

應用管制圖建構試題品質評估之指標

USING CONTROL CHART TO CONSTRUCT THE INDEX OF EVALUATION FOR THE TEST QUALITY

蘇純縉

國立雲林科技大學工業管理系教授

丁秋瑤

國立雲林科技大學管理研究所博士生

Chwen-Tzeng Su

*Department of Industrial Management Professor
National Yunlin University of Science and Technology*

Chiu-Yao Ting

*Graduate School of Management
National Yunlin University of Science and Technology*

摘要

本研究主要以考科單元能力 (Ability) 的觀點, 提出應用管制圖 (Control Chart) 工具之品質管制特性, 並且結合相關的統計方法, 據以評估測驗 (Test) 命題之情形。本研究除了將試題品質 (Test Quality) 指標 (Index) 依照一般測驗的評鑑標準進行探討外, 也將以標準差 (Standard Deviation) 作為管制圖的管制界限, 以建立一個試題品質的評鑑標準。研究結果顯示, 本文所使用的評鑑標準, 不僅能更清楚地了解品質等級, 並且可以更明確地指出試題的品質情形。

關鍵字: 能力、管制圖、指標、統計方法、試題品質

ABSTRACT

Based on the view of the unit ability, the study purposes to employ features of control chart and to integrate relevant statistical methods for assessment of the test quality in the

joint examination. This article achieves the establishment of criteria for assessing test quality by exploring the index of test quality in the general test area, and using standard deviation for the control limit of control chart. The result of the research indicates that the criteria can demonstrate the levels of test quality in finer state and more accurately point out the situation of quality.

Key Words: ability, control chart, index, statistical methods, test quality

壹、前言

測驗 (Test) 為一種衡量一個人在某一領域的能力或知識的方法 (Brown, 1994), 聯合招生考試的主要目的是用來診斷學生能力, 以作為考生分發學校的依據。測驗中的每一個試題均扮演重要角色, 假若各個試題不能夠發揮功能, 則整個測驗的功能自然會受到影響 (范德鑫, 1990)。

試題分析 (Item Analysis) 為檢視考生對於每一個試題的反應, 以判定試題品質的過程 (Mehrens & Lehmann, 1984)。由於試題可以重複運用, 若每次能修改部分試題, 那麼最後整個測驗品質將有所改進, 亦可作為建立題庫 (Item Bank) 之用。透過試題品質分析, 每次保留優良的試題, 修正較差的試題, 或者刪掉不良的題目, 如此一來, 必能累積大量的優良試題, 提供未來測驗命題之使用。

目前針對每一題考題品質之評估方法, 在學術研究已多有著墨, 而入學測驗個考科命題原則乃根據雙向細目表 (Two-way Specification Table) 中的知識能力為基礎, 一份試題主要由幾種題型所構成, 如本文章所探討的生產管理科試卷係由產能規劃、品質管理等單元所組成。然而, 由於測驗係鎖定能力來命題, 才能引導教學以能力為導向, 讓學生真正學習到應該具備的能力, 因此, 以單元能力角度做試題品質評估分析的相關研究仍顯不足。

現代測驗理論的試題分析有一個重要的特點, 那就是以圖形呈現測驗試題的某種特徵 (Wainer, 1989; 李正源, 1996)。管制圖為統計製程管制 (Statistical Process Control, SPC) 的主要工具之一 (Chen & Liao, 2004), 經常被使用於品質管制上, 而本研究所欲分析的測驗資料為難易度 (Item Difficulty), 屬於計量值 (Variable) 之品

質特性 (Quality Characteristics)。此外，亦可應用管制圖的規格界限 (Specification Limits) 作為試題品質等級的分類標準。因此，本文章基於上述幾點原因，以整個能力的觀點加以審視，並且提出應用管制圖工具之品質管制特性，以及結合相關的統計方法，進行試題的量化分析 (Quantitative Analysis)，據以評估生產管理科測驗命題之優劣。

本研究除了將試題品質指標依照一般測驗的評鑑標準進行探討外，也將以標準差作為管制圖的管制界限，以建立一個試題品質的評鑑標準，進而了解各單元能力是否達到良好或可接受程度之試題品質，並且提供改善試題品質的建議。在文章安排上，依序為前言、文獻探討、研究方法、實證分析、研究結果對試題品質的改善，以及結論。

貳、文獻探討

文獻探討部分將針對試題分析及管制圖相關文獻進行探討。

一、試題分析

要能夠真正落實考出學生的程度不外乎就是試卷品質，而試卷品質的優劣攸關了眾多的考生權益，對於影響更是深遠。一般而言，試題分析分為質的分析 (Qualitative Analysis) 與量的分析兩部分，說明如下：

(一) 質的分析

質的分析是指在編製試卷時，以有邏輯思考的方式對試題取材的適切來做評估鑑定指標。編製試卷者一般在編製試卷時，也應該著重考量學生的「教材內容」與「教學方針」兩個因素，並且用此兩因素來作為測驗編製的藍圖架構 (Blueprint)，建立雙向細目表，來作為測驗編製時的根據，對於教學目標的分類依據，一般是參考 Bloom et al. (1956) 所提出的六個認知層次。雙向細目表可以使得整份試卷在內容上及認知層次上能有適當的分配，也就是說試題的安排可藉由雙向細目表而顯得比較有結構。

(二) 量的分析

量的分析係採用統計方式來分析試題的品質，包括難易度及鑑別度 (Discrimination) 分析，一個測驗信度及效度的高低完全取決於試題的品質，因此藉由題目分析，可以提高測驗的信度和效度。量的分析亦為本研究所欲探討的主題，以下針對難易度及鑑別度進行探討：

1. 難易度

編製測驗時，分析試題的難易度最主要的目的在於要選取難度適當的題目。Kerlinger and Lee(2000)指出難易度若稱為容易度或許更為貼切。Reynolds, Livingston, and Willson (2006)提及難易度為評估能力測驗中重要的考量。難易度適當的試題是構成優良測驗的必要條件，通常以全體受測試者通過或是答對該試題的百分比來表現，題目越容易，通過的百分比的比例也就越高。

P 值就是直接將答對該題的人數比例表示出來，P 值本身並無法指出試題的好壞，而只是單純地指出對於受試團體而言，該題的難度，因此，如果換成另外一組人作答，所得到的 P 值也就不同，P 值這種團體參照 (group-referent) 的特性，稱為取樣依賴 (sample dependence) (Maller, 1997)，為了避免取樣依賴的特性，測驗或量表編製者必須慎選預測的樣本，亦即樣本之選取必須具備代表性，如此，所得到的 P 值才有分析的價值。如果試題本身為二分 (對或錯)，則最常被使用的試題難度指數就是 P 值 (李正源，1996)。

難易度的計算方式可概括分為三種方式 (簡茂發，1991)，公式(1)與公式(2)均為難易度之計算公式，P 值代表試題難易度，N 值代表全體受測試者人數，R 值為該答對題的人數，其中， P_H 高分組是指測驗總分中最高 27% 之受測試者的難易度，而 P_L 低分組是指測驗總分最低的 27% 之受測試者的難易度：

$$P = R/N \times 100\% \quad (1)$$

$$P = (P_H + P_L) / 2 \quad (2)$$

P 值介於 0 與 1 之間，而 P 值的計算出來結果越大則表示試題對受測試者越是容易。反之，P 值越接近 0，表示試題越困難。公式(2)的算法只使用到高、低分組考生所得到的 P，精確性不及使用所有受試者的公式(1)計算所得的 P。

第三種難易度計算方式為美國教育測驗服務社 (Educational Testing Service, ETS) 創立一種等距量尺 (Interval Scale) 特性的難易度指數，其不僅可以表示出試題難度

的相對位址，還可以指出各難度之間差異性的大小。是假定一個試題所測量的特質呈常態分配，它是一種以 13 為平均數、4 為標準差、下限為 1，上限為 25。值越大難度越高；反之，值越小難度則越小，Z 值為該試題難度的標準差值，可由范氏項目分析表查出 (Fan, 1952)，其計算公式如下：

$$= 13 + 4Z \quad (3)$$

2. 鑑別度

所謂試題鑑別度係指能區分成就高的考生（高分組）及成就低的考生（低分組）的能力，試題的難度與鑑別力有著密切的關係，如圖 1(郭生玉, 1990)及表 1(Reynolds et al., 2006)所示，當 $P=1.0$ 或 0.0 時，鑑別度將為 0；而當 $P=0.5$ 時，可得到最大的鑑別度 +1.0。為了使試題鑑別度達到最大，必須選擇難易度適中的試題來組合測驗，測驗才能達到區別各種不同能力水準的功用。至於試題的難易度應視測驗的目的而定， P 越接近 0.5 時最適宜，因為此時鑑別度可達到最大 (Gronlund, 1976)。另外，Osterlind (1989) 提及如何選擇試題並無一定的標準，大致上難易度在 0.4 至 0.8 之間，即可加以考慮。因此，選擇試題時，似乎盡可能地選擇難度接近 0.5 的題目為宜。

二、管制圖

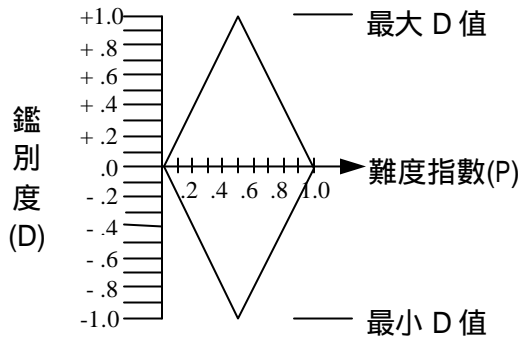
典型的管制圖包含中心線 (Center Line, CL) 及兩個管制界限：上管制界限 (Upper Control Limits, UCL) 和下管制界限 (Lower Control Limits, LCL) (Yang & Yang, 2005)。管制圖可用公式 (5) 至公式 (7) 之通式來表示，設 X 為品質特性的樣本統計量， X 的平均數為 μ_x ，標準差為 σ_x ，中心線表示在沒有任何不正常的變數干擾下的製程特徵， K 為管制界限至中心線的距離，以標準差的倍數來表示，如圖 2 所示。

$$UCL = \mu_x + K\sigma_x \quad (5)$$

$$CL = \mu_x \quad (6)$$

$$LCL = \mu_x - K\sigma_x \quad (7)$$

管制圖可以分成兩類，假若品質特性可以量化且為連續數值，則屬於計量值，很多品質特性可以計量值表示，一個可量測的品質特性，如尺寸、重量，以及本研究所探討的難易度等均屬於計量值，此種品質特性的管制圖，稱為計量值管制圖。假若品質特性不可量化，則以某些屬性來表示，如合格品或不合格品，或是每一件產品上的不合格點數，此種品質特性的管制圖稱為計數值管制圖 (Montgomery, 2001)。



資料來源：郭生玉 (1990)，心理與教育測驗，第五版，台北：精準。

圖 1 難度與鑑別力的關係

表 1 不同難易度之最大 D 值

P	Maximum D value
1.0	0
0.9	0.2
0.8	0.4
0.7	0.6
0.6	0.7
0.5	1.0
0.4	0.7
0.3	0.6
0.2	0.4
0.1	0.2
0.0	0



圖 2 一個典型的管制圖

管制圖假設某一品質特性服從常態分配，平均數為 μ ，標準差為 σ ， x_1, x_2, \dots, x_n 表示樣本數為 n 的樣本，樣本的平均數如公式(8)所示：

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (8)$$

由於 \bar{x} 為常態分配，平均數及標準差分別為 μ 和 $\frac{\sigma}{\sqrt{n}}$ ，因此，樣本有 $1 - \alpha$ 的機率落在製程平均數的上下管制界限，如公式(9)所示：

$$\mu \pm Z_{\alpha/2} \sigma_{\bar{x}} = \mu \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad (9)$$

【定義】：管制界限

管制界限受制於製程的自然變異，以製程標準差計算之，以及受到製程的自然允差界限 (Natural Tolerance Limits) 影響，通常取三倍標準差所做的界限稱為自然允差上下界限，以 UNTL、LNTL 表示。

【定義】：規格界限

規格界限則是由管理人員、顧客甚至是產品開發者來訂定，整體來說，管制界限與規格界限之間並無任何數字或統計上的關係，只有在繪製單一觀測值 (非平均值) 時，繪上規格界限才具有意義，圖 3 為管制界限、規格界限與自然允差界限之關係。

參、研究方法

本研究主要以考科單元能力的觀點，提出應用管制圖工具之品質管制特性，並且結合相關的統計方法，據以評估品質之情形。在研究工具方面，採用統計軟體 Minitab。

一、試題資料之常態性、隨機性及獨立性檢定

應用管制圖作為評估試題品質的工具，除研究資料需服從常態分配外，尚需對隨機性及獨立性加以檢定，才能夠利用管制圖的特性，進行品質管制。在常態性檢定 (Normality Test) 方面，可以繪製常態機率圖 (Normal Probability Plot) 決定資料是否服從常態分配，以下為分析資料是否符合管制圖常態性的假設檢定：

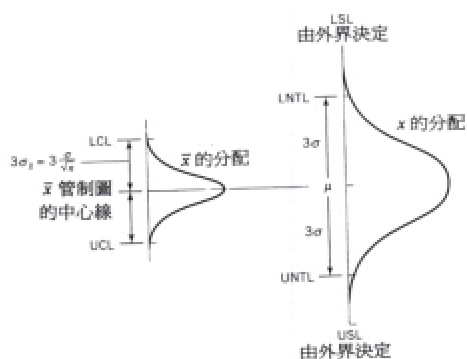


圖 3 管制界限、規格界限及自然允差界限

H_0 ：資料服從常態分配

H_1 ：資料不服從常態分配

在隨機性方面，檢定資料殘差項是否具有隨機性可使用連串檢定（Runs Test），以下為分析資料是否符合隨機性的假設檢定：

H_0 ：資料符合隨機性

H_1 ：資料不符合隨機性

在獨立性檢定方面，可利用時間數列的自相關函數（Auto Correlation Function, ACF）檢定資料是否具獨立性。

二、單元能力試題難易度之信賴區間檢定

本研究將品質評估分為兩種情形，第一種情形為測驗領域常用的評鑑標準，假設為 P 情形，第二種情形為本研究所提出的評鑑標準，即標準化管制圖，稱為 Z_p 情形，進行試題品質情形之判斷。

在 P 情形方面，根據文獻探討的內容，本研究根據過去學者訂定的難易度指標值（ $P < 0.39$ ：難， $0.40 < P < 0.80$ ：中， $0.8 < P$ ：易， $P = 0.5$ ：最佳品質），作為試題品質判斷的標準，以及繪製管制圖的依據。依照上述的指標值，將個別管制圖之上界限（Upper

Bound, UB), 假設為 0.8, 位於上界限以上的試題屬於「易」之試題, 以及下界限(Lower Bound, LB), 假設為 0.4, 位於下界限以下的範圍屬於「難」之品質, 若位於上界限及下界限之試題, 則定義為「中」品質的試題。其中, Target 表示為最佳試題品質 $P=0.5$, Mean 表示此單元各試題之平均難易度。

在 Z_p 情形方面, 係以應用管制圖中的標準差倍數作為評鑑標準, 即標準化管制圖, 因此, 需將難易度 P 值轉換成標準常態 Z 值, 以進行管制圖之繪製。由上文中可知 $P=0.5$ 為最佳的品質, 若將 P 值轉換成標準常態 Z 值時, 則中心線 (CL: $Z=0$) 代表 $P=0.5$, 如果 $P < 0.5$, 則將落於中心線之下, 反之, 則落於中心線之上。

管制圖上可以使用兩組管制界限, 外部的界限為 3σ , 為常用的行動界限 (Action Limit), 至於內部界限則可視使用目的為何, 選擇倍數的標準差, 稱為警告界限 (Warning Limit), 警告界限的使用可以增加管制圖的敏感度, 即它可以更迅速的察覺製程中參數的偏移 (Montgomery, 2001) 本研究主要目的在於建構一個新的試題品質評鑑之標準, 使用管制界限 3σ 的原因為落在平均數加減三個標準差的界限內有 .9973 的難易度之試題。本研究依據標準差將評鑑標準分成在中心線以上的 PA ($+1\sigma$) 區: 佳、PB ($+2\sigma$) 區: 易, 及 PC ($+3\sigma$) 區: 極易, 在中心線以下分成 NA (-1σ) 區: 佳、NB (-2σ) 區: 難, 及 NC (-3σ) 區: 極難, 總共分類成六個品質區域。

在警告界限方面, 則使用 1σ , 原因為一個標準差的界限內有 0.6828 的數值, 若對照 Z 表, 則可知道在正負一個標準差時, 難易度介於 0.1587~0.8413, 大多數試題的難易度包含於此範圍, 假若使用 2σ , 則難易度將介於 0.0228~0.9772, 故本研究使用 1σ 作為警告界限。

本研究所測量的觀察值來自常態母體, 標準差已知, 由於各單元答題數量為小樣本, 故檢定統計量 \bar{X} 的抽樣分配屬於 Z 分配, 在雙尾檢定中, 二個臨界值 c_1 與 c_2 所構成的之區間為

$$\left(\bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right) \quad (10)$$

信賴區間檢定的觀念為在雙尾檢定中, $H_0: \mu = \mu_0$; $H_1: \mu \neq \mu_0$, 若母體平均數 μ 的 $(1-\alpha) 100\%$ 之信賴區間公式(10)包含 μ_0 , 則 \bar{X} 的觀察值會落於接受域, 此時將做出接受 H_0 的假設。

信賴區間估計檢定意指利用信賴區間的觀念做檢定，計算母體參數的信賴區間，若此區間包含 H_0 成立時的假想值（如 μ_0 ），則接受 H_0 ，反之，則拒絕 H_0 。

肆、實證分析

本研究主要目的在於分析生產管理科各單元能力之試題品質，實證分析將針對問題描述、資料處理、試題資料之常態性檢定、試題品質之管制圖與信賴區間檢定，進行說明。

一、問題描述

雙向細目表示測驗的架構藍圖，描述一份測驗中所應該包含的內容以及所評量到的能力，也是命題的依據。它是以教學目標和學習內容為兩個軸，分別說明各項評量目標，建立雙向細目表可以幫助命題者釐清教學目標和學習內容的關係，以確保測驗能反映教材的內容，並能夠評量到預期之學習結果。本研究以參加生產管理科測驗之考生為研究對象，考生人數為 4,544 人，由於樣本具備代表性，因此，可以使用難易度 P 作為試題品質判斷之指標。此外， P 介於 0~1 之間，且期望值為 .5，若 P 符合常態性，則可將 P 值轉換成標準常態 Z 值之管制圖，進行試題品質之評估。表 2 生產管理科的雙向細目表中，每一個試題係依照七個單元能力類別作分類。

二、資料處理

表 3 係根據表 2 的各單元能力作考生難易度的整理，其中 P 為所有考生的試題難易度， Z_p 則表示為將難易度轉換成標準常態後的 Z 值， Z 值則可透過累積標準常態表得之。

三、試題資料之常態性、隨機性及獨立性檢定

難易度指標 P 的數據資料經過常態分配檢定後，由圖 4 可知 P 值為 $0.065 > \alpha = 0.05$ ，因此不拒絕 H_0 ，故資料可合理地假設服從常態分配。本研究以參加生產管理科測驗之考生為研究對象，考生人數為 4,544 人，屬於大樣本，而每一題試題為 P 值所有考生測驗的結果，因此每一題的 P 值均服從常態分配，同理，七個單元係由數題試題所組成，應服從常態分配。

表 2 生產管理科之雙向細目表

測驗內容	認知層次			
	知識	理解	應用	綜合分析
1 導論		1,2,3,4,5,6,7,8	10	9
2 預測		11,12,14,15,16,17,19,20		
3 產能規劃		13,18		
4 製程選擇	22	23	21,24,25	
5 設施佈置		26,27,28,29,30,31,33		
6 存貨管理		32,36	34,35,37,38	39,40
7 品質管制				41,42,43,44,45,46,47,48,49,50

表 3 生產管理科各單元能力之難易度

導論			預測			產能規劃			製程選擇			設施佈置			存貨管理			品質管制		
題項	P	Z _P	題項	P	Z _P	題項	P	Z _P	題項	P	Z _P	題項	P	Z _P	題項	P	Z _P	題項	P	Z _P
1	0.27	-0.6127	11	0.43	-0.1764	13	0.3	-0.5244	21	0.41	-0.2274	26	0.28	-0.5829	32	0.42	-0.2018	41	0.31	-0.4958
2	0.2	-0.8418	12	0.35	-0.3854	18	0.32	-0.4678	22	0.48	-0.0503	27	0.29	-0.5534	34	0.2	-0.8418	42	0.26	-0.6433
3	0.38	-0.3055	14	0.3	-0.5244				23	0.62	0.3055	28	0.3	-0.5244	35	0.34	-0.4124	43	0.39	-0.2792
4	0.34	-0.4124	15	0.24	-0.7065				24	0.55	0.1256	29	0.28	-0.5829	36	0.24	-0.7065	44	0.32	-0.4678
5	0.21	-0.8066	16	0.31	-0.4958				25	0.5	0	30	0.43	-0.1764	37	0.44	-0.1510	45	0.37	-0.3318
6	0.44	-0.1510	17	0.26	-0.6433							31	0.26	-0.6433	38	0.39	-0.2792	46	0.45	-0.1256
7	0.26	-0.6433	19	0.31	-0.4958							33	0.35	-0.3854	39	0.29	-0.5534	47	0.29	-0.5534
8	0.24	-0.7065	20	0.32	-0.4678										40	0.45	-0.1256	48	0.14	-1.0805
9	0.21	-0.8066																49	0.41	-0.2274
10	0.3	-0.5244																50	0.24	-0.7065

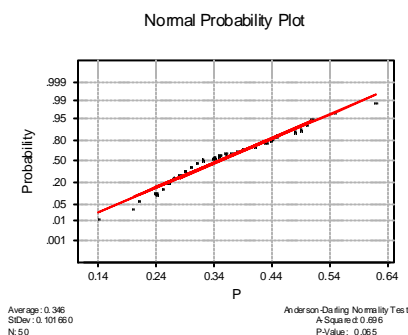


圖 4 常態機率圖

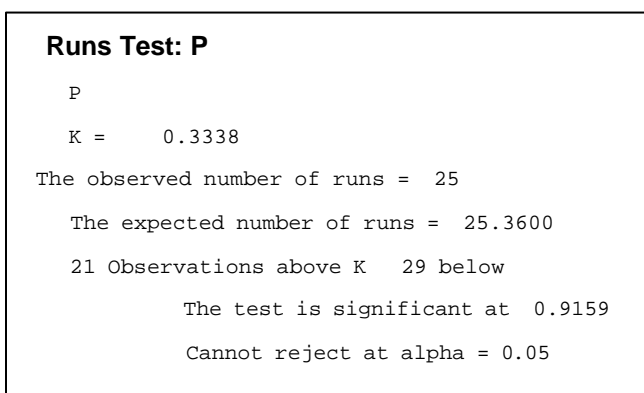


圖 5 連串檢定

在隨機性檢定方面，使用連串檢定，圖 5 顯示顯著水準為 0.9159，因此不拒絕 H_0 的假設，意指資料可合理地假設資料具有隨機性。

圖 6 為自我相關函數圖，顯示觀察值並沒有嚴重的自我相關，資料符合獨立性。綜合上述結果，可知研究資料符合常態性、隨機性及獨立性，換言之，此一指標之數據資料符合管制圖的基本假設。因此，本研究所探討的資料可將管制圖作為試題品質分析的工具。

四、試題品質之管制圖

本文章除了針對生產管理科五十題試題的資料符合管制圖的基本假設，以進行試題品質分析，為了進一步瞭解個單元之試題品質，將整份試題依據雙向細目表分成七個單元進行品質評鑑，然而本研究所欲探討的各單元分類後題數減少，雖然每個單元

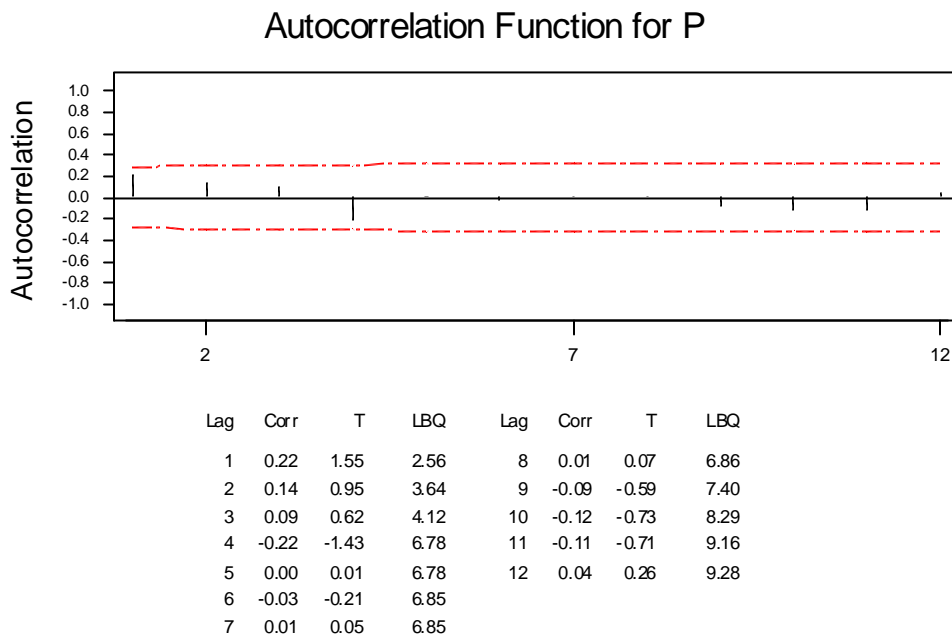


圖6 自我相關函數圖

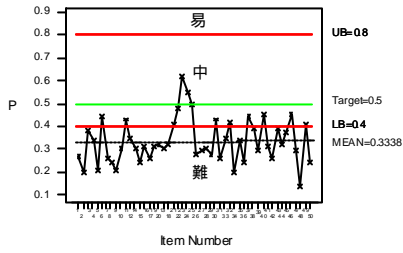
均符合常態性，但是存在著每個單元題數較少的問題，使得資料呈現非隨機性及非獨立性的情形。然而，由於測驗係鎖定能力來命題，才能引導教學以能力為導向，讓學生真正學習到應該具備的能力，以單元能力角度做試題品質評估分析有其重要性，基於此理由，本研究將各個單元假設為符合隨機性及獨立性，此乃為本研究之限制，主要目的為運用管制圖的概念評估試題各單元之品質，以進一步瞭解各單元試題之品質情形。試題品質之管制圖分成整份試卷之試題品質管制圖及各單元能力之試題品質管制圖兩部分進行探討。

(一) 整份試卷之試題品質管制圖

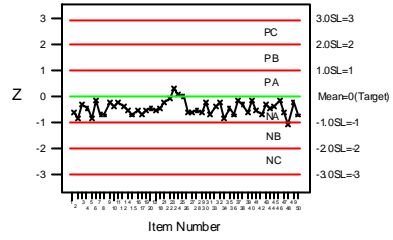
圖 7a 及圖 7b 為整份試題繪製後之 P (難易度)、難易度 Z 值之 Z_p 管制圖。

(二) 各單元之試題品質管制圖

圖 8a 至圖 8n 係依照表 2 所繪製而成的七個單元能力 P 及 Z_p 管制圖。

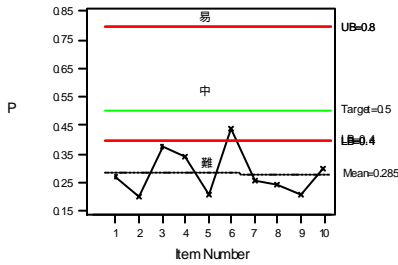


(a) 整份試題之 P 管制圖

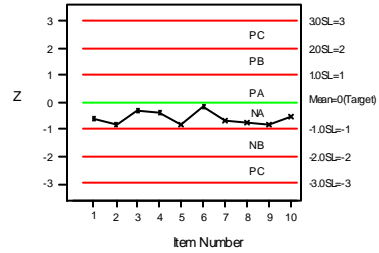


(b) 整份試題之 Z_p 管制圖

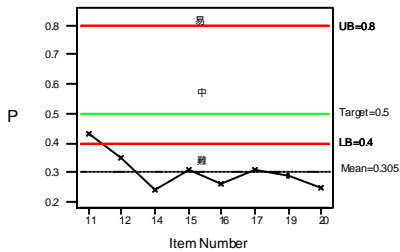
圖 7 整份試題之 P 及 Z_p 管制圖



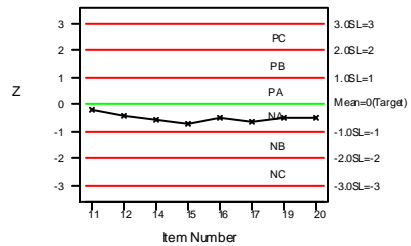
(a) 導論單元 P 管制圖



(b) 導論單元 Z_p 管制圖

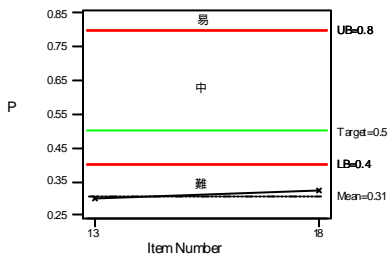


(c) 預測單元之 P 管制圖

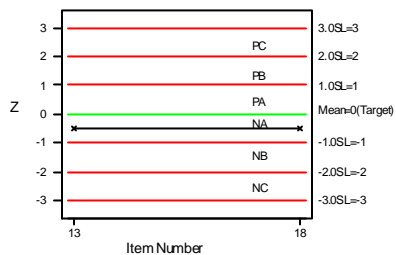


(d) 預測單元之 Z_p 管制圖

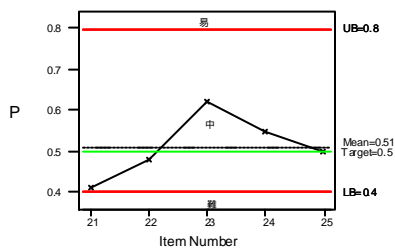
圖 8 各單元之 P 及 Z_p 管制圖(續)



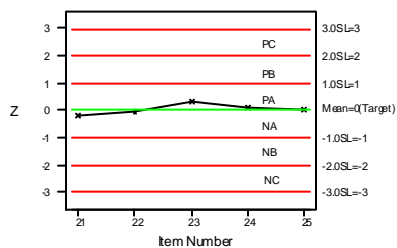
(e) 產能規劃單元之 P 管制圖



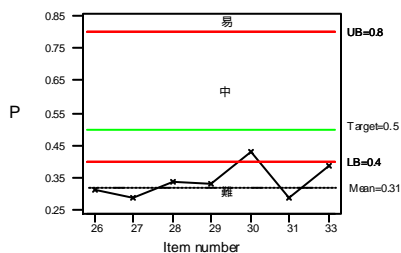
(f) 產能規劃單元之 Z_p 管制圖



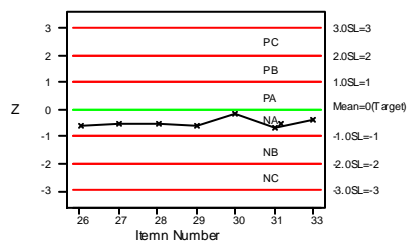
(g) 製程選擇單元之 P 管制圖



(h) 製程選擇單元之 Z_p 管制圖

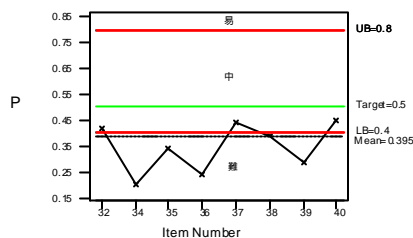


(i) 設施佈置單元之 P 管制圖

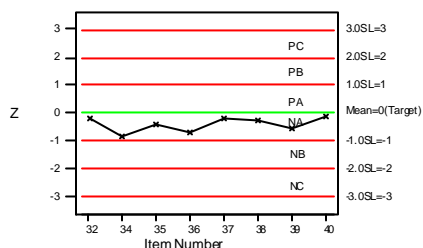
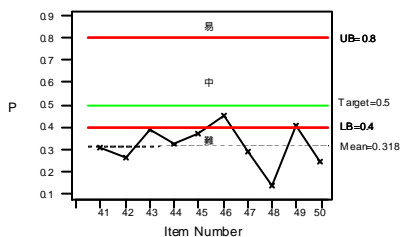


(j) 設施佈置單元之 Z_p 管制圖

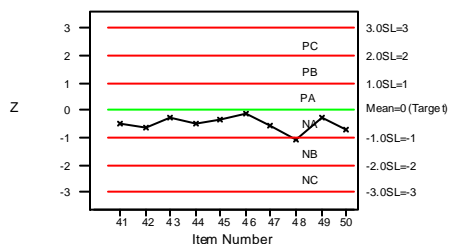
圖 8 各單元之 P 及 Z_p 管制圖(續)



(k) 存貨管理單元之 P 管制圖

(l) 存貨管理單元之 Z_p 管制圖

(m) 品質管制單元之 P 管制圖

(n) 品質管制單元之 Z_p 管制圖圖 8 各單元之 P 及 Z_p 管制圖

將 P 管制圖及 Z_p 管制圖的分析結果，整理如下幾點：

1. 以一般測驗領域中常用的品質判斷標準之 P (難易度) 管制圖來說，整份試題品質共有 13 題位於「中」之品質 (圖 7a)，包括題項 6、11、21、22、23、24、25、30、32、37、40、46、49，其餘 37 題均位於「難」之品質，且未有屬於「易」之品質，顯示整份試題大多呈現難的情形。在七個單元能力方面，除了製程選擇單元(圖 8i)，難易度屬於「中」之品質，其餘六個單元能力之難易度幾乎都位於「難」之品質。
2. 以 Z_p 管制圖來說，整份試題品質共有 3 題位於 PA 區，包括題項 23、24、25，其餘試題均位於 NA 區，顯示整份試題品質跨越整個 A 區 (PA 區、NA 區)，表示試題品質亦屬佳。在七個單元能力方面，除製程能力與品質製程能力兩個單元外，其餘

五個單元能力之試題品質均位於 NA 區（難易度低於 $P=0.5$ ），難易度較 PA 區來得高，未超過 1σ ，顯示試題品質尚屬佳。

3. 根據第 2 點，由圖 8h 可知，製程選擇單元能力跨越整個 A 區（PA 區、NA 區），表示試題品質亦屬佳。至於品質管制單元能力方面（圖 8n），由於第 48 題位於 NB 區，此題評鑑為難，但整體單元之品質，則需要透過信賴區間檢定的結果，進一步地評估整體單元能力之試題品質。

表 4 為生產管理科各單元能力難易度之 Z 檢定結果，在 P 情形中，試題品質評鑑指標之優良難易度區間為（0.4, 0.8），表示優良難度之區間介於 0.4 至 0.8 間，且最佳品質試題為 $P=0.5$ ，由表 4 可知整份試題不論 P 情形或 Z_p 情形，95% 信賴區間均未包含最佳之難易度（ $P=0.5$ 及 $Z=0$ ），意指整份試題之試題品質情形不佳。另外，從各單元能力方面來看，製程選擇單元之 95% 信賴區間為（0.4776, 0.5939），此信賴區間不僅靠近最佳之難易度 $P=0.5$ ，而且也符合 Aiken（2000）建議的難易度範圍 0.4 至 0.6。因此，將此單元能力（製程選擇單元）評估之為品質優良之試題。而在 Z_p 情形中，製程選擇單元能力之 95% 信賴區間為（-0.1438, 0.2052），包含了最佳之難易度（ $Z=0$ ，表示難易度為 0.5），並且位於試題品質良好的 A 區，故此單元能力亦評鑑優良之品質優良之試題。換言之，在兩種品質評鑑情形下，製程選擇單元均擁有優良品質之試題。

整體來說，本研究以 3σ 為管制界限、 1σ 為警告界限的 Z_p 管制圖，不僅擁有與 P 情形的品質評鑑能力，並且可以更清楚、明確地指出各單元試題品質位於哪一個等級，所以，此管制圖為一個良好的試題評鑑方法。

伍、研究結果對試題品質的改善

在研究結果對試題品質的改善方面，將分別針對 P 情形及 Z_p 情形做說明。以 P 情形來看，整份試題及各單元能力的難易度大多為偏難的情形，並未出現容易的試題，表示不管何種能力的受試者，都很少答對大部分的單元，因此，必須加以診斷與修改，才能對於試題品質有所改善。從 Z_p 管制圖來看，整份試題及各單元能力之品質多位於 A 區，且未有位於與超過 3σ 的單元，表示本研究所分析的試題資料品質程度均屬良好，並無任何一個單元能力的試題品質呈現不佳的情形。

從 Z_p 管制圖即可知道整份試卷及單元中的試題位於最佳品質 $Z=0$ （ $P=0.5$ ）的上

表4 生產管理科各單元能力難易度之t檢定結果

	N	Mean		StDev		SE Mean		95% CI (信賴區間)	
		P	Z _P	P	Z _P	P	Z _P	P	Z _P
整份試題	50	0.3338	-0.4281	0.0967	0.2667	0.0137	0.0377	(0.3063, 0.3613)	(-0.5039, -0.3524)
1. 導論	10	0.2850	-0.5811	0.0800	0.2316	0.0253	0.0732	(0.2354, 0.3346)	(0.7246, -0.4375)
2. 預測	8	0.3050	-0.4869	0.0623	0.1609	0.0220	0.0569	(0.2618, 0.3482)	(-0.5984, -0.3754)
3. 產能規劃	2	0.3100	-0.4961	0.0141	0.0400	0.0100	0.0283	(0.29046, 0.32954)	(-0.5515, -0.4407)
單元能力 4. 製程選擇	5	0.5120	0.0307	0.0785	0.1991	0.0351	0.0891	(0.4776, 0.5939)	(-0.1438, 0.2052)
5. 設施佈置	7	0.3129	-0.4927	0.0588	0.1608	0.0222	0.0608	(0.2693, 0.3564)	(-0.6118, -0.3736)
6. 存貨管理	8	0.3950	-0.4090	0.0981	0.2682	0.0347	0.0948	(0.2783, 0.4142)	(-0.5948, -0.2231)
7. 品質管制	10	0.3180	-0.4911	0.0917	0.2779	0.0290	0.0879	(0.2652, 0.3788)	(-0.6646, -0.3176)

面區域(評鑑符號為P)或下面的區域(評鑑符號為N),以了解此試題屬於較簡單或較困難的品質。一般而言,影響試題品質異常的因素為試卷(如是否超過考試範圍)、學生(只有針對能力高及能力低的學生作分析)及教師,一旦發現試題品質呈現極難或極易的情形,則可從此三個方面進行探討。

陸、結論

現代測驗理論的試題分析有一個重要的特點,那就是將測驗試題圖形化,其具有容易了解的優點。本研究目的在於從管制圖的品質管制特性,探討考科單元的試題品質。本研究所欲分析的測驗資料為難易度,屬於計量值之品質特性,因此,本研究以

整個能力的觀點加以審視，提出應用管制圖工具之品質管制特性，以及結合相關的統計方法，進而評估生產管理科測驗命題之品質情形。

本文章主要為應用管制圖的圖示化、統計基礎等品質管理特性於單元試題品質的判斷，透過本研究，不僅可以提供試題品質分析的一個方法，更可以讓命題者瞭解到命題時應注意的重點，如以能力為命題核心，以提供如何改善試題品質的建議及作法。

本研究除了將試題品質指標評鑑標準依照一般測驗的評鑑標準進行探討外，也將以 3σ 作為管制圖的管制界限，建立一個試題品質的評鑑標準。研究結果顯示，本文所使用的評鑑標準，不僅能更清楚地了解品質等級，並且可以更明確地指出試題的品質情形。

在入學測驗中，每一個考科的命題方式與本文所提及的原則均相似，故本研究非但可用於生產管理科，也可運用於其他考科，或其他入學能力測驗的考科試題品質分析上。

參考文獻

一、中文部分

1. 李正源(1996)，試題品質的決定 - 測驗與量表的試題分析實務，中等教育，47(4)，44-61。
2. 范德鑫(1990)，試題分析的功能、方法及其使用限制，中等教育，41(1)，68-78。
3. 郭生玉(1990)，心理與教育測驗，第五版，台北：精準。
4. 簡茂發(1991)，教育的心理學基礎，於黃光雄主編，教育概論（頁 127-143），台北：師大書苑。

二、英文部分

1. Aiken, L. R. (2000). Psychological testing and assessment. Boston: Allyn & Bacon.
2. Bloom, B. S., Englehart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. White Plains, NY: Longman.

3. Brown, H. D. (1994). Principles of language learning and teaching (3rd ed.). Englewood Cliffs, NJ: Prentice Hall Regents.
4. Chen, Y. K., & Liao, H. C. (2004). Multi-criteria design of an \bar{X} control chart. Computers & Industrial Engineering, 46, 877-891.
5. Fan, C. T. (1952). Item analysis table. Princeton, N.J.: Educational Testing Service.
6. Gronlund, N. E. (1976). Measurement and evaluation in teaching (3rd ed.). New York: Macmillan.
7. Kerlinger, F N., & Lee, H B. (2000). Foundations of behavioral research (4th ed.). Orlando, FL: Harcourt College Publishers.
8. Maller, S. J. (1997). Deafness and WISC- item difficulty: invariance and fit. Journal of School Psychology, 35(3), 299-314.
9. Mehrens, W. A., & Lehmann, I. J. (1984). Measurement and evaluation in education and psychology (3rd ed.). New York: Holt, Rinehart, and Winston.
10. Montgomery, D. C. (2001). Statistical quality control (4th ed.). John Wiley and Sons.
11. Osterlind, S. J. (1989). Constructing test items. Boston: Kluwer Academic Publisher.
12. Reynolds, C. R., Livingston, R. B., & Willson, V. (2006). Measurement and assessment in education. Boston: Pearson Education.
13. Wainer, H. (1989). The future of item analysis. Journal of Educational Measurement, 26(2), 191-208.
14. Yang, J. H., & Yang, M. S. (2005). A control chart pattern recognition system using a statistical correlation coefficient method. Computers & Industrial Engineering, 48, 205-221.

2006年03月15日收稿

2006年05月15日初審

2006年06月15日複審

2006年06月30日接受