

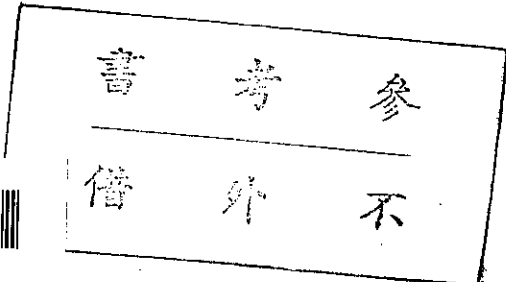
TR-88-006

標準鍵盤中文輸入法之評估

中研院資訊所圖書室



3 0330 03 00081 9



0081

TR-88-006

標準鍵盤中文輸入法之評估

中央研究院資訊研究所

鄭國揚

參與人員：曾志明

中華民國七十七年六月

目 錄

<誌	謝 >	i
<摘	要 >	ii
第	一 章	簡 介 ..	1-1
第	二 章	介紹整個評估系統的架構	2-1
第	三 章	評估之服務項目的由來與計算	3-1
	第 一 節	輸入法學習階段	3-1
	第 二 節	鍵盤學習階段	3-7
	第 三 節	記憶強度與輸入法靜態服務項目分析 ..	3-8
第	四 章	S-P 圖表 (CHART) 的介紹，製作與功能	4-1
	第 一 節	何謂 S - P 圖表 (CHART)	4-2
	第 二 節	S - P 表的作表步驟	4-4
	第 三 節	S - P 圖表的基本類型	4-7
	第 四 節	S - P 表的判斷要領	4-9
	第 五 節	用以個別的學習診斷而判斷	4-13
第	五 章	透過 S-P CHART 的差異係數與迷糊集 (FUZZY SET) $F(U)$ 的動態分析	5-1
	第 一 節	從學習函數 $f(u)$ 與實際正確值的差距來了解學習輸入法的困難之處	5-2

第 二 節	從學習函數 $f(u)$ 與實際正確值的差距 和鍵盤熟悉來了解鍵盤排列的困難度..5-6
第 三 節	由吉尼 (GINI) 係數來判斷分佈狀況 .5-11
第 四 節	動態變法分析說明.....5-13
第 六 章	結論與建議.....6-1
< 參 考 文 獻 >R-1

摘 要

中文輸入法的研究是發展『中文電腦化』一個重要的課題。對一個輸入法而言，學習時間與輸入速度正確且快是為最佳的方法。然而今日在以鍵盤輸入法為主，而種類相當多的狀況下，卻還無一種輸入速度既快，又不需太多時間學習，由於各種輸入法各自擁有不同的特性。因此對一個使用者而言常會無所適從地不知如何來選擇一個合適的中文輸入法而困擾。因而評估需求應運而生，而評估的重要性在於 1. 提供選擇輸入法的依據 2. 指出有待改進之處 3. 加速電腦中文化的步伐。

在過去，雖然亦有人作此方面的評估，但一般的評估方式大都祇是作『量的評估』而沒作『質的分析與診斷』。所以在我們的評估法中，不但作量的評估外，亦作質的分析與診斷。但對量的評估中，我們只求出評估的服務項目作為參考用。而不再像過去對各個服務項目給予一個加權數的算法，而把加權數留給專家去計算。在質的分析與診斷方面我們利用 S-P 圖表 (CHART) 的技術分析，來彌補此方面的不足。另外，我們針對一般評量方式所採用二分法方式 (即不是答對就是為錯的評量法) 的缺失。提出以迷糊集 (Fuzzy Set) 為概念基礎之學習函數 $F(U)$ 的架構作法與公式，並可同時輔助 S-P 圖表中差異係數解釋的不足，作進一步的分析與探討。

第一章 簡介

一般將中文字的造字方法分爲六種：象形、指事、會意、形聲、轉注、假借、即所爲的“六書”。早期的中文字以象形文字爲主，所表達的僅是簡單的意念。後來由於應用上的需要，中文文字逐漸演進、擴充。由簡單的文字擴充到複雜的文字。在整個過程中，中文字朝美觀、易於學習及了解的方向發展。

然而正因爲中文字爲表意文字，屬於字形語系，而不是拼音文字，無法以少數的字母來正確的組合文字，可以說與電腦鍵盤所善於處理的文字特性正好相反，以致利用電腦直接處理中文資料的理想一直遭遇到實際上的困難，爲了瞭解及克服這些困難，必須先對中文與一般拼音語文字（如英，美，俄，德等）之結構特性有所了解，茲以中文與英文不同點作一比較如表 1-1。

由表 1-1 可知，英文的文字數量雖多，但均可由 26 個英文字母按自左而右的順序組合而成，電腦只要能處理該 26 個字母及一些配合符號，即可依簡單的邏輯來組合所有的字，而中文卻是一字一形，而且其字形與數量仍隨社會之變化而變動，如何利用電腦之特性，以有效而經濟的來儲存及處理中文資料，確是頗費周章。

表 1-1 中文與英文之比較

比較項目	英 文	中 文
語 系	拼音語系	字形語系
文字結構	由 26 個字母組成，屬於一度空間的線性結構	沒有字母，但有部首及偏旁，屬於二度空間的平面結構
筆 順	固定由左到右	為方塊字，由上到下，由左到右，由中間向左右等不定
字 數	一百多萬字 (韋氏字典)	49,905 字 (張其昀編 "中華大辭典")
文 字 衍生性	造字要受母音限制，不能隨便造字。	只要懂得某物的屬性，加上偏旁，便可造字，易於瞭解其 "意" 與 "聲"。

所以，應用電腦鍵盤輸入上不似英文字的輸入具有唯一的字碼也無法如英文能快速而有效的輸入文字，因此在中文電腦資料處理中文輸入法是相當基本而重要的一環，故中文電腦方面的輸入問題一直是很多專家學者研究的重要課題之一，根據資訊策進會于民國72年2月所作的“中文電腦市場調查分析報告”指出，當時在台灣已商品化的中文輸入法的與其相關鍵盤已從1981年的24種增加到1982年的31種，增加幅度達28%，如圖1-1。[1]。

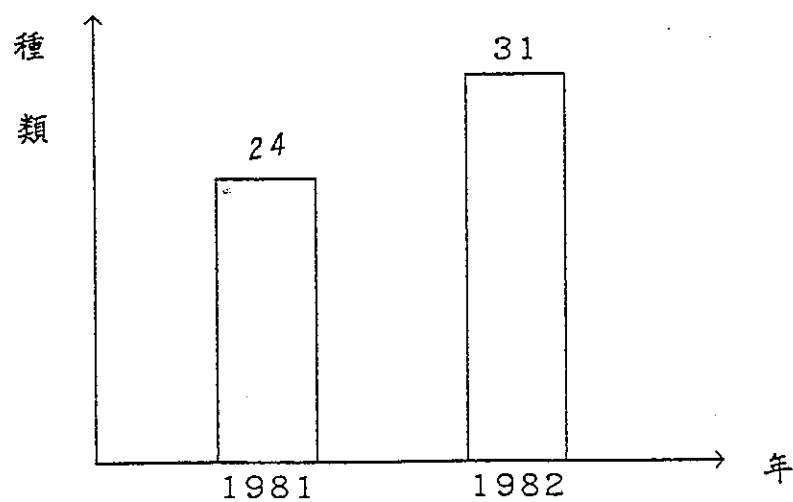


圖 1-1.

從以上已知發展的中文資料輸入方法有很多種，而目前輸入部分，傳統的鍵盤式中文資料輸入可說已相當成熟，如依輸入鍵盤之大小及種類區分，可分為大鍵盤，字根鍵盤，注音鍵盤及一般標準英文鍵盤等，製作成本以大鍵盤最高，並依次減低。如依輸入方式來區分，目前鍵盤式中文輸入法研究的主流，大致可分為三大類[2]，優劣點分別述說如下：

<-> 形碼輸入法

形碼輸入法就是將中文文字依構造分析歸納為若干個字根，根據特定規則，以組合中文字的方法。所以其優點，在於即使不認識的字，根據字的構造判斷，就可以得知輸入碼，因此，能輸入的字集就沒有限制。但是它也有缺點，就是需要投入相當多的時間去記憶字根在輸入器的及拆字組字規則，同時，輸入時需要較多的思考判斷時間來尋找所要輸入字根的位置，如此很難達到類似英打所謂的觸摸打字『TOUCH TYPING』的程度，所以輸入速度會受到限制。

<二> 音碼輸入法

音碼輸入法以注音符號為輸入碼，該法最大的優點是相當容易學習，這與我們多年來接受中文教育有關，而它的缺點是不會唸的字就不會輸入，欠缺彈性；此外，因同音字的問題，在輸入注音符號後仍需目視螢幕檢字，因此輸入速度並不是很理想。

<三> 混合輸入法

混合輸入法是將形碼與音碼混合使用，嚴格來說，該法介於形碼輸入法與音碼輸入法之間，雖然學習上較音碼輸入法為困難，但輸入速度卻較純粹的音碼輸入法為快，不過仍無法達到很理想的水準。

由上述的輸入法種類與特性，可了解不同的輸入法難免性能優劣不齊，造成一般用戶選擇中文輸入法的困擾。因此蘊育起對輸入法的評估需求，下列我們來看評估之重要性：

1. 提供選擇輸入法的依據，2. 指出有待改進的地方，3. 加速電腦中文化的步伐。所以，便有人提出一種測量方法模式。

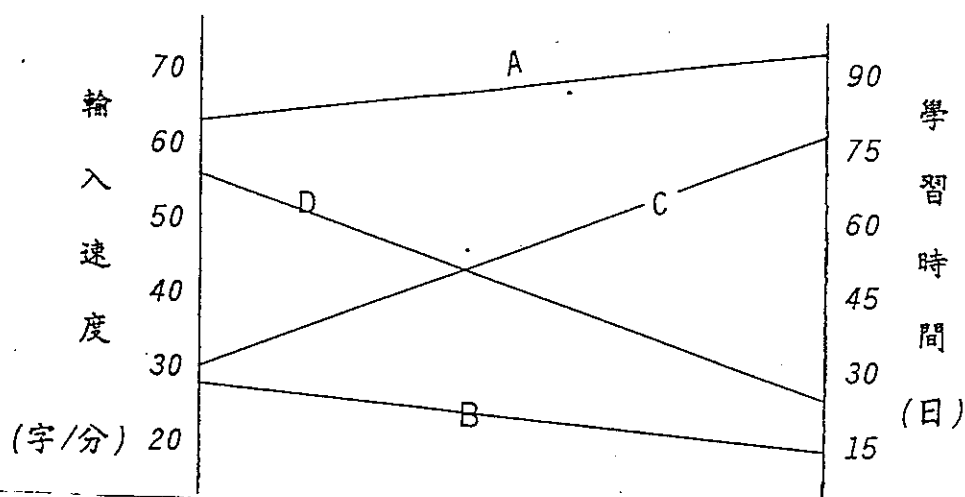


圖 1-2.

- 輸入法 A：由於輸入速度快，但相對學習時間長，所以較適合專業操作者。
- 輸入法 B：在學習時間較低，但輸入速度慢，所以較適合一般的操作者。
- 輸入法 C：無論在學習時間或速度上均不適合專業或一般操作者。
- 輸入法 D：不論在學習時間或速度上均適合專業或一般操作者，都為最佳的系統。

然而以目前研究的方向來看，沒有一種鍵盤式中文輸入法可輸入速度既快，又不需太多時間學習，亦即，沒有任何一種輸入法既可適合專業輸入人員使用，又適合一般使用者使用，這是鍵盤式中文輸入法始終無法有很大突破的原因，不過，在具親和力的中文輸入方式（如語音輸入，光學印刷字型掃描輸入及手寫字型輸入等）尚未正式問世及普遍應用以前，鍵盤式中文輸入法將仍是中文輸入的主要方式。

所以，針對輸入法的評估，幾年前即有人在作，祇是相當少，以下我們將略述其方法與優缺點：

1. 1984年CHEN和GONG [3] 所提出的評估法中，他們的評估方式為經由電腦公司實際訓練五位學生，每天施予3小時的學習輸入法，連續五天，再練習敲鍵輸入達2個星期，並每天作測試並作成記錄，其主要測試項目如下：

- (1) *average input speed*
- (2) *input error rate*
- (3) *conflict code rate*
- (4) *size of the character set*
- (5) *average no. of strokes for per character*
- (6) *ratio of rejected character*
- (7) *average learning time*

等七個評估因素，然後再對不同的評估因子給於不同的加權數 (WEIGHT)，如此評估因子乘上加權數，再加總，然後以分數高低作為選擇的好壞。其優點在於將所有一般評估法的評估因子 (服務項目) 均予考慮，相當詳細，然而其缺點在於測試人員過少，有失公平性。受測者的學習心態、狀況均難以了解，而且對於加權數 (WEIGHT) 全憑經驗而來，不免令人懷疑。

2. 1985 年 S.Y.LO [4] 提出 "使用科學模式來評估中文輸入法"，其純粹以數學理論在 (1) 鍵盤大小 (2) 系統符號大小 (3) 輸入時每階段的步驟分析總數 (4) 輸入速度 (5) 找尋資料的方便性等五方面對中文輸入法作一評估。其優點在於能對於人拆解成字碼的動作步驟。純粹以數學理論分析，適用於過去、現在、或未來的輸入法作評估，但其缺點正因為其純粹以數學理論作評估，忽略人為因素，不免流於理論而不實際。

3. 1987 年資訊工業策進會，亦再以類似 1984 年 CHEN 與 GONG 的作法，以實際測驗一群程度近似的學生。但著重的評估因子 (服務項目) 卻只有學習時間、輸入速度、與記憶強度 (停止練習一段時間後，所還能記憶的程度)，除此之外，他們還作了偏好感覺的問卷調查。其優點在於除了考慮實際測驗外，亦作人在學習過程中所產生對此種輸入法的喜好程度的調查。然而其缺點在於加權數 (WEIGHT) 的計算方式是由問卷調查而來，所以加權數的大小，不免令人質疑。

在上述的評估方法中，我們認為應除了一般的考慮項目(如學習時間，正確字數.....等)外，對於人的感覺的細部分析亦相當重要，所以我們亦提出一個評估方式。基本上，我們祇將針對在標準鍵盤輸入的工作環境 (ENVIRONMENT)，將整個評估分為三個階段，分別為學習中文輸入法階段、鍵盤輸入速度測定階段與對輸入法記憶強度階段，在每一個階段均予以不同時間的多次學習與測驗，而這些測驗試題，我們的以隨機的方式產生，同時不同階段可得到一些測驗的服務項目值。另外，我們在服務項目內亦針對答題的結果提出一個學習函數 $F(U)$ 的迷糊集 (FUZZY SET) 以作為評估對此輸入法的切確了解程度。於是我們在不同的輸入法得到不同的服務項目值，然而我們將不似過去評估方式，將不同的服務項目(評估因子)給予乘上不同的加權數，再加總來考量。我們的作法祇提出服務項目值，而其加權數的設定由於考慮到人性的喜好程度與觀點不同，故我們將把此問題留給專家去討論。而且一般學習評鑑以『百分比』或『等第法』來表示學習成就，至於學習困難的發現與診斷卻始終未能獲的突破性的技術。於是我們除了『量的評估』外，我們亦希望能對其『質的分析與診斷』。於是我們將在不同階段，不同時間得到的測驗結果值以 $S-P$ 圖表 ($S-P$ CHART) 的統計方式對每次測驗均有一個學習評鑑值差異係數 D^* 。以從試題答案中去明瞭學生對此輸入法所呈現的學習情況與是否有瓶頸所

在，並且我們試著以所提的 $F(U)$ 功能去解釋何種原因的存
在。於是將這些圖形與訊息傳達給輸入法設計者可以改良其
輸入法外，亦可提供一般使用者更多比較參考與選擇的訊息
。現在我們把本篇論文的内容簡述如下：

第二章 主要介紹評估系統的架構。

第三章 服務項目的計算。

第四章 $S-P$ CHART 的介紹，製作與功能。

第五章 透過 $S-P$ CHART 的差異係數與迷糊集 (FUZZY
SET) $F(U)$ 的動態分析。

第六章 結論與建議。

第二章 介紹整個評估系統的架構

以下我們對整個評估架構運作作一說明

大體上，我們的系統除具有一般的評估能力（亦即除作一般評量）外，我們亦將針對測驗結果的狀況作分析（即質的分析），利用量的評定與質的探討以供輸入法設計者可再度作改良輸入法或鍵盤排列的不適宜之處。

以下我們就整個處理方式說明，整個評估過程分為三個階段，分列于下：如圖 2-1

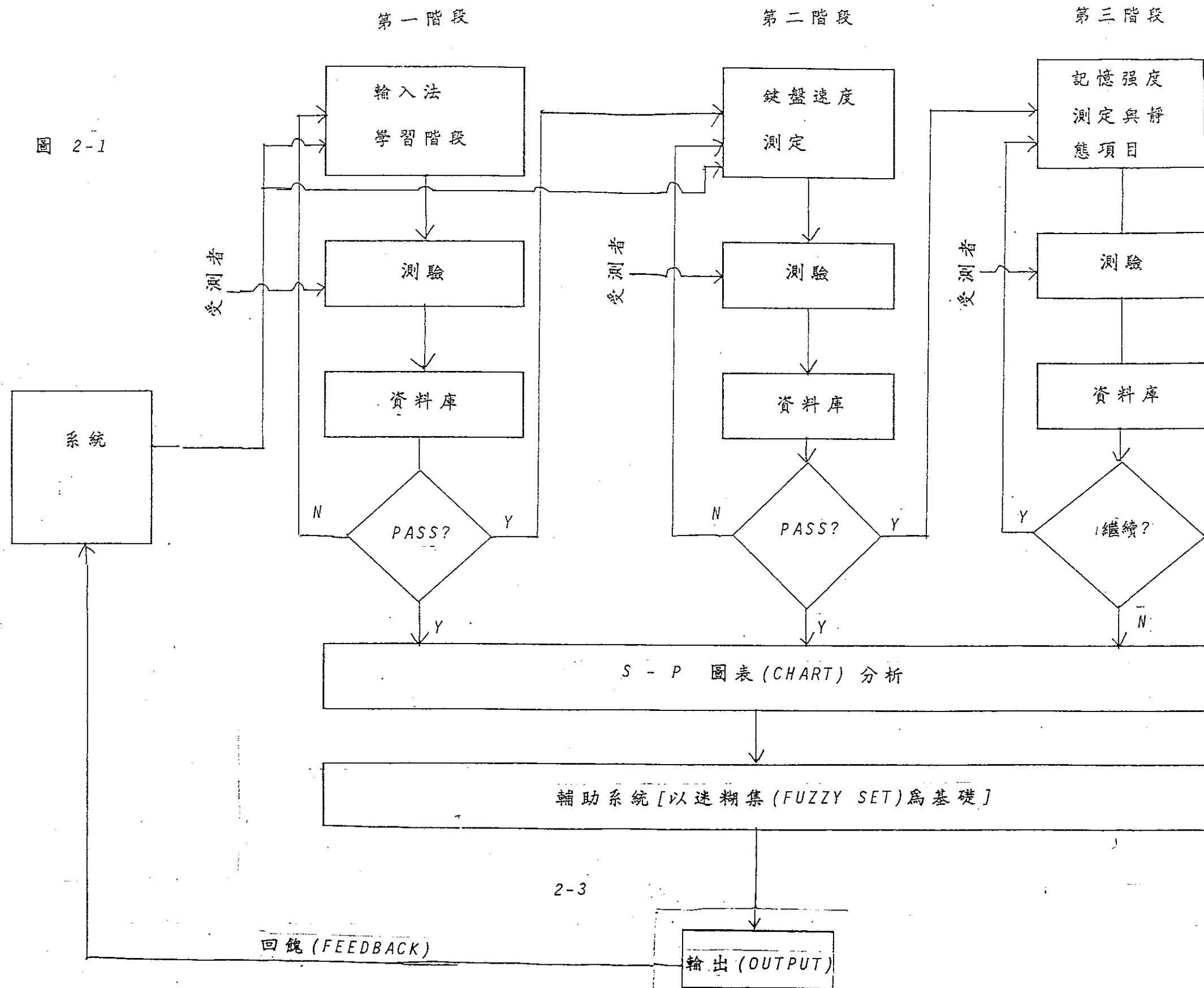
第一階段為學習中文輸入法階段，主要為學習基本字碼有那些，與如何將中文字析解為基本字碼的規則學習，並驗證其學習狀況，在我們的測試資料的字集完全以通用漢字標準碼，並以隨機方式出題目。而在測試期間內，受測者必須將測試資料全部輸入完畢，並以通過一定程度的標準，方可進入下階段。至此階段結束前，我們將每次測試結果，以建立一個資料庫 (data base) 儲存我們所需要的試題結果值與及服務項目內容 (service items)，如學習時間、平均正確字數率、學習函數 $F(U)$ 。當此階段結束後，我們進入下一階段。

第二階段為透過電腦標準鍵盤來作輸入基本字碼的速度學習與測定，其測試系統狀況一如第一階段，同樣地我們將測驗結果值與一些服務項目如輸入速度率，鍵盤熟析度等儲存於資料庫內，當此階段結束後，我們再進行下一階段。

第三階段為對受測者停止練習一段時間後，對於原輸入法輸入速度所還能記憶的程度如何，所以我們同時進行第一階段步驟與第二階段步驟，並將所有的結果與服務項目分別儲存於資料庫內。

最後，當所有的測驗結束後，我們將透過 *S-P CHART* 的差異係數作學習評鑑，並透過我們設定的迷糊集 (*FUZZY SET*) 公式來作輔助的動態分析說明以供一般使用者選擇與輸入法設計者重新思考改良其原先在系統設計內容。

圖 2-1



第三章 評估之服務項目及計算

我們經由測試而得到一些服務項目 (SERVICE ITEMS)，在本章中我們將討論如何來計算這些項目的值。在下列各節，我們將針對各階段的執行分別加以討論，其中包括輸入法學習階段、鍵盤學習輸入階段、與記憶強度及靜態服務項目分析階段。

第一節 輸入法學習階段

由於對不同的輸入法，教學者會有不同教學方式、教學時間，而且必須達到某一特定之水準 k (例如必須達到 80 分以上。)，而 k 由專家們來定。

我們採取下列流程：

1. 每次學習 T_j 時間

2. 測驗 P_{ij} 時間

(在第 i 次測驗中，受測者 j 每次完成試題，所需的測驗時間)。

3. 檢定結果以判知是否達到所要求的程度，否則重複 1. 的步驟。

所以當受測者達到要求程度時，我們便可計算得到下列服務項目。

1. 學習時間 = $\sum_{j=1}^m T_j$ 代表學習者經過 m 次學習後由不會到了解到某種程度時所花費的總時間。

2. 學習函數 $F(U)$

在輸入法的析解中文與字碼對應，我們認為在思考上有類似樹狀傾向，因此對於絕大多數的中文輸入法，其字碼形成方式均可看成一個樹狀結構 (tree structure)。首先我們來看中文字與對應字碼析解圖 (Character Decomposition Graph)。(CDG)

[定義]

$$CGD = (V, E)$$

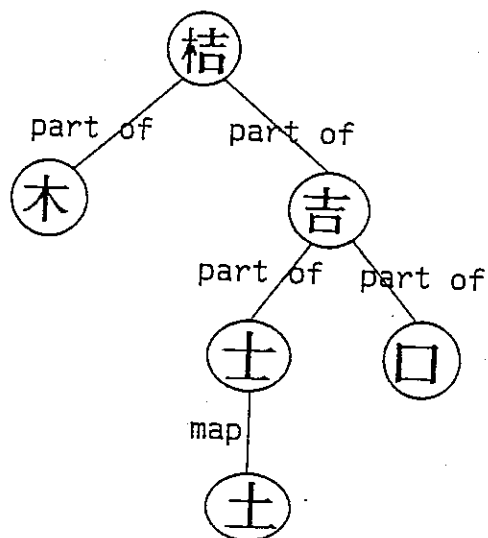
$$V = TN \cup NTN$$

V 為節點 (NODES)，代表中文字形

包括終節點字集 (TERMINAL NODE SET) TN 與非終節點 (NON-TERMINAL NODE SET) NTN 。

E 為連線 (EDGE)，代表節點之間的階層關係 (HIERARCHICAL RELATION)。

如圖 3-1，『桔』的析解過程如下



$$NTN = \underline{\text{桔, 吉, 士}}$$

$$TN = \underline{\text{木, 土, 口}}$$

$$E = \underline{\text{PART-OF, MAP}}$$

圖 3-1

其最後的 *TERMINAL NODES* 為我們所欲答的字碼，而其餘的 *NON-TERMINAL NODES* 為我們思考過程中所呈現腦海的影像符號，然而在作答上，一般的評定方式是認為所有的 *TERMINAL NODES* 均須答對才算是正確，否則只要其中有一字碼答錯，即使得得分為零，這即是心理學上二分法。然而近來的心理測驗則採用所謂的多分法，即對某事務有某種程度的認識，如為 0, 0.25, 0.5, 0.75, ... 等等。如例『老』字可以去定義所有人類的年齡，有的人認為超過七十或是大於七十則是 1.0，因此也就不明顯的表示。六十歲算不算老呢？也許我們可以說這種個別的是屬於部分的老，對於老的程度評估為 0.7，在這種情況，將『老』一詞的含糊成分以一種演算法則的式樣來表示。同樣地，對一個輸入法的字碼了解性上，受測者亦不應以二分法來看待，而應有某一程度的了解，而我們此種了解程度即一般的迷糊集 (*Fuzzy set*) [5]。以下便可定義迷糊集 (*Fuzzy set*) 和中文輸入法中中文字與字碼間的階層結構 (*Hierarchical Structure*)。

[定義]

令 W 為中文字與對應字碼析解圖 (*CDG*) 中，所有的字碼 [*(TERMINAL NODES)* 和 *(NON-TERMINAL NODES)*] 組合的集合，即 $W = (TERMINAL NODES, NON-TERMINAL NODES)$ 。

令 U 為 W 的冪集合，即 $U = 2^W$ 。假設 F 為能從一個中文字了解到某輸入法的字碼集合，則 F 包含於 U 中， F 是 U 的一個迷糊集，即

$$F(U) = \frac{\sum_{i=1}^n f(U_i)}{U_i}$$

其中， $f(U_i)$ 是 U_i 在 f 中的成員關係函數值， $U_i \in U$ 即 $f : U \rightarrow [0, 1]$

所以我們便產生了對一個中文字碼的了解程度，其值介於 0 與 1 之間，我們的作法為任一個中文字碼圖。

其架構型式如下圖 3-2

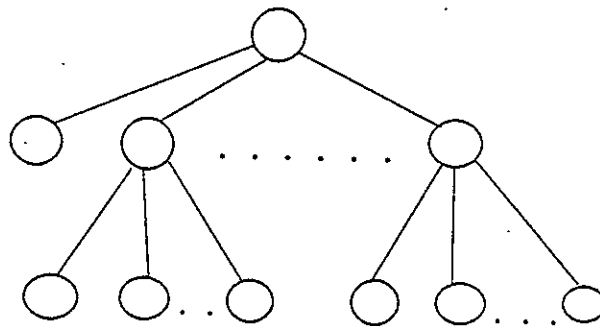


圖 3-2

由於 *TERMINAL NODES* 為其答對或答錯的結果，即不是 0 就是 1。但任一 *TERMINAL NODES* 均由上一層之 *NON-TERMINAL NODES* 所拆解或得到的，因此對一個 *NON-TERMINAL NODE* 而言，其學習函數值均由下一層的 *TERMINAL NODES* 或 *NON-TERMINAL NODES* 所答對或答錯的狀況來決定其認識此字形的程度。所以我們計算時，實際上除 *TERMINAL NODES* 答對者為 1，答錯者為零外，其餘的 *NON-TERMINAL NODES* 為下一層的 *TERMINAL NODES* 或 *NON-TERMINAL NODES* 總和的平均值。所以我們定義一個中文字對應字碼的熟悉度為。

0 OR 1

IF U_i IS Terminal Nodes

(基本字碼所在之節點)

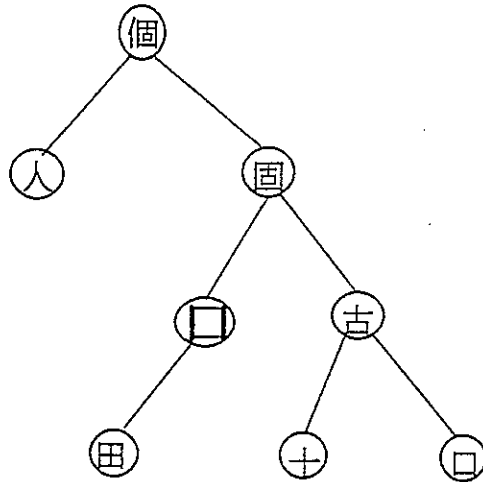
$F(U_i)_1 =$

$$\sum_{j=1}^M F(U_{i+1})_j / M \quad \text{IF } U_i \text{ IS NON-Terminal Nodes}$$

(非基本字碼所在之節點)

即除 *TERMINAL NODES* 的值不是 0 就是 1 外，如果某一個 *NON-TERMINAL NODES* 的下一階層擁有 M 個 *NODES*，則其值為 M 個 *NODES* 值的加總，再求平均值，如此遞迴 (*recursive*) 一層一層地由最底層回到原中文字，而產生對此中文字的了解程度。

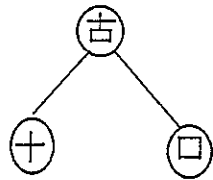
例如：



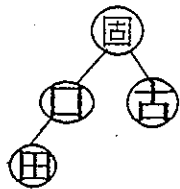
如果 Terminal Nodes 答對或答錯的狀況如下：

人 : 0
 田 : 1
 十 : 0
 口 : 1

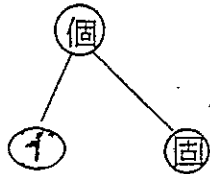
則



$$\text{古} = (\text{十} + \text{口}) / 2 = \frac{0+1}{2} = \frac{1}{2}$$



$$\text{固} = (\text{田} + \text{古}) / 2 = \frac{1 + \frac{1}{2}}{2} = \frac{3}{4}$$



$$\text{個} = (\text{人} + \text{固}) / 2 = \frac{\frac{3}{4} + 0}{2} = \frac{3}{8}$$

所以根據其答題的狀況，可知其對『個』字的了解程度為 3/8。

從上述迷糊集的定義，我們知道此種學習函數 $F(U)$ 表示對一個中文字與對應碼的了解程度，當 $F(U) = 1$ 表示其能完全被了解； $0 < F(U) < 1$ 表示無法立即了解其對應，必須經過不同程度的學習過程才能完全了解， $F(U)$ 值愈小，能夠了解的程度愈低；當 $F(U) = 0$ 表示無法被了解。

3. 在第 i 次測驗，平均每分鐘正確字數率 = NL_j

我們將每次測驗的結果以對或錯的方式來得到受測者在本次測驗時所答對的正確字數，而事實其計算方式即為對學習函數中每一個字中，除非 $F(U)$ 為 1，否則即為 0。

n_{ij} : 在第 i 次測驗時，第 j 個人所答對字數

P_{ij} : 在第 i 次測驗時，第 j 個人所花費的測驗時間

$$NL_j = \frac{\sum_{j=1}^m (n_{ij} / P_{ij})}{m}$$

m 個人

m

第二節 鍵盤學習階段

第二階段我們採取

1. 測驗 P_{ij} 時間

2. 檢定結果以判知是否達到所要求得程度，否則重複 1.

所以當受測者達到要求程度時，我們可得下列服務項目

。

$$\text{平均敲鍵速度} = NT_i$$

為平均每次受測者於 P_{ij} 時間內所敲打正確的字數
 nt_{ij} : 在第 i 次測驗時, 第 j 個人所打對字數

P_{ij} : 在第 i 次測驗時, 第 j 個人所花費的測驗時間

$$NT_i = \frac{\sum_{j=1}^m (nt_{ij} / P_{ij})}{m}$$

m 個人

m

第三節 記憶強度與輸入法靜態服務項目分析

< 一 >. 第三階段 記憶強度

目的: 在於探討當使用者停打一段時間的敲打與練習後, 那些輸入法遺忘程度最大, 那些輸入法遺忘程度最小。

作法: 予以測驗, 施以同時作學習輸入階段與鍵盤速度階段時的測驗, 再以其結果作分析。

結果: 得到在第一節的服務項目如: 平均每分鐘正確字數率, 學習函數 $F(U)$ 。得到在第二節的服務項目如: 平均敲鍵速度。

< 二 >. 靜態服務項目分析

1. 字碼數目多寡

由系統設計者所定義的基本字碼數。

$$\sum_{i=1}^m (N_i * F_i)$$

2. 重複率 = -----

$$\sum_{i=1}^m F_i$$

所謂重複率，即當一個中文輸入法系統中，對應同一組字碼的中文字之個數。例如 注音：

ㄇ ㄨ ㄥ ㄛ：{名，明，冥，茗，.....，螟} 共 10 個。

ㄇ ㄣ ㄛ：{人，儿，仁，壬，任} 共 5 個。

而一般重複率愈高，目視螢幕檢字，所需時間愈長。

N_i ：對應同一堆字碼的中文字個數

F_i ：此中文字在一般常用中文字中所佔的使用頻率 [6]。

$$\sum_{i=1}^m (F_i * C_i)$$

3. 平均字碼數 = -----

$$\sum_{i=1}^m F_i$$

代表平均每個中文字要鍵入 (key in) 的字碼數。

F_i ：此中文字在一般常用中文字中所佔的使用頻率。

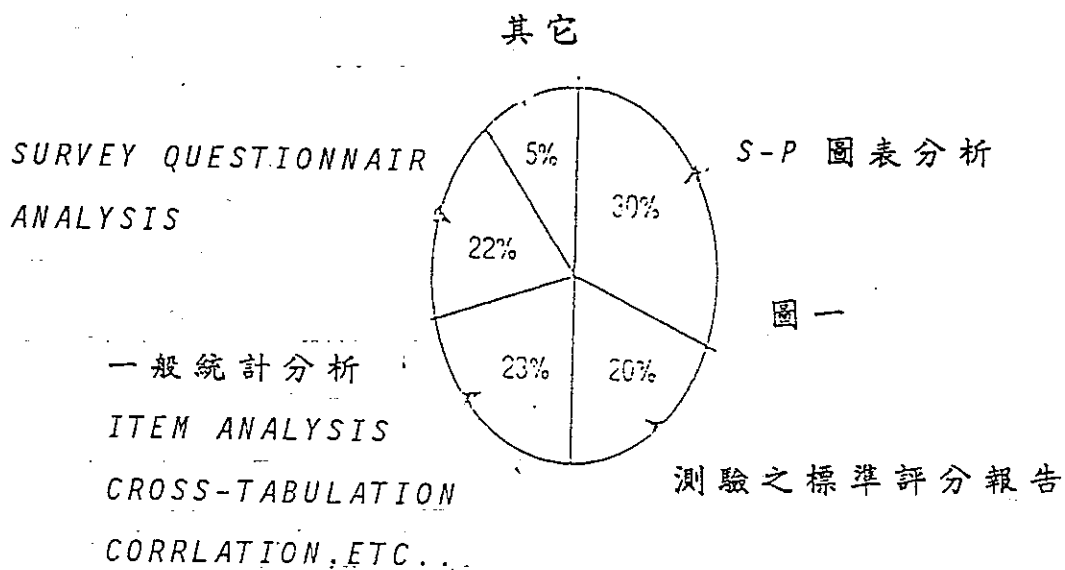
C_i ：此中文字所對應的字碼數。

在我們求出上述的服務項目後，這即為我們所謂的『量的評估』，提出作為外界的參考，然而我們對於服務項目間加權數的設定將留給對此方面有所精研的專家們來解決。

第四章：S-P 圖表分析

柯柏勒(R.J.Ribler, 1974)於其著作 "教學目標與評鑑 (Objectives for Qnotruction and Evalution)", "一書中指出：教學的基本過程可分為教學目標、學前評鑑、教學活動、評鑑等四個過程。同時強調此四個過程不僅互有關係，而且其中評鑑更應具有回餽作用的積極功能。[7] 由於一般學習評鑑卻依然沿用傳統方法，以『百分比』或『等第法』表示學生的學習成就之地位。至於學習困難的發現與診斷卻始終未能獲得突破性的技術。其實評鑑並不是教學的終點，其主要目的，在於檢討與分析教學的缺失，診斷學生學習的困難所在，以便進行適當的補救或個別輔導的依據。現行的學習評鑑方式主要有S-P 圖表分析(S-P CHART ANALYSIS)、一般統計分析(GENERAL STATISTICAL ANALYSIS)、SURVEY QUESTIONNAIRE 分析以及測驗之標準評分報告(STANDARDIZED TEST SCORING REPORT)等。圖一所示為目前在日本使用各種學習評鑑方式的比例[8][9]。由圖一中可看出 S-P 圖表分析的使用最為廣汎。

因此我們在評估的過程中，除了上節所介紹的基本參數 (Parameters)，因此另外我們將利用 S - P 的評鑑技術 [10]來分析受測驗者的學習狀況以及輸入法的特異情形。



第一節 何謂 S-P 圖表 (CHART)

所謂 S-P 表係指英文 STUDENT 的『S』，和 PROBLEM 的『P』等兩個字母的簡稱，是指學生與試題得分一覽表。然而，此表的排列，乃是根據學生作答的實際情況，予以有系統的重排，以便進一步診斷之用。換言之，將學生對某種考試，如期考，平時考，月考……乃至標準化測驗等，均可作成此表。不過，其排列的順序，並不是依據學生原來的名冊或座位表。其縱座標是依學生得分的高底而排，也就是高得分者排在上端，低得分者排在下端，依序排列。橫座標則依試題的難易順序而排，左邊是答對率高的試題，右邊是答對率低的試題，由易而難依次排列，如此作成縱橫一覽表，即可判斷或診斷學生的學習困難與成就的情形。因為將學生與試題得分，予以重排並圖表化，所以教師一見就能察學習是否有困難，應否加以個別化處理的資料。舉例言之，從圖 4.1.1(a)，4.1.2(a)，圖 4.1.1(b)，4.1.2(b) 所示，可知，圖 4.1.1(a) 與圖 4.1.2(a)，不容易一看就查知兩表有何差異。然而，若將其重排，依成績的高低由上而下，答對率的多寡由左(易)而右(難)，即可明瞭其差異的所在。

至於 S-P 表在學習評鑑上的功用，依據佐滕隆博實驗指出 [11]。

1. S-P 表的處理與分析方法，簡單容易做到。
2. S-P 表的判斷，一見就能了解，並能把握整體結構。
3. S-P 表對學習診斷上『質的分析』，極為方便。

4. S-P 表不僅對學習的診斷，而且對教學或輔導、發問等兩方面的檢討與改進，均有很大的功用。
5. S-P 表不獨可用於測驗結果的處理，項目的分析。同時對練習題的考試，事前測驗 (PRE TEST)，事後測驗 (POST TEST) 等的得分，均可成 S-P 一覽表，予以解釋、研判，所以對學習診斷的功能甚鉅。

		試題				
		1	2	3	4	
學生	a	1	0	0	0	2
	b	0	1	1	1	3
	c	1	0	1	1	3
	d	0	1	1	0	2
		1	2	3	2	

圖 4.1.1 (A)

		試題				
		1	2	3	4	
學生	a'	1	0	1	1	3
	b'	0	0	1	0	1
	c'	1	1	1	1	4
	d'	1	0	1	0	2
		3	1	4	2	

圖 4.1.2 (A)

		試題				
		1	2	3	4	
學生	a	1	1	0	1	3
	b	1	0	1	1	3
	c	1	1	0	0	2
	d	1	0	1	0	2
		1	2	3	2	

圖 4.1.1 (B)

		試題				
		3	1	4	2	
學生	a'	1	1	1	1	4
	b'	1	1	1	0	3
	c'	1	1	0	0	2
	d'	1	0	1	0	1
		3	1	4	2	

圖 4.1.2 (B)

第二節 S-P 表的作表步驟

(一) 作學生的得分原表

1. 事先作學生人數與題數的空格表，學生為縱座標，試題為橫座標。
2. 將試卷予以評閱，答對的試題給 1，答錯或未答的試題給 0。依每位學生實際作答情形逐題填入空表格內。
3. 若是題目較大，而且可以分為若干小題目的數目而設橫座標的格子。
4. 學習診斷上，答錯與未答，其意義應予區別。因之，亦可將未答的，給 B (即表示 BLANK 之意)。另外，對答錯的題目，依其作答內容，分為幾種類型而各以其它符號表示的地方，也是可取的。
5. 如圖 4.2.1

(二) 重新排出各題得分次序

1. 依學生得分 (總分) 高的排上端，低的排在下端，依次排列，如圖 4.2.2。
2. 依學生答對率高的排在左端，低的排在右，而將其圖 4.2.2 重新排列，作成如圖 4.2.3。
3. 從圖 4.2.3 可見左上方『1』居多，右下方『0』較多。
4. 相同的得分及相同的答對率之排列順序，不必拘泥。但是，求出了『注意係數』，則應得分的注意係數較小排在上端，题目的注意係數較小者，可排在左邊，依此類推。

(三) 畫 S 曲線和 P 曲線

1. S 曲線的畫法

- (1) 將 $S-P$ 表中各個學生的答題數(亦指每位學生的得分)，從左邊算起，並畫出區分線，如圖 4.2.5， G 生得分 14，就在 14 題右旁畫區分線，依此類推而畫出每位學生的得分區分線。
- (2) 其次，將各生的區分線加以連線，就可畫得分的分佈曲線。
- (3) 此種曲線往往形成階段狀的線條。然而，全體觀之，卻成爲 S 型彎曲狀。因此，稱之爲 S 曲線。如圖 4.2.5。

2. P 曲線的畫法

- (1) 在 $S-P$ 表中將各個题目的答對人數，加以算並由上而下，畫其曲分線。如 i 題答對者有 26 人，就由上而下算，並畫其曲分線。
 - (2) 然後將各題的答對數作區分線，並加以連接，就成試題答對率的 P 曲線。
 - (3) 此種曲線表示答對人數的分佈情形。因此，稱之爲 P 曲線，如圖 4.2.6。
3. S 曲線與 P 曲線是否相接近，或是相一致，或分離多遠，此種分離情形或程度，對 $S-P$ 表特性的了解，頗具研究的價值。

		試題編號										合計得分
學生編號	5	1	2	4	5	6	7	8	9	10		
1	6	1	1	0	0	0	1	0	1	1	5	
2	9	1	1	1	1	1	0	0	1	6	6	
3	0	1	0	0	0	0	1	0	1	5	3	
4	1	1	1	1	0	1	1	0	1	6	7	
5	1	1	1	1	1	1	0	1	1	9	9	
6	1	1	0	1	1	0	0	0	1	5	5	
7	1	1	1	1	1	1	1	1	1	10	10	
8	1	0	0	0	0	0	1	0	1	2	2	
9	1	1	1	1	1	1	1	1	1	6	6	
10	1	1	0	0	1	1	0	1	1	4	4	
11	1	1	0	0	1	0	0	1	0	4	4	
12	0	0	1	0	0	0	0	0	0	1	1	
13	0	1	1	1	0	1	0	1	0	5	5	
14	1	1	1	1	0	0	1	0	0	5	5	
15	1	0	1	1	0	0	1	0	0	4	4	
		8	12	11	9	6	7	10	4	8	5	

答對人數

圖 4.2.1

由上而下排出得分高低

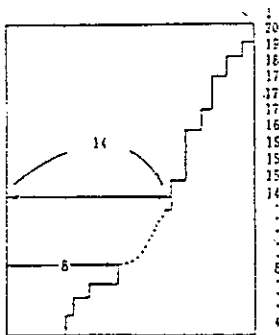


圖 4.2.5 S 曲線 (S-P) 表
左邊算起答對題數處，
加以畫出區分線，然後
將其連接的曲線)

		試題編號										合計得分
學生編號	5	1	2	3	4	5	6	7	8	9	10	
7	1	1	1	1	1	1	1	1	1	1	10	
5	1	1	1	1	1	1	1	0	1	1	9	
2	0	1	1	1	1	1	1	1	1	0	8	
4	1	1	1	1	0	1	1	0	1	0	7	
10	0	1	1	0	0	1	1	0	1	1	6	
2	0	1	1	1	1	1	0	0	1	0	6	
14	1	1	1	1	0	0	1	0	0	0	5	
1	0	1	1	0	0	0	1	0	1	1	5	
13	0	1	1	1	0	1	0	1	0	0	5	
5	1	1	0	1	1	0	0	0	0	1	5	
15	1	0	1	1	0	0	1	0	0	0	4	
11	1	1	0	0	1	0	0	1	0	0	4	
3	0	1	0	0	0	0	1	0	1	0	3	
8	1	0	0	0	0	0	1	0	0	0	2	
12	0	0	1	0	0	0	0	0	0	0	1	
		8	12	11	9	6	7	10	4	8	5	

得分高者排在
上端

(左邊排答對數多的排序

圖 4.2.2

排出)

		試題編號										合計得分
學生編號	5	1	2	3	4	9	1	6	5	10		
7	1	1	1	1	1	1	1	1	1	1	10	
5	1	1	1	1	1	1	1	1	1	0	9	
9	1	1	1	1	1	0	1	1	0	1	8	
4	1	1	1	1	1	1	1	0	0	0	7	
10	1	1	1	0	1	0	1	0	1	0	6	
2	1	1	0	1	1	0	1	1	0	0	6	
14	1	1	1	0	1	0	0	0	0	0	5	
1	1	1	1	0	1	0	0	0	1	0	5	
13	1	1	0	1	0	0	1	0	0	1	5	
6	1	0	0	1	0	1	0	1	1	0	5	
15	0	1	1	1	0	1	0	0	0	0	4	
11	1	0	0	0	0	1	0	1	0	1	4	
3	1	0	1	0	1	0	0	0	0	0	3	
8	0	0	1	0	0	1	0	0	0	0	2	
12	0	1	0	0	0	0	0	0	0	0	1	
		12	11	10	9	8	8	7	6	5	4	

圖 4.2.3

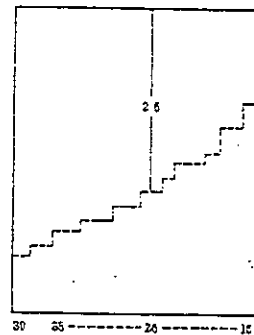


圖 4.2.6 P 曲線 (從 S-P 表
上端算起答對人數處加
以劃出區分線，然後將
其連接的曲線)

第三節 S - P 圖表的基本類型

< 1 >. 標準化測驗的 S - P 圖表如圖 4.3.1

從 S 曲線和 P 曲線所佔的面積，我們可以看出其平均答對率為 50%，而從 P 曲線其試題答對率大致分佈在 20% ~ 80% 之間，且得分率在 50% 的學生最多，愈靠近滿分及零分的學生人數也愈少，並且下降之弧度不快。因此從整個 S - P 圖表看來，無論在試題之困難度或學生學習的分佈情形，都顯得相當調和，此乃標準化測驗之 S - P 圖表。

< 2 >. 達到度測驗的 S - P 圖表如圖 4.3.2

從 S、P 曲線所佔的面積來看，其平均答對率亦是在 50% 左右，且由 S 曲線可以看出學生答對率，由高而低形成向左下成直線的傾斜，對於鑑別到達度的個別差異，可一目了然，極為方便。

< 3 >. 部分學生到達度急降的 S - P 圖表如圖 4.3.3

從 S、P 曲線所佔的面積來看，其平均答對率在 70% ~ 75% 之間，但從 S 曲線的下半部可以看出少數學生得分有急速下降現象，是值得注意的地方。

< 4 >. 高低得分率顯然分離的 S - P 圖表如圖 4.3.4

知其平均答對率在 60% 左右，但從 P 曲線來看，我們可以明顯的看出一奇怪的現象——在試題中，分成高答對率的試題群和低答對率的試題群，而且是明顯的分離，這對試題的編排來說是一個不良的現象。

< 5 >. 編序學習後的 S - P 圖表如圖 4.3.5

平均答對率大約在 80 % 左右，除少數學生的答對率稍低以外，其餘的學生都有相當不錯的答對率，此類型是屬於編序學習 (Programmed Learning) 所常見的 S - P 圖表。

< 6 >. 學前測驗的 S - P 圖表如圖 4.3.6

平均答對率在 25 % 左右，試題的答對率以及學生的達到度都非常的低，這是屬於標準的學前測驗的 S - P 圖表。

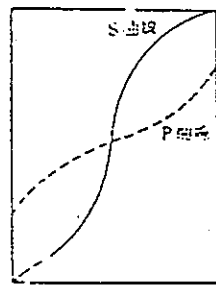


圖 4.3.1

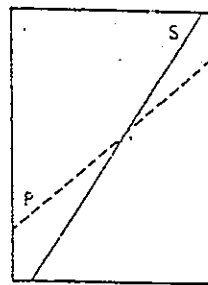


圖 4.3.2

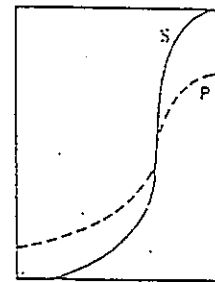


圖 4.3.3

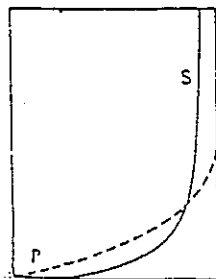


圖 4.3.4

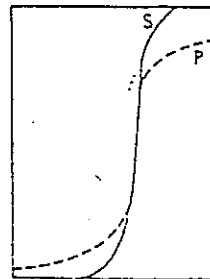


圖 4.3.5

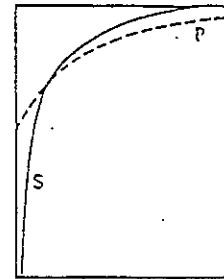


圖 4.3.6

第四節 S-P表的判斷要領

從S曲線與P曲線相分離情形而判斷

一. S-P兩曲線分離的標準，不能一概而論。通常是適當的分離，過與不及，均屬欠當。至於其標準，必須從經驗而判斷。換言之，依科目、測驗種類及方式等而有些差異。

二. S-P兩曲線的定量表示，根據佐藤隆博研究指出：差異係數 D^* 的算法 [12]、[13]、[14]，可供參考：

(一) 差異係數的意義——所謂差異係數是指『實際所得S-P表(大小 N (學生數) * n (試題數)，平均答對率 P) 的S、P曲線所包圍部分面積(偏差)與S-P(大小 N (學生數) * n (試題數)，平均答對率 P)，及隨機S-P表(大小 N (學生數) n (試題數)，平均答對率 P) 等各S-P兩曲線所包圍面積的差距，加以標準化之值。』 [15]。換言之，差異係數是S-P表全體非等質程度的係數。學生數為 N ，試題數為 n ，平均答對率 P 的實際S-P表之S-P兩曲線所包圍部分的面積 (I 與 O 合計)，則為 $S(N, n, P)$ 。其次，以學生數 N ，試題數 n ，平均答對率 P 為條件而作成隨機S-P表時之S、P兩曲線所包圍部分的之期待值，則可用 $E[Sr(N, n, P)]$ 表示之，其差異係數 D^* 為：

$$S(N, n, P)$$

$$D^* = \frac{\quad}{\quad}$$

$$E[Sr(N, n, P)]$$

因此， $D^* \geq 0$ ，完全 S-P 表時，其 $D^* = 0$ ，隨機 S-P 表時，其 $D^* = 1$ ，至於從經驗所得，要考察總和學力時，以 $D^* = 0.5$ 左右為標準。假如 D^* 超過 0.6，則『要注意』。其意義，就是表示含有異質的因素。因此，此種情況的試題，應設法修改。至於以練習為主的測驗因其試題群和學生群，均相當等質性，其標準以 0.4 左右，而超過 0.5 則要注意了。其一般電腦計算公式可參考佐滕隆博[16]。

(二) 差異係數的判斷與應用

1. 可供作學習輔導和命題改進的線索--通常以 $D^* = 0.5$ 為宜，如果 $D^* \geq 0.5$ ，則要檢討命題或指導技術。具體而言之，編序式學習的細目可能太粗，太難，或是試題的文字表達或敘述不清楚、含糊。若是練習，可能由於指導順序及教材排列欠當，乃至要求學生學習行為，如發問內容含糊不清等..... 所致。因此， D^* 的大小可供作發現此種原因的線索。茲將 D^* 值的標準列舉如下：

綜合測驗

$D^* = 0.5$ 標準

$D^* > 0.6$ 要注意

平時測驗

$D^* = 0.4$ 標準

$D^* > 0.5$ 要注意

2. 可供作一連串的資料能否當作一整個 $S-P$ 表而考察的線索。例如對某校新生入學當時，一連串指導的評鑑結果，製作 $S-P$ 表。但是，其 $D^* > 0.5$ ，於是再查構成當時 $S-P$ 表的資料，發現其資料是由二群不同內容性質的資料構成。因此，將其分別加以處理之後，其 D^* 值卻變成相當小了。由此可知，即使是單位時間或一連串的學習輔導資料，均有其提示教材，發問的目的與性質等，加以分別考察的必要。茲將其示例如下：

某項指導的測驗結果

平均答對率 $P = 0.52$

差異係數 $D^* = 0.46$

再指導的測驗結果

平均答對率 $P = 0.58$

差異係數 $D^* = 0.23$

3. 學業成就測驗結果的 S 曲線和 P 曲線，若是過於分離，不妨認為『可能有某些原因作祟』。因此，必須進一步加以調查或探究。通常形成 $S-P$ 兩曲線太分離的原因，可能下列九項：

(1) 教學不徹底，不清楚。

(2) 學習目標，學習內容與成就測驗的內容不相對應，或是測驗內容的編製，未能反映教學內容。

(3) 命題的順序或是各題之間的前後關係配置欠當，以及試題的內容，表達敘述，提示等不適當。

(4) 教學時期與測驗實施時期，在時間上相隔甚久。

(5) 學生的演練或練習不夠，或學生未達固定的層次。

(6)中等程度或低劣學生的學習上有不穩定的事情發生。

(7)學生的動機或意願低下。

(8)學生的心理上或生理上較大的變化。

(9)評分不一貫，評分標準不一，或評分的重點不明確，乃至評分錯誤、誤解等。

4. $S-P$ 兩曲線太接近也是有問題，應加注意例如學生的學習達成度(水準)甚低，或學習達成度測驗太難時， $S-P$ 兩線，多半會呈現相接近的形狀。

第五節 用以個別的學習診斷而判斷

- 一. 學習診斷不僅應了解個別學生的得分，而且應考察其對那些試題能通過，那些未能通過的情況。
- 二. $S-P$ 表的應用，對學習個別診斷，頗具價值。例如圖 4.5.1 可知 A 、 B 兩位學生的學習診斷結果，各有其特徵。在 $S-P$ 表上， A 生與 B 生，其得分均為 15 分，但是其內容，也就是反應模式，卻差異甚大。就 P 曲線（試題的試題率）分佈而考察。 B 生對全班同學幾乎答對的試題，也都能答對。全班同學答錯的試題（答對率低），也是答錯。至於 A 生則對全班同學答錯的試題（較難），卻能答對。反之，全班同學都能答對的試題（較易），卻答錯了，此種反應模式，可說是學習上欠缺穩定的現象。因此，對其答錯的試題內容為何？答錯的試題之內有何共同的性質，均有深加診斷的必要。

所謂注意係數是表示某生的反應模式異於其它同學的反應式達成度或未達成的傾向的量數。

- (一) 注意係數的意義 [17] --- 欲探究 $S-P$ 表中的各種反應模式，應以區分線為界而考察其『1』和『0』的分佈情形。因此，注意係數的定義可說：『所謂某反應的注意係數，是指以 $S-P$ 表中的區分線為界，並以其『1』和『0』在完全反應模式之下為基準，而表示實際反應模式與完全反應模式相差程度之值』。

實際反應模式與完全反應模式差異

$$\text{注意係數} = \frac{\text{與完全反應模式最大差異}}{\text{與完全反應模式最大差異}}$$

依據上述的定義，	注意係數 ≥ 0
完全的反應模式時	注意係數 $= 0$
隨機的完全反應模式時	注意係數 $= 1$

由此可知，實際的反應模式之注意係數，幾乎存在 0 與 1 之間。係數的值為 1 以上之機會，乃是『1』和『0』倒轉得特殊反應模式。就 S-P 表內而考察，顯然是特殊的模式。其一般電腦計算公式可參考佐滕隆博 [18]。

(二) 注意係數的標準

我們要從某次測驗中，了解學生的反應模式，是否異於其它同學，可使用學生注意係數 (Student Caution Index) 來判定。而其判定的基準從教育實際的案例探討結果如下：

1. $0.75 > C.Si \geq 0.50$ =====> 要注意。

2. $C.Si \geq 0.75$ =====> 應特別注意。

3. 此值為經驗所得的基準。

4. 0.50 基準值恰等於隨機反應模式的注意係數之 1/2 值。

5. 多因子的 S-P 表，如標準化學力測驗，其差異係數 D^* 值往往較大，因之，C.Si 的判定也較大。設 $C.Si \geq 0.60$ 時，要注意就可以了。

(三) 注意係數的判斷與應用

所謂注意係數大，就是指『應加注意』之意。但是，判斷時應留意下列：

1. 判斷注意係數時應與S或P曲線相對照。
2. 判斷學生的注意係數應與答對率相對照而判斷。

以下為注意係數的判斷方式，圖 4.5.2。

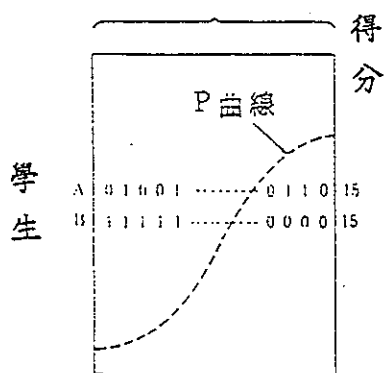


圖 4.5.1 從 P 曲線判斷學生正誤答的類型 (B 生是無問題，A 生可能不屬於學習穩定)

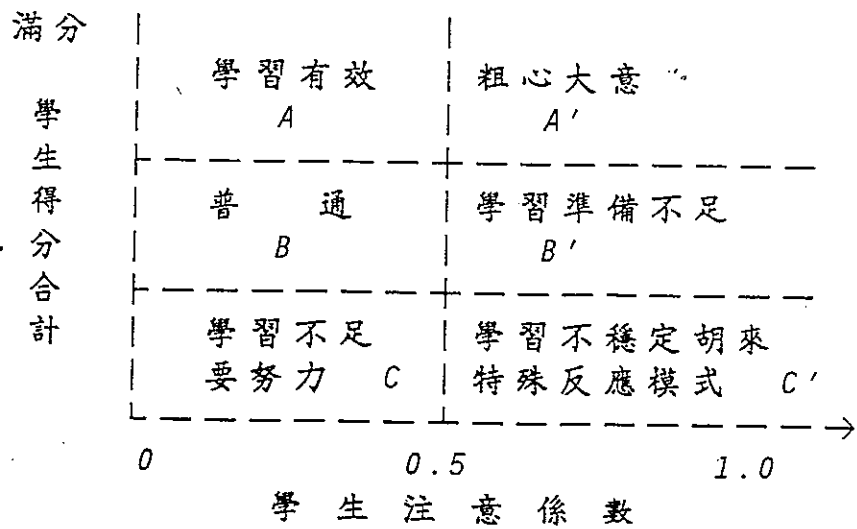


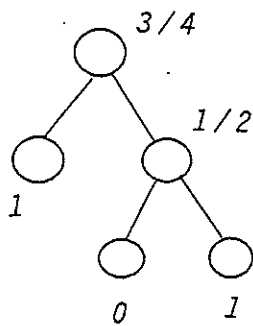
圖 4.5.2。學生注意係數之判別 (C.Si)

第五章 動態分析

在上章中，我們談述了 $S-P$ 圖表對學習評鑑方面所具有的功能特性後，同樣地，我們將此功能特性利用在評鑑輸入法上。經由測驗的結果以 $S-P$ 方式排列，以了解其狀況，而我們最主要利用差異係數的特性，因差異係數 D^* 的值猶如一個指標，指示我們對於某次測驗的狀況是否標準，或是否應注意。以讓我們對此輸入法不但可涉及一般評量方式，亦可對質的探討有所悟覺，但美中不足的是差異係數 D^* 祇能較含糊地指出可能情況原因，必須在配合其它資料訊息，因此，我們試著利用在第三章所談述的學習函數 $F(U)$ 及其它的服務項目以進行輔助的工作幫助我們了解評鑑上質的探討，然而在探討上，由於教學方式與心理因素為專業化知識 [如由學生注意係數 ($C.S_i$) 對每個學生的探討]，我們將不去涉及而保留給教育、心理專家外，我們主要針對本身教材的因素，由於欠缺實驗性的研究，所以我們以主觀觀點來探討。然而如果只對某次測驗來作評鑑分析，不免有所偏頗，是故我們應建立經由許多次測驗結果來作一個動態分析，以期我們的系統不但能摒棄過去祇求結果的比較外，使得提供一般使用者去選擇一個所希望的輸入法，也希望能提供輸入法系統設計者能瞭解如何改良或設計一個好的輸入法系統。我們將在下述各節討論我們的方式。

第一節 從學習函數 $f(u)$ 與實際正確值的差距來了解學習輸入法的困難之處。

我們知道每次的測驗結果皆有一個學習函數 $f(u)$ 值與一個正確值對應，然而 $f(u)$ 為我們對輸入法的了解程度，為一個迷糊集 (fuzzy set) 值，而因為一般的測驗當中得分方式，為絕對二分法即，不是 0 就是 1，所以 $F(U)$ 值與 S 值，(其中 $\sum_{i=1}^N \sum_{j=1}^M f(u)_{ij} = F(U)$ 為 N 個受測者，每個人在 M 題測驗中各題所得的 $f(u)$ 值，之總和。而 S 為全體總得分數。) 之間的差距為在得分為零的試題中，部分對應正確字碼的了解度所以 $F(U) - S$ 當其值頗大時，表示可能含有某些缺失，如字碼的析解對應中大的原則可以掌握，細微部分無法瞭解或字碼過多，部分難以記憶，而導致部份的析解對應字碼錯誤。

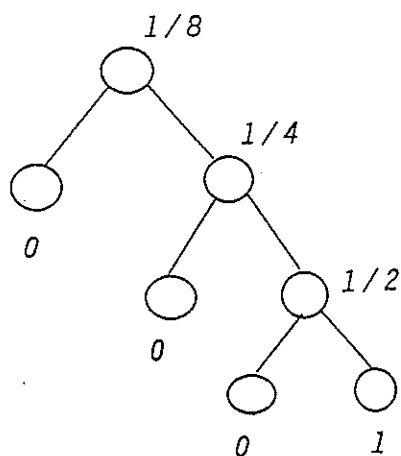


$$f(U) = 3/4$$

原因：細部規則不熟悉
而導致之錯誤

當其值頗小時，其可能原因：(1) 所瞭解規則皆可完整答對，而其它的規則則難以解出字碼，(2) 字碼過多難以記憶，或輔助字碼過多。

例如



$$f(U) = 1/8$$

原因：規則難以了解

然而對於其差的評定，我們須由實際試驗與經由教育專家來作一個評定，設定一些分界以判定其差距值應以何者方式來評定。

接著我們將配合 $F(U) - S$ 和 S 的關係來說明輸入法的一些特性而我們主要對 $F(U) - S$ 與 S 的極端值時的討論，即其值為極大、極小的情況。所謂極大當 $F(U) > K1$ 、 $S > K2$ ，所謂極小當 $F(U) < K3$ 、 $S < K4$ ，而 $K1$ 、 $K2$ 、 $K3$ 、 $K4$ 由專家來設定。而介於其間的值，我們視為一般情況，不在此討論。我們將分為三種 cases，而每個 case 又由兩個情況來探討，如下：

$$\text{CASE 1: } F(U) - S > K1, S < K4。$$

每個人分散在某些試題零星答對，而在其餘試題上大家均呈現了解度極高卻無法圓滿答對。

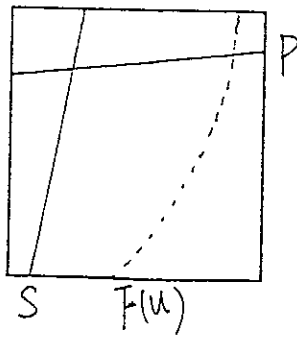


FIG 1.

所以我們認為對此輸入法大體上規則均可以掌握，但如圖所示由於 D^* 值極大，此種輸入法值得去注意，因為大家了解度極高，但呈現此種現象答對率，所以除輸入法本身教材外，是否因為教學方式與學生心理所影響是值得去斟酌的，而由於 $F(U)$ 值極大的狀況下，所以我們可以預測下一次得分率會高。

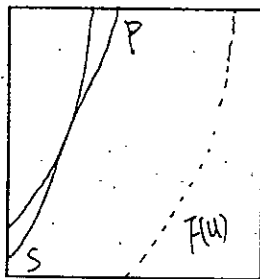


FIG 2.

CASE 2:

$$F(U) - S < K3 \text{ 和 } S > K2$$

此種輸入法造成分野太明顯，適於高程度的人使用，而 $F(U) - S$ 值極小，明顯顯示不會的題目頗難，呈現了解度極低，所以可預測下一次增加的分數不會太多。

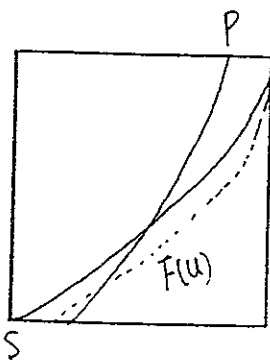


FIG 1.

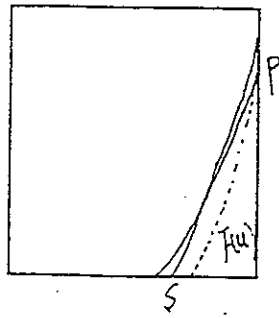
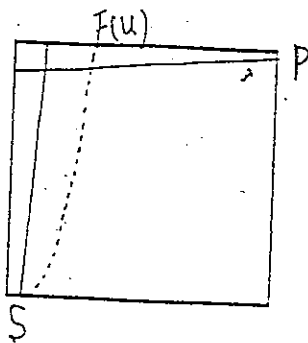


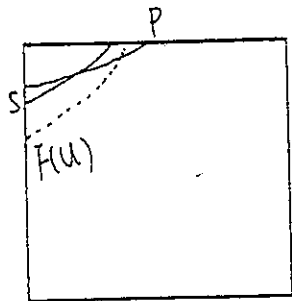
FIG 2.

題目大部分對大多數人均能答對，祇是對少部分試題了解度極低。而如圖所示 D^* 小為適中的輸入法。

CASE 3: $F(U) - S < K3$, $S < K4$ 。



每個人分散在某些試題零星答對而其餘試題上大多呈現了解度極低，由於圖示 D^* 值極大，值得去注意，除輸入法本身教材外，是否因為教學方式與學生心理所影響是值得去斟酌的，而 $F(U)$ 值極小，可以預測下一次得分率會低。



少部分簡易的題目，大家均能答對但對其餘的題目了解度卻極低，而且由圖示 D^* 值極小顯示此輸入法極不易學習。而 $F(U)$ 值極小，可以預測下一次得分率會低。

FIG 1.

第二節 從學習函數 $f(u)$ 與實際正確值的差距和鍵盤熟悉度來了解鍵盤排列的困難度

透過鍵盤來輸入中文字，所以鍵盤熟悉度亦為我們關心的一個問題，因為鍵盤字的多寡與位置的排列常為影響熟悉度之因。以下為我們得公式

$$X = \frac{\sum_{i=1}^M \frac{n_i - e_i}{n_i}}{M}$$

其中 n_i : 表示字碼 i 在此次測驗中，出現的總次數。

e_i : 表示字碼 i 在此次測驗中，錯誤的總次數。

M : 表示此次測驗出現的基本字碼個數。

也就是說，我們以字碼為單位(取代以字為單位)，並加入其字碼出現頻率，來探討對鍵盤的熟悉度。

例如：

某次測驗中出見的基本字碼為 A, B, C, D, E 而其出現頻率與對、錯頻率狀況如下表。

	A	B	C	D	E
出現總數	1	2	3	4	5
錯誤次數	1	1	2	2	2
正確次數	0	1	1	2	3

所以其對鍵盤排列熟悉度算法為

$$X = \frac{(0 + 1 + 1 + 2 + 3 + 29) + (1 + 2 + 3 + 4 + 5 + 75)}{5} = \frac{102}{5} = 20.4$$

對 X 值的得到，我們可作為判斷對鍵盤的熟悉度，但對於熟悉度實際的分野判斷感受我們希望由專家或學者們來判定。譬如我們假設 $IF X > \text{某一個值 } W5 \text{ THEN 熟悉}$ ， $IF X < W6 \text{ THEN 陌生}$ 。而其產生原因大約分類如下：

- (1) 鍵盤位置排列 [19] [20]。
- (2) 字碼數多寡對記憶能力的負荷。

同樣的，在考慮熟悉度之餘，我們對於輸入速度 (V)，亦需考慮，其受下列影響：

- (1) 思考時間。
- (2) 重複率。
- (3) 平均字碼數多寡，導致鍵入 (KEY-IN) 字碼時間長短。
- (4) 鍵盤位置安排，導致左右手負荷是否平衡。

而一個中文字對應鍵盤字碼學習函數 $f(u)$ ，而

$$F(U) = \sum_{i=1}^n \sum_{j=1}^m f(u)_{ij} \text{ 我們均呈現下列型式:}$$

即一個中文字的 *NON-TERMINAL NODE* 即直接為其基本字碼 *TERMINAL NODES* 的組合。而其定義公式同前所數。

接著我們將配合 $F(U) - S$ ， S ， X 和 V 的關係來說明輸入法的一些特性而我們主要對 $F(U) - S$ ， S ， X 與 V 的極端值時的討論，即其值為極大、極小的情況。所謂極大當 $F(U) > W1$ 、 $S > W3$ 、 $X > W5$ 、 $V > W7$ ，所謂極小當 $F(U) < W2$ 、 $S < W4$ 、 $X < W6$ 、 $V < W8$ ，而 $W1$ 、 $W2$ 、 $W3$ 、 $W4$ 、 $W5$ 、 $W6$ 、 $W7$ 、 $W8$ 由專家來設定。而介於其間的值，我們視為一般情況，不在此討論。我們將分為三種 cases，而每個 case 又由兩、三個情況來探討，如下：

CASE 1: $F(U) - S > W1$ ， $S < W4$ ， $X > W5$ ， $V < W8$

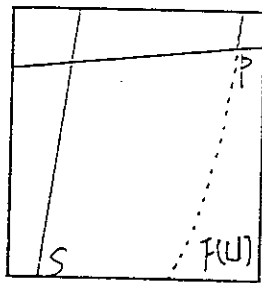


FIG 1.

每個人分散在某些試題零星答對，而在其餘試題上大家均呈現了解度極高卻無法圓滿答對。

由熟悉度 X 知極大，即對鍵盤排列大體上均可以掌握，但由 S 極小下我們不禁感到訥悶，為什麼在熟悉度極大下會有如此情況。是否因鍵盤某些字碼位置無法牢記，或者學生心理因素影響。而且如圖所示由於 D^* 值極大，此種輸入法值得去注意。由於 $F(U)$ 的極大下，我們可以預測下一次得分率會高。

答對題目均集中某些試題，但其餘試題上，大家了解度極高，而卻無法圓滿答對。

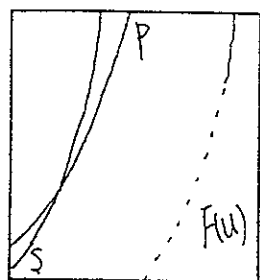


FIG 2.

由熟悉度 X 知極大，即對鍵盤排列大體上均可以掌握，但由 S 極小下我們不禁感到訥悶，為什麼在熟悉度極大下會有如此情況。

但由圖示 D^* 值極小，而 $F(U)$ 值極大，故為適宜的排列，而由於 $F(U)$ 值的極大下，我們可以預測下一次得分率會高。

CASE 2 $F(U) - S < W2$, $S > W3$, $X > W5$

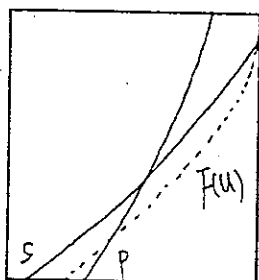


FIG 1.

此種情況造成分野太明顯，適於高程度的人使用，而 $F(U) - S$ 極小，明顯顯示不會的題目頗難，呈現了解度極低，所以可預測下一次增加的分數不會太多。而如果 $V < W8$ \Rightarrow 應注意上述影響原因。

題目大部分對大多數人均能答對，祇是對少部分試題了解度極低。而如圖所示 D^* 小，為適中的輸入法。而速度上又分為兩種：

(1) $V > W7 \Rightarrow$ 正常。

(2) $V < W8 \Rightarrow$ 應注意上述影響原因。

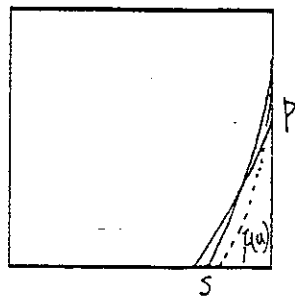


FIG 2.

CASE 3 $F(U) - S < W2, S < W4, X < W6, V < W8$

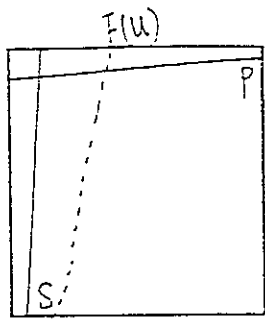


FIG 1.

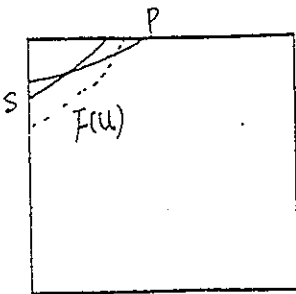


FIG 2.

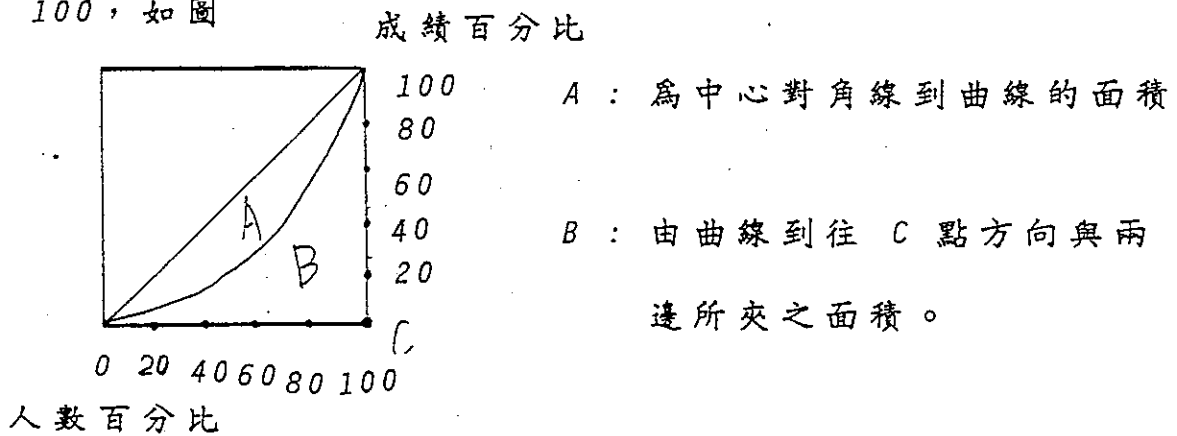
每個人分散在某些試題零星答對而其餘試題上大多呈現了解度極低，由於圖示 D^* 值極大，值得去注意，除排列方式外，是否因為學生心理所影響是值得去斟酌的，而 $F(U)$ 極小，可以預測下一次得分率會低。少部分簡易的題目，大家均能答對但對其餘的題目了解度卻極低，而 S, X, V 均低而且由圖示 D^* 值極小，顯示此輸入法極不易學習。而 $F(U)$ 極小，可以預測下一次得分率會低。

利用 $S-P$ CHART 的差異係數 D^* ，一般可以作為學習評鑑，一般 D^* 值以愈大愈不理想，但 D^* 愈小也不一定好，所以搭配 S 和 $F(U) - S$ 值，整理出可能發生的狀況，更深入的去探討發生各種現象的原因所在，以便做進一步之改良與修正。綜合上節與本節之討論，整理如下：

D^*	$F(U) - S$	S	圖形
大	大	小	CASE 一 . FIG1
	小	大	CASE 二 . FIG2
	小	小	CASE 三 . FIG1
小	大	小	CASE 一 . FIG2
	小	大	CASE 二 . FIG1
	小	小	CASE 三 . FIG2

第三節 由吉尼 (GINI) 係數來判斷分佈狀況

不論在輸入法學習階段或鍵盤速度階段我們均可得到一些結果值，於是我們希望由這些結果來分析使用者的成績分佈狀況以明瞭其是否呈現特異現象，以了解是否此輸入法會造成高低程度明顯分明，我們使用類似羅峇士曲線與吉尼係數的作法 [21]，我們作法是以人數百分比為橫軸，以成績百分比為縱軸，軸間之區分單位為百分之 20, 40, 60, 80, 100，如圖



因此資料在圖上便成羅峇士曲線，而呈 A, B 兩區域，當圖形曲線愈偏向 C 點時愈呈現不平均，而愈偏向中心對角線時為均勻，而其測度公式我們引用吉尼公式為

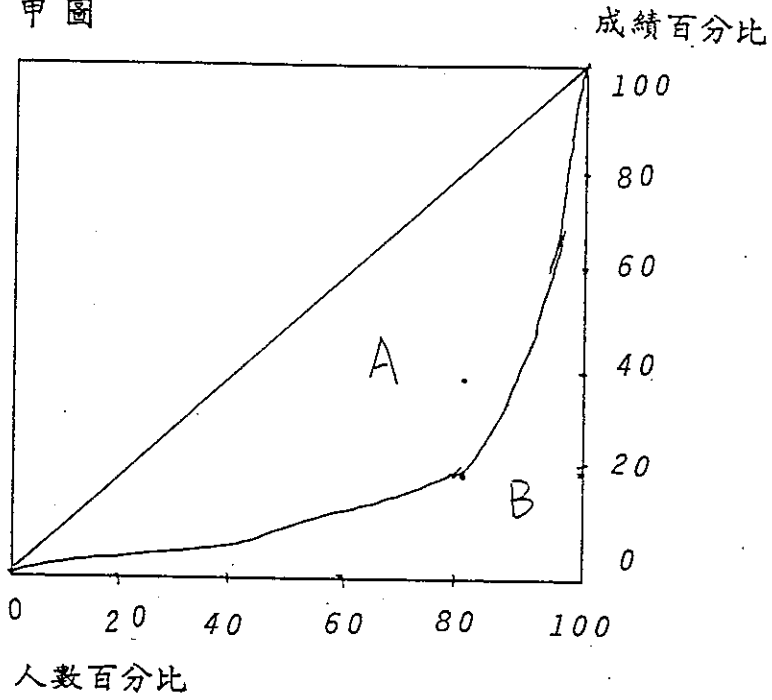
$$\text{吉尼值} = A / (A + B)$$

其值介於 $0 \leq gini \leq 1$ ，當愈趨近 0 愈平均，愈靠近 1 愈不平均，是應注意，例如：

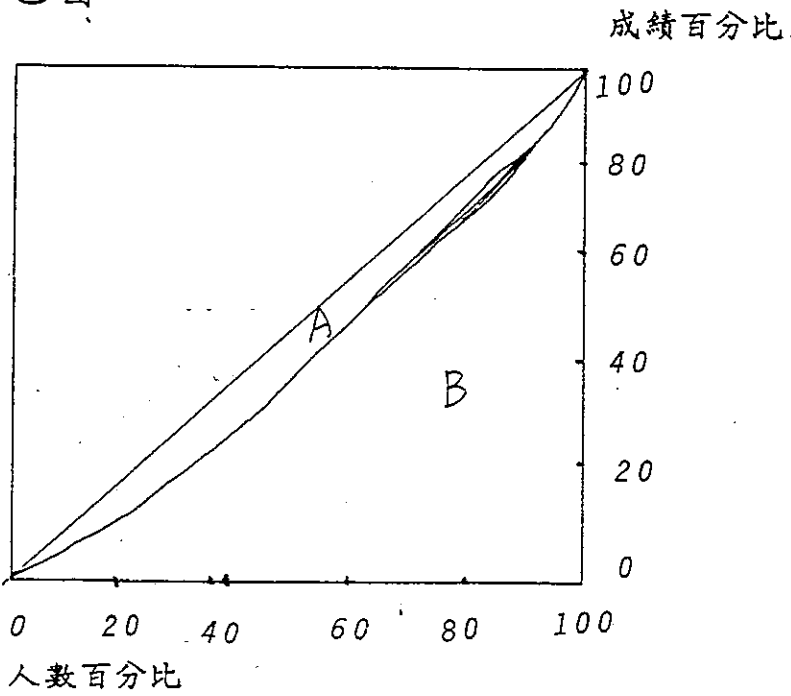
如甲圖：大部分的人得分占總人數得分 20% 的百分比，而其餘少部分的人卻得大部分的分數，所以曲線偏向 C 點，而其吉尼值為而此種明顯分開的輸入法值得考慮。而乙圖：曲線偏向中心線，顯示很平均的輸入法。

例如

甲圖



乙圖



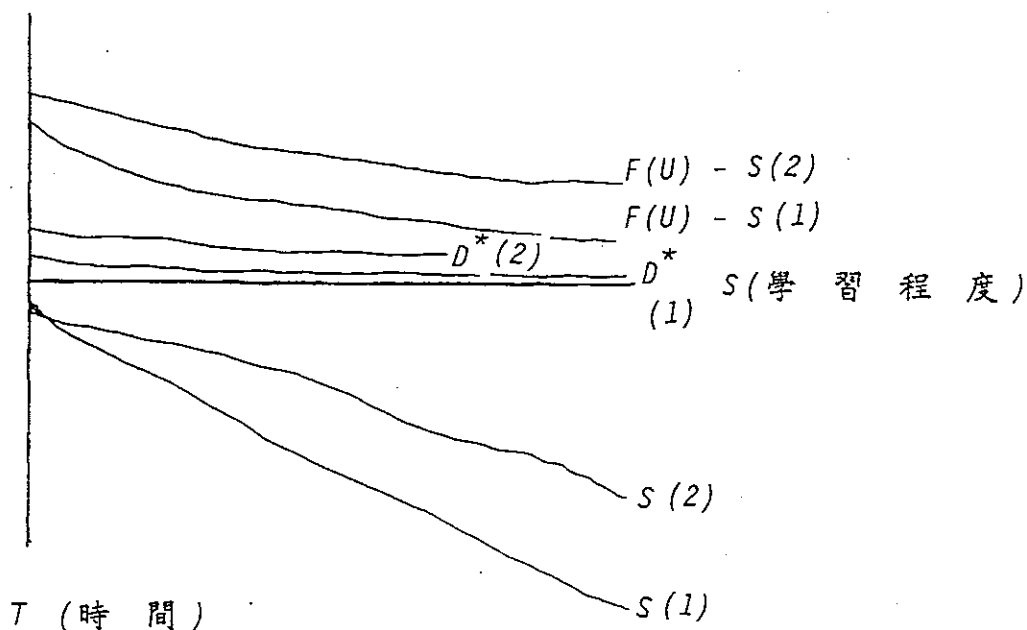
第四節

動態變法分析說明

單由一次測驗結果分析很難決定這個輸入法的優缺點，所以我們必須以多次的動態測驗分析結果來作評定，以使得使用者選擇與輸入法設計者得以改良。

一. 由 $F(U) - S$ 和 D^* 值中的變化來觀察學習程度與時間的狀況。如圖

$D^*, F(U) - S$

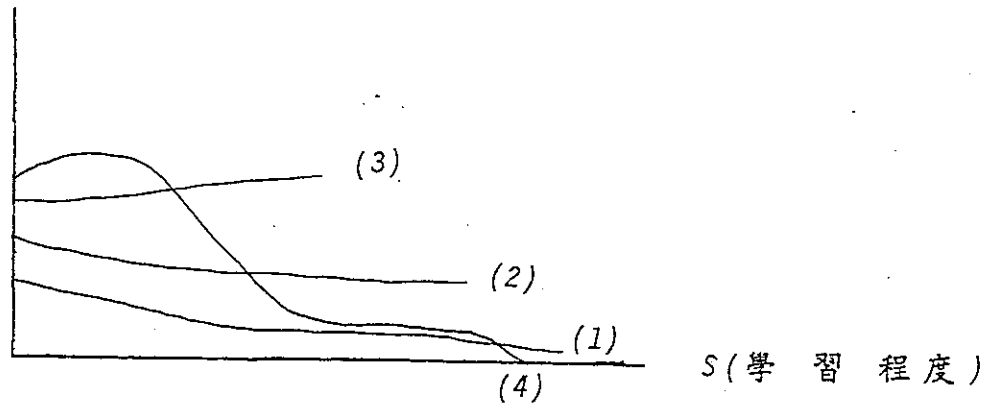


由 (1) 可知， D^* 值一直相當接近標準值，而 $F(U) - S$ 隨著 S 的增長而差距愈小，而且 S 成長較快，所以為一個較能被使用者接受的輸入法或鍵盤。

由 (2) 可知， D^* 值卻一直遠離標準值，可見必須注意為何如此，而 $F(U) - S$ 值雖然隨 S 的增長而差距愈小，但卻緩慢減小，而且 S 成長較緩，所以此種輸入法較無法被使用者所接受。

二. 以多次測驗結果的利用 GINI 係數來看成績高低分佈狀況。

吉尼係數



(1) : 呈現緩慢而平穩的朝向均勻分佈。

(2) : 呈現明顯分野，顯示此種輸入法值的考慮。

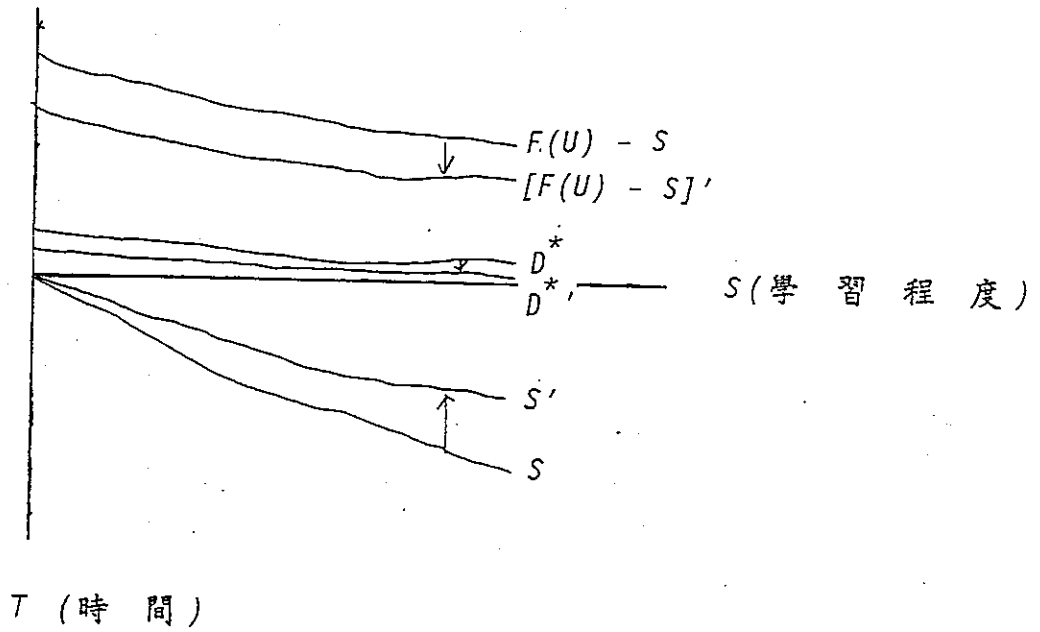
(3) : 呈現極端不平均，顯示此種輸入法值的考慮。

(4) : 呈現快速朝向均勻分佈。

所以 (1)、(4) 為可接受的輸入法，而 (2)、(3) 則難以接受。

三. 經改良後的狀況。

下圖我們假設為假如原本學習評鑑後的結果為 $F(U) - S$ ， D^* 和 S ，而經過改良成為 $(F(U) - S)'$ ， D^{*} ， S' ， D^{*} ， $F(U) - S$



第六章 結論與建議

在這篇論文中，我們針對一般評估方法祇考慮『量的評鑑』而無『質的探討』的缺失，而加入以 $S-P$ 圖表分析的技術，在值方面作更進一步的探討。另外針對一般考量還是以測驗二分法(即不是答對就是答錯的評判方式)的缺點，提出一個以迷糊集(Fuzzy set)概念為基礎的學習函數 $F(U)$ 之作法與公式。並同時輔助 $S-P$ 圖表中差異係數 (D^*) 解釋的不足，作進一步的分析與探討。在我們全部的處理過程中，由於我們本身採取主觀看法來解釋各種現象，所以在某些方面還需要加以深入研究。

1. 由於學習函數 $F(U)$ 的公式導出為直覺作法，欠缺客觀，譬如是否應考慮字碼出現頻率，階層(LEVEL)的深、淺度或寬度是否應給予一個加權數(WEIGHT)是值得注意。因此我們希望在以後能夠將這些因素考慮進去，以期能夠找到一個更客觀的公式。
2. 由於 D^* , $F(U)-S$, S , P 所產生的變化相當多所以我們希望能由專家或學者，將極端值間的一些變化情形作實驗性的判斷以建成一個表格(TABLE)使得測驗結果在輸出後，即可知其狀況。
3. 關於鍵盤的安排，我們建議輸入法設計者可參考[19][20]。
4. 希望對加權數(WEIGHT)能提出一個客觀的公式。

參 考 文 獻 (REFERENCES)

- [1] C.K.CHEN , " INTRODUCTION TO CHINESE LANGUAGE COMPUTER AND VERDORS," III TECHNICAL REPORT C16 ,PP. 6-21,39-44,TAIPEI,MARCH 1982.
- [2] 1987 財團法人資訊工業策進會 : 中文電腦輸入法與輸入器調查評估報告 .
- [3] C.K.CHEN AND R.W.GONG , " EVALUTION OF CHINESE INPUT METHODS ", COMPUTER PROCESSING OF CHINESE & ORIENTAL LANGUAGES VO1.1 , NO.4, NOVEMBER 1984.
- [4] S.Y.LO , " A SCIENTIFIC MODEL FOR COMPARING VARIOUS METHODS OF INPUTTING CHINESE CHARACTERS INTO COMPUTER", COMPUTER PROCESSING OF CHINESE & ORIENTAL LANGUAGES. VO1.2, NO.1, MAY 1985.
- [5] 盧東華, 胡明森, 和鄭國揚 "從表意文字的觀點看圖像的形成", 全國計算機會議, 76年.
- [6] 林樹, "中文電腦基本用字研究", 交通大學工學院, 科技研究報告 CC-601 號.
- [7] 簡茂發 "教學評量原理與方法" 國民中學學習成就評量理論與命題示例國文科 臺灣省政府教育廳 73年.

[21] 郭婉容著, "總體經濟學", "個體經濟學".