

試題反應理論的介紹(二)

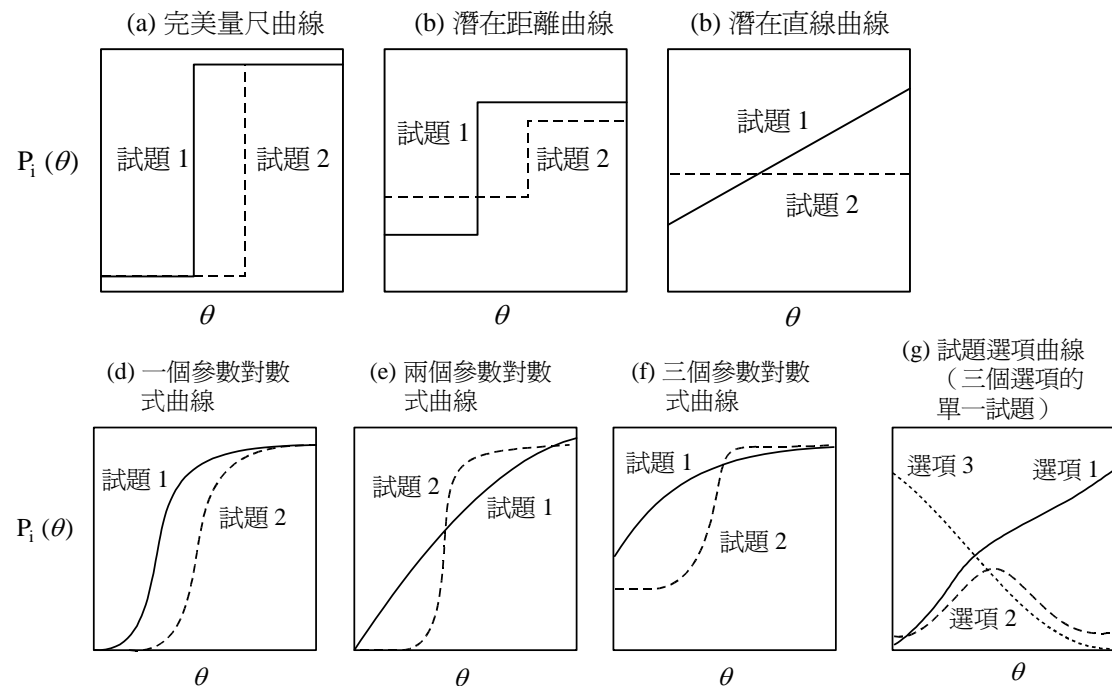
——基本概念和假設

余民寧 著

基本概念

試題反應理論(item response theory)建立在兩個基本概念上：(1)考生(examinee)在某一測驗試題上的表現情形，可由一組因素來加以預測或解釋，這組因素叫作潛在特質(latent traits)或能力(abilities)；(2)考生的表現情形與這組潛在特質間的關係，可透過一條連續性遞增的函數來加以詮釋，這個函數便叫作試題特徵曲線(item characteristic curve, 簡寫為 ICC)。其實，我們把能力不同的考生得分點連接起來所構成的曲線，便是能力不同的考生在某一測驗試題上的試題特徵曲線，把各試題的試題特徵曲線加總起來，便構成所謂的試卷特徵曲線(test characteristic curve, 簡寫為 TCC)。因此，試題特徵曲線即是一條試題得分對能力因素所作的迴歸線，這條迴歸線在基本上是非直線的，但直線的試題特徵曲線也是有可能的，端視所選用的試題反應模式(item response model)而定。

試題特徵曲線所表示的涵義，即是某種潛在特質的程度與其在某一試題上正確反應的機率，二者之間的關係；這種潛在特質的程度愈高（或愈強），其在某一試題上的正確反應機率便愈大。在試題反應理論中，每一種試題反應模式就有其相對應的一條試題特徵曲線，此一曲線通常包含一個或多個參數來描述試題的特性，以及一個或多個參數來描述考生的潛在特質；因此，所選用的試題反應模式所具有的參數個數及其數值的不同，所畫出的試題特徵曲線形狀便不同。常見的曲線形狀，如圖一所示。



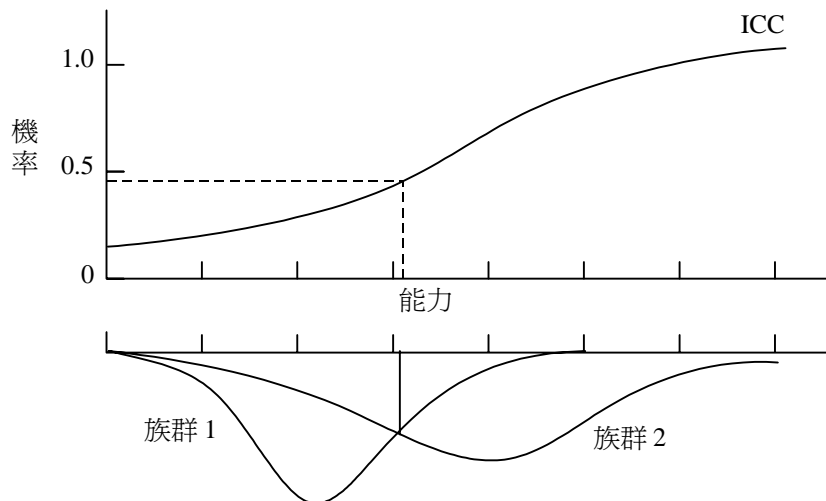
圖一 七個不同的試題特徵曲線例子 (資料來源：Hambleton & Cook, 1977)

圖一所示的七種曲線中， θ 表示考生或受試者的能力或潛在特質， $P_i(\theta)$ 則表

示能力或潛在特質為 θ 的考生或受試者，其答對或正確反應某一試題的機率。例(a)所示，代表 Guttman 的完美量尺模式(perfect scale model)，它是一個階段式函數(step functions)所形成的曲線，表示在某一關鍵值 θ^* 以右的機率為 1，以左的機率為 0；換句話說，這種模式具有完美的鑑別能力，而 θ^* 即為區別出有能力組和無能力組的關鍵值。例(b)為例(a)的一種變形，叫作潛在距離模式(latent distance model)，為社會心理學者常用來測量人的態度的一種模式，其正確與不正確反應的機率，在 0 與 1 之間變動不已。例(c)所示即為古典測驗理論下的試題特徵曲線，截距的大小反映出試題的難度，而曲線的斜率即代表試題的鑑別度，在所考慮的條件相等的情況下，正確反應的機率與 θ 值成正比。例(d)到例(f)所示，即代表試題反應理論中的一個參數、兩個參數、與三個參數的對數式模式(logistic model)，它們的涵義分別代表著：某一試題的正確反應機率除了受考生或受試者的 θ 值所決定外，並且分別受一個參數（即難度）、二個參數(即難度和鑑別度)、或三個參數（即難度、鑑別度、和猜測度）的試題參數所決定，其正確反應的機率值亦介於 0 與 1 之間。例(g)所示為特殊的試題反應模式，如：Samejima (1969)的等級反應模式(graded response model)，Masters (1982)和 Yu (1991)的部份計分模式(partial credit model)等，即是採用試題選項特徵曲線(item option characteristic curves)，所代表的意思是試題中每一選項被選中的機率，它也是能力或潛在特質的一種函數，它有個基本假設，亦即就某一固定能力的考生而言，他／她在同一試題上所有的試題選項特徵曲線的總和為 1。

試題反應模式不像古典的真實分數模式，它是可能作假的模式(falsifiable models)；換句話說，任何一種試題反應模式都有可能適用或不適用於某份特殊的測驗資料，亦即模式可能會不當地預測或解釋資料。因此，在應用試題反應理論時，我們必須先估計出模式與考生的參數值外，還需要考驗模式與資料間的適合度(model-data fit)。這兩者留待後文補充說明。

當某一種試題反應模式適用於某種測驗資料時，一些試題反應理論的基本特性也會跟著產生。首先，從不同組的試題估計而得的考生能力估計值，除了測量誤差外，不會受所使用的測驗種類的影響，亦即，它是試題獨立(item-independent)的能力估計值；其次，從不同族群的考生估計而得的試題參數估計值，除了測量誤差外，亦不受參與測驗的考生族群的影響，亦即，它是樣本獨立(sample-independent)的試題參數估計值。上述兩種特性，在試題反應理論中叫作「不變性」(invariant)，這些不變性是從把試題的訊息(information)考慮在能力估計的過程中，把考生能力的訊息考慮在試題參數估計的過程中而得。典型的試題參數不變性例子，如圖二所示。在圖二中，不管考生所來自的族群為何，只要他們具有相同的能力，他們答對（或正確反應）某一試題的機率便相同；由於某特定能力的考生答對某一試題的機率是由試題參數所決定，試題參數對這兩族群的考生而言也必定相同。



圖二 試題特徵曲線與兩族群考生的能力分配曲線

除了上述的特性外，試題反應理論還可以針對個別的（亦即每一位能力不同的考生或受試者）能力估計值，提供其測量的估計標準誤(standard errors)，這點作法不同於古典測驗理論僅提供所有考生單一的誤差估計值的作法。此外，試題反應理論把能力測量的估計標準誤之平方的倒數，定義為試題的訊息函數(item information function)，它可以用來作為評量能力估計值之精確度的指標，大有取代古典測驗理論中「信度」(reliability)指標之勢(Wright & Masters, 1982; Wright & Stone, 1979)。這些作法及概念續待後文補充說明。

基本假設

任何一條試題特徵曲線所代表的涵義是：答對某一試題的機率，是由考生的能力和試題的特性所共同決定。因此，試題反應理論具有下列幾項基本假設，唯有在這些假設都成立的前提下，試題反應模式才能被用來分析所有的測驗資料。

(一)單向度(unidimensionality)：試題反應理論中的各種模式有個最常用的共同假設，那就是測驗中的各個試題都測量到同一種共同的能力或潛在特質；這種單一能力或潛在特質（因素）必須包含在測驗試題裡的假設，便是單向度的假設。

其實，在實際的測驗情境裡，考生在測驗上的表現情形很少是純粹受到一種因素的影響，其他因素如：成就動機、考試焦慮、應試技巧、及人格特質等，也都會影響到測驗的結果；因此，試題反應理論中對測驗必須具有單向度因素的基本看法，認為只要該測驗具有能夠影響測驗結果的一個「主要成份或因素」(dominant component or factor)，便算符合單向度假設的基本要求，而這個主要因素所指的，即是該測驗所測量到的單一能力或潛在特質。

適用於含有單一主要因素測驗資料的試題反應模式，便稱作單向度模式。適用於含有多種主要因素的試題反應模式，便叫作多向度(multidimensional)模式。多向度模式的數學公式複雜難懂，而且模式也還在發展中，本系列文章不擬介紹，有興趣的讀者可參閱 McDonald (1981)和 Ackerman (1989)的文章。

(二)局部獨立性(local independence)：它的涵義是說，當影響測驗表現的能力被

固定不變時，考生在任何一對試題上的反應，在統計學上而言是獨立的；換句話說，在考慮考生的能力因素後，考生在不同試題上的反應間沒有任何關係存在。簡單地說，這意謂著涵蓋在試題反應模式裡的能力因素，才是唯一影響考生在測驗試題上做反應的因素；這組能力因素代表整個潛在空間(complete latent space)，當單向度基本假設成立時，這整個潛在空間僅包含一種能力因素。

假設 θ 為能力因素， U_i 代表某位考生在第 i 試題上的反應， $P(U_i | \theta)$ 代表具有能力為 θ 的考生在第 i 試題上的反應機率，且 $P(U_i = 1 | \theta)$ 為正確反應的機率， $P(U_i = 0 | \theta)$ 為錯誤反應的機率，那麼，局部獨立性的涵義即是：

$$P(U_1, U_2, \dots, U_n | \theta) = P(U_1 | \theta) P(U_2 | \theta) \cdots P(U_n | \theta) \\ = \prod_{i=1}^n P(U_i | \theta)$$

這條公式即是說明，對某一特定能力的考生而言，他／她在某份測驗上的反應組型(response pattern)的機率，等於他／她在單獨一題試題上反應機率的連乘積。例如，某位考生在一組三個試題測驗的反應組型為 (1, 1, 0)，其中 $U_1 = 1, U_2 = 1, U_3 = 0$ ，那麼，局部獨立性所要表達的意思即為：

$$P(U_1 = 1, U_2 = 1, U_3 = 0 | \theta) = P_1(U_1 = 1 | \theta) P_2(U_2 = 1 | \theta) P(U_3 = 0 | \theta) \\ = P_1 P_2 Q_3$$

其中，

$$P_i = P(U_i = 1 | \theta) \text{ 且 } Q_i = 1 - P_i$$

由於 $P(U_i | \theta)$ 是一種條件機率(conditional probabilities)的表達方式，因此，局部獨立性假設又稱為條件獨立性(conditional independence)假設。

通常，當單向度假設獲得成立時，局部獨立性假設也會獲得成立，就這一項涵義而言，這兩個概念是相通的(Lord, 1980; Lord & Novick, 1968)，甚至於，即使資料不是單向度的，局部獨立性也可以獲得成立。只要整個潛在空間被界定清楚，亦即當所有影響表現的能力向度都考慮之後，局部獨立性便可獲得成立。局部獨立性在下列情況下無法成立：影響測驗表現的能力向度不只一種時，連鎖性試題，以及試題本身提供作答的線索等，在這種情況下，試題反應模式也就無法適用於該筆測驗資料。

(三)非速度測驗：試題反應模式所適用的情況有個隱含的基本假設，那就是測驗的實施不是在速度限制下完成的；換句話說，考生的考試成績不理想，是由於能力不足所引起，而不是由於時間不夠答完所有試題所致。由於這項假設是隱含在單向度假設裡，所以不常被試題反應理論學者所提起，但是在選用試題反應模式時，這項基本假設亦必須要考慮到才行。

(四)知道——正確假設(know--correct assumption)：如果考生知道某一試題的正確答案，他／她必然會答對該試題；換句話說，如果他／她答錯某一試題，他／她必然不知道該試題的答案。當然，把正確答案填錯在別的格子上以致整個試卷都錯的例子，不在本假設所考慮的範圍內，因為人為的疏忽不是任何測驗理論所能顧及到的。此外，省略不答的試題(omitted items)和未答完的試題(unreached items)有所不同，前者是受能力影響所致，後者是受施測速度影響所致。本假設僅能適用於前

者，它和前個假設一樣，都隱含在單向度假設裡，故殊少被提及。

參考文獻

- Ackerman, T. A. (1989). Unidimensional IRT calibration of compensatory and noncompensatory multidimensional items. Applied Psychological Measurement, 13, 113-127.
- Hambleton, R. K., & Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. Journal of Educational Measurement, 14, 75-96.
- Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston, MA: Kluwer-Nijhoff.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: SAGE.
- McDonald, R. P. (1981). The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology, 34, 100-117.
- Masters, G. N. (1982). A Rasch model for partial credit model. Psychometrika, 47, 149-174.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometric Monograph, No. 17.
- Wright, B. D., & Masters, G. N. (1982). Rating scale analysis. Chicago: MESA Press.
- Wright, B. D., & Stone, M. H. (1979). Best test design. Chicago: MESA Press.
- Yu, M. (1991). A two-parameter partial credit model. Unpublished doctoral dissertation of University of Illinois at Urbana-Champaign.