技能檢定密切配合與(6) 社會各界的認同與支持(黃燈淵，2003)。

目前國內各項證照中，以行政院勞工委員會所主辦的全國技術士檢定考試最受肯定，持有證照的人數也多。業界對持有技術士證照的專業人士，多予肯定或認同，此一檢定屬於國家級技術士檢定。一般技職體系的學校，均設定目標，讓學生在畢業前能取得技術士證照。近年來勞委會更因應資訊化的進展，

增加數項資訊相關的檢定職類和檢定職級，職類如，電腦軟體應用、電腦軟體設計、電腦硬體安裝；職級如，甲、乙、丙級。本研究的範圍是電腦丙檢中，術科第二站文書處理實作測驗的部分。

二、文書編輯操作技能試題

文書處理是電腦軟體應用中最普遍的一種，目前國內中小學電腦教育，普遍以文書處理為入門課程。文書處理操作技能的試題型態是多樣化的，可以是一般紙筆測驗，也可以是以電腦實際完成文書處理操作的實作測驗。以下將簡介全國技術士技能檢定在電腦丙級檢定中所採用的文書處理評量試題。

(一) 文書處理操作技能的評量

電腦的操作課程若以選擇題或填充題等紙筆測驗作為測驗方式，往往不能測得學生的實際應用能力。採用類似檢定考試的「作品完成測驗」即實作評量題型則較適合。無論是紙筆測驗或實作測驗，大部分文書編輯測驗通常指定某一編輯軟體，例如，微軟公司的 Microsoft Word。本研究即以文書編輯課程為測驗主體，跨越各編輯軟體的藩籬。故採用不限制使用何種軟體的電腦丙檢術科第二站文書編輯實作試題作為研究的基礎。

文書處理軟體的檢定是操作性的技能檢定，同時具備難度測驗(power test)與速度測驗(speed test)的特性，即評量學生獲得知識與技能的程度的同時評量學生反應的速度。本研究僅探討難度測驗的領域，所有實驗將給予受試者足夠的時間完成文書編輯操作。

(二) 電腦丙檢的術科第二站試題

目前電腦丙檢是高中職學生參與最多的國家級檢定考試，其術科部分是以電腦文書處理軟體為測驗主體。一般考生均學習微軟的 Word 系統參

加考試，但考試辦法中並不限制考生使用的軟體。關於電腦丙檢的術科第二站試題的測驗方法、評分方式及測驗元素，分項說明如下：

1. 測驗方式

- (1) 測驗時間：45 分鐘
- (2) 測驗項目：軟體安裝能力、輸入能力、編輯能力、列印及儲存能力等。
- (3) 題庫包含 30 題，隨機抽 1 題測驗。
- (4) 測驗軟體：不限定測驗時使用的軟體。
- (5) 試題型式：以完成指定作品為試題型式
- (6) 試題：包含測驗操作說明文件及答案範本文件(如附錄 A)。

3. 評分方式

- (1) 採人工閱卷，比較學生作品與標準答案的相似度。
- (2) 參考評分標準(如附錄 A)列出的操作項目及分數權重，以扣分方式進行評分。

4. 測驗元素分析

可將操作分為必考、必選及選考三部分，分別如表 2-1(a)、表 2-1(b)及表 2-1(c)中所列。其中必選操作為「插入表格」及「插入圖片」二選一(試題範例詳列於附錄 A)。

表 2-1(a) 電腦丙檢所包含「必考」文書編輯測驗元素

測 驗 元 素		測 驗 元 素	
1	安裝應用軟體	14	編輯頁首/頁尾
2	開啟檔案	15	插入頁碼
3	設定紙張大小	16	插入日期/時間
4	設定紙張邊界	17	插入換行符號
5	設定列高	18	列印文件
6	設定字元間距	19	輸入文字
7	設定中文字型	20	中/英文輸入法切換
8	設定英文字型	21	全/半形輸入法切換
9	替換指定標點符號之全半形	22	選取字元
10	設定定位點	23	複製字元
11	設定列首不得出現標點符號	24	搬移字元
12	設定整段文字的首行縮排	25	刪除字元
13	設定整段文字的對齊方式	26	另存新檔

表 2-1(b) 電腦丙檢所包含「必選」文書編輯測驗元素

測 驗 元 素		測 驗 元 素	
圖 片 部 分	1	插入圖片	表 格 部 分
	2	設定圖片位置	
	3	設定圖片大小	
	4	設定圖片框線	
	5	設定文字圍繞圖片的方式	
	6	設定文字與圖片間距	
	1	插入純文字檔	
	2	插入表格	
	3	設定表格邊界	
	4	設定表格欄寬	
	5	設定表格內容水平對齊方式	
	6	設定表格內容垂直對齊方式	
	7	設定表格框線	
	8	設定表格底色	
	9	設定表格字型	

表 2-1(c) 電腦丙檢所包含「選考」文書編輯測驗元素

測 驗 元 素		測 驗 元 素	
1	設定整段文字的左右縮排	12	設定字型粗體
2	設定紙張方向	13	設定字型斜體
3	設定全文書寫方向	14	設定字型加底線
4	設定整段文字的框線	15	設定字型加框線
5	設定整段文字的底色	16	設定字型加底色
6	設定整段文字的框線粗細	17	設定字型大小
7	替換指定文字	18	設定表格文字書寫方向
8	替換指定英文字串之大小寫	19	設定表格文字左右縮排
9	替換指定文字的字型格式	20	設定表格文字字型大小
10	設定項目編號	21	設定表格列高
11	設定項目符號	22	設定分欄

第三節 試題難度

為說明本實驗所採用之實作試題難度計算原理，本節分為測驗原理、古典測驗難度、及實作試題難度值三部分說明於后。

一、測驗理論

測驗理論(test theory)是一種解釋測驗資料間實證關係(empirical relationships)的有系統的理論學說，它的發展，迄今已邁入新紀元，測驗理論學者通常把它劃分為古典測驗理論(classical test theory)，主要是以真實分數模式(true score model) (Gullikson, 1987; Lord & Novick, 1968)為骨幹；另一為現代測驗理論(modern test theory)，主要是以試題反應理論(item response theory) (Hambleton & Swaminathan, 1985; Hambleton, Swaminathan, & Rogers, 1991; Lord, 1982)為架構。此兩種理論目前並行流通於測驗學界，但試題反應理論卻有後來居上，逐漸凌駕古典測驗理論之上，甚至進而取而代之之勢。

古典測驗理論的內涵，主要是以真實分數模式為理論架構，依據弱勢假設(weak assumption)而來；其理論模式的發展為時甚久，且已發展得相當具有規模，所採用的計算公式簡單明瞭、淺顯易懂，適用於大多數的教育與心理測驗資料，以及社會科學資料的分析，為目前測驗學界使用與流通最廣的理論依據。

然而，除上述各項優點外，古典測驗理論卻有下列諸項先天的限制(Guion & Ironson, 1983; Wright, 1979)：

1. 所採用的係數，諸如難度、鑑別度、和信度(reliability)等，皆為樣本依賴(sample dependent)。

2. 以一個相同的測量標準誤(standard error of measurement)，作為每位受試者的測量誤差指標，沒有考慮受試者能力的個別差異，對高、低能力兩極端組的受試者而言，這種指標不合理且不準確，致使理論假設的適當性受到懷疑。
3. 對於功能相同的非平行(nonparallel)測驗所測得的分數，無法提供有意義的比較，有意義的比較僅侷限於相同測驗的前後測分數或複本測驗的分數。
4. 對信度的假設，是建立在複本(parallel forms)測量的概念假設上，但是這種假設往往不存在於實際測驗情境裡。
5. 忽視受試者的試題反應組型(item response pattern)，認為原始得分相同的受試者，其能力必定一樣，其實不然。

一般說來，為了克服古典測驗理論的缺失，現代測驗理論應運而生。現代測驗理論的內涵，主要是以試題反應理論為理論架構，依據強勢假設(strong assumptions)而來，理論模式正不斷的在發展當中，所採用的計算公式複雜深奧、艱澀難懂，為一立論與假設均合理與嚴謹的學說，所適用的測驗資料種類雖屬有限，但深受測驗學者的青睞，已有逐漸凌駕古典測驗理論之上，甚至進而取而代之之勢。

現代測驗理論具有下列幾項特點，這些特點正是古典測驗理論所缺(Hambleton & Cook, 1977; Hambleton & Swaminathan, 1985; Hambleton, Swaminathan, & Rogers, 1991; Lord, 1982)：

1. 所採用的試題參數(item parameters)，不受樣本影響(sample-free)。
2. 能夠針對每位受試者，提供個別差異的測量誤差指標，而非單一相同的測量標準誤，因此能夠精確推估受試者的能力估計值。

3. 可經由適用的同質性試題組成的分測驗，測量估計出受試者個人的能力，不受測驗的影響(test-free)，並且對於不同受試者間的分數，亦可進行有意義的比較。
4. 以試題訊息量(item information)及測驗訊息量(test information)的概念，來作為評定某個試題或整份測驗的測量準確性，似有取代古典測驗理論的「信度」，作為評定試卷內部一致性指標之勢。
5. 同時考慮受試者的反應組型與試題參數等特性，因此在估計個人能力時，除了能夠提供一個較精確的估計值外，對於原始得分相同的受試者，也往往給予不同的能力估計值。
6. 所採用的適合度考驗值(statistic of goodness-of-fit)，可以提供考驗模式與資料間之適合度、受試者反應是否為非尋常(unusual)等參考指標。

綜合上述，現代測驗理論似乎是優於古典測驗理論，但是事實上，現代測驗理論被採用於解決真實測驗資料者，比起古典測驗理論廣泛地被應用的情形而言，為數較少，其主要原因有下列諸點：

1. 現代測驗理論係建立在理論假設嚴謹的數理統計學機率模式上，是一種複雜深奧、艱澀難懂的測驗理論，這對於在數學方面訓練較有限者而言，無非是一大挑戰。
2. 礙於嚴苛的基本假設，現代測驗理論所能適用的測驗資料較有限，並且需要大樣本的配合，因此使得它的應用性大打折扣，未獲一般測驗使用者的全力擁護。

由上述兩種測驗理論的比較可知，古典測驗理論雖然較欠嚴謹，但理論淺顯易懂，便於在實際測驗情境實施；現代測驗理論雖然嚴謹，但理論艱深難

懂，僅適用於大樣本測驗資料的分析。所以，此兩種測驗理論各有所長，在應用上也各有其限制。

二、古典難度係數

難度(difficulty)是針對試題統計特徵的量化分析(quantitative analysis)之一。在常模參照測驗中，測驗的目的是為了區分受試者學習成就的高低，而測驗試題的難度取決於測驗的目的。因此，為了發揮測驗的鑑別功能，在編製試題時，教師多選擇難度適中的試題，即難度係數接近 0.500 的試題，此種試題最能夠區別學生的能力高低，其鑑別度指數最高。

試題的難度係數有兩種表示方式：答對百分比法(number correct ratio)及范氏試題分析表法(Fan's item analysis table)。其中，答對百分比法既簡單又實用，本研究採用此難度係數表示方式。

答對百分比法是計算全體學生中答對該題人數(R_i)佔總人數(N)的百分比值作為該試題的難度係數(P_i)，計算公式如式(2.1)：

$$P_i = \frac{R_i}{N} \quad (2.1)$$

通常研究者會取兩個極端族群的答對百分比平均值做為該題的 P_i 值。此種方法所計算出來的 P_i 值無法顯示試題的區別功能。因為無論答對的人是高分組還是低分組，只要答對百分比相同的題目，被視為難度相同的試題。因此，做難度係數分析方法有下列四個步驟：

1. 將學生的反應資料評分後，依原始總分降冪排序。
2. 取前 1/4 到 1/3 的學生作為高分組；取後 1/4 到 1/3 的學生作為低分組。將學生區分為高、中、低分三組。
3. 分別計算高分組及低分組受試者在每一題答對人數及其百分比。

4. 計算每一題高分組答對百分比和低分組答對百分比的平均，即為該試題之難度值(P_i)。計算公式如式(2.2)、(2.3)與(2.4)：

$$P_{iH} = \frac{R_{iH}}{N_{iH}} \quad (2.2)$$

$$P_{iL} = \frac{R_{iL}}{N_{iL}} \quad (2.3)$$

$$P_i = \frac{P_{iH} + P_{iL}}{2} \quad (2.4)$$

上三式中，L 為低分組，H 為高分組； R_{iL} 低分組在第 i 題的答對人數、 R_{iH} 高分組在第 i 題的答對人數； N_{iL} 低分組總人數、 N_{iH} 高分組總人數、 P_{iL} 低分組在第 i 題的難度係數、 P_{iH} 高分組在第 i 題的難度係數。

三、實作試題的難度

本研究所採用的實作評量是以評分標準(即前節所述之「檢核表」)作為計分方法。評分標準中列出所有的評分項目。每個評分項目被視為分數權重相等之測驗元素，以受試者表現正確的測驗元素佔所有測驗元素之百分比，作為受試者得分。實作試題原始得分計算公式如式(2.5)：

$$S = \frac{X}{N} \times 100 \quad (2.5)$$

其中， S 為實作試題原始得分； X 為受試者表現正確的測驗元素數量； N 為測驗元素的總數。

實作評量的難度計算，是以評分標準上所有項測驗元素的難度值平均而得。測驗元素的難度值計算，是參考古典測驗理論的難度計算方式。以實作試題的原始總分，取高分及低分兩組。再以各組受試者在某一測驗元素的答對人

數，代入古典難度計算公式，計算該測驗元素之難度值。最後，依據評分標準上每一項測驗元素的難度值平均作為該實作題之難度值。