

試題反應理論的介紹(六)

——能力量尺

英譯：(The ability scale)

余民寧 著

英譯名：(Yu Min-Ning)

測驗的最終目的往往在於給考生打分數，該分數便代表考生習得（或熟習）某種技能的程度。因此，謹慎地打分數以及小心翼翼地解釋測驗分數，便成為教育或心理測量中一個重要的課題。

在古典測驗理論裡，考生在某份測驗上答對（或正確反應）的題數和，即代表他在該測驗上的真實分數(true score)的不偏估計值(unbiased estimate)，亦即是他在該測驗所測量之技能上的表現程度。而在試題反應理論裡，考生的能力並不是由答對的題數和來表示，而是必須估計的；亦即經由某種適當模式以及考生的反應組型(response pattern)，來估計出考生應有的能力估計值。

一般說來，決定考生能力估計值大小的常用步驟，如下所述：

- (1)先取得考生在一組試題上的反應資料，答對者給 1 分，答錯者給 0 分。
- (2)若試題參數（如：難度、鑑別度、及猜測度）為已知時，可用前文所提到的最大近似值估計法(maximum likelihood estimation)來進行估計考生的能力。
- (3)若試題參數為未知時，則試題與能力參數就必須一起同時進行估計，此時亦可用最大近似值估計法來估計。
- (4)再將估計好的能力估計值，經由直線或非直線轉換，以便換算成較為一般大眾所熟悉的量尺分數，增加解釋測驗分數的便利；例如：TOEFL 分數便是一例。

經由上述步驟，我們可以獲得考生的能力分數，有助於我們解釋考生在某份測驗上的表現好壞。然而，在考慮解釋能力分數的意義之前，能力分數的本質是什麼？什麼樣的分數轉換才有意義？該用何種量尺來表示？我們也必須要有所瞭解才行，方不致誤用或濫用了試題反應理論的優點。

能力量尺的本質

前面說過，答對題數和分數 X 是真實分數 τ 的不偏估計值，亦即是 $E(X) = \tau$ （公式一）

答對題數和分數 X 除以總題數（亦即經過直線轉換），即可獲得一個比例正確分數(proportion-correct score)。當某測驗包含許多分測驗，且每一個分測驗都包含不同的題數，測量到不同的目標時，使用比例正確分數則較具有意義與適當性。這種情形通常出現在效標參照測驗(criterion-referenced test)裡，而在常模參照測驗(norm-referenced test)裡，往往需要使用別種直線轉換，才能獲得所需的標準分數。當然，答對題數和分數 X 也可以透過非直線的轉換，以換算成標準九分或百分位數等分數，以便於進行比較考生在某測驗上的表現差異情形。

然而，分數 X 有個天生的缺點，那就是它不是一種試題獨立的分數，因此，經過轉換過的分數也不是一種考生族群獨立的分數，亦即它會受到不同試題與不同

考生的反應的影響。另一方面，能力分數 θ 卻具有試題獨立與樣本獨立等不變性的特質（請參考前文關於試題反應理論的基本概念之說明），它與 X 分數不同。所以，我們可以使用 θ 分數來比較回答不同試題的考生能力，而 θ 分數所用的量尺 (scale) 也可以被看成是測量特質或能力的絕對量尺。

其實，測驗試題所欲測量的特質或能力，可以被廣義的定義成態度或成就、一種狹義的成就變項（如：四則運算的能力）、或一種人格變項（如：自我概念、成就動機等），它們未必是天生的、或是一成不變的。事實上，能力或特質一詞被看成是考生的一種固定特徵時，多少都有一點不恰當或誤導的意味在裡頭。在許多情境裡，使用精熟程度 (proficiency level) 一詞，也許會比較恰當些。

另者，定義 θ 分數所用的量尺，又具有什麼樣的本質呢？很明顯的，觀察分數 X 不是定義在比率量尺 (ratio scale) 上，也不是定義在等距量尺 (interval scale) 上，充其量，它最多僅被定義在次序量尺 (ordinal scale) 上。同樣的， θ 分數亦被定義在次序量尺上。然而在某些情境中， θ 量尺被作為有限的比率量尺的解釋，也是有可能的。

θ 量尺的轉換

轉換可分成兩種：直線轉換和非直線轉換。轉換的目的在於使測驗分數的解釋和涵義的瞭解，能廣被一般大眾所接納。以下就以這兩種轉換來說明 θ 量尺的涵義。

讀者們可還記得：在試題反應理論裡，正確反應的機率是以試題反應函數 $P(\theta)$ 來表示。以二個參數模式為例：

$$P(\theta) = e^{a(\theta-b)} / [1 + e^{a(\theta-b)}] \quad (\text{公式二})$$

若將 θ 、 b 、及 a 加以轉換成： $\theta^* = \alpha\theta + \beta$ ， $b^* = \alpha b + \beta$ ，和 $a^* = a / \alpha$ ，則

$$P(\theta^*) = P(\theta) \quad (\text{公式三})$$

亦即經過直線轉換後，一個正確反應的機率不會改變；它意謂著，只要試題參數也經過適當的轉換，我們便可針對 θ 量尺進行直線轉換，而仍不改變其正確反應的機率值。

例如，Woodcock(1978) 的心理教育測驗庫 (Woodcock-Johnson Psycho-Educational Battery) 所用的量尺，便是以一個參數模式求得之 θ 值，經轉換成以 9 為底的對數量尺：

$$W_\theta = 20 \log_9(e^\theta) + 500 \quad (\text{公式三})$$

或

$$W_\theta = 9.1\theta + 500 \quad (\text{公式四})$$

因此，它是一種直線量尺 (linear scale)。同理，試題難度也可以轉換成

$$W_b = 9.1b + 500 \quad (\text{公式五})$$

而 W_θ 量尺的特性之一便是 $(W_\theta - W_b) = 20, 10, 0, -10, -20$ 的差值，其正確反應的機率分別剛好是 .90, .75, .50, .25 和 .10。Wright(1977) 曾把這個量尺修改成

$$W = 9.1\theta + 100 \quad (\text{公式六})$$

並把它叫作「智慧」量尺(WITs scale)。

有時候在某些情境裡，進行非直線轉換(nonlinear transformation)反而有助於我們對參數的估算和解釋。茲以一個參數模式為例，說明如下：

$$P(\theta) = e^{(\theta-b)} / [1 + e^{(\theta-b)}] \quad (\text{公式七})$$

如果我們把 θ 和 b 值經由非直線方式轉換成新的 θ^* 和 b^* 值如下：

$$\theta^* = e^\theta, b^* = e^b \quad (\text{公式八})$$

則公式七可以轉變成

$$\begin{aligned} P(\theta) &= \frac{e^{\theta-b}}{1 + e^{\theta-b}} \times \frac{e^b}{e^b} \\ &= \frac{e^\theta}{e^b + e^\theta} \\ &= \frac{\theta^*}{b^* + \theta^*} \quad (\text{公式九}) \end{aligned}$$

因此，

$$P(\theta^*) = \theta^* / (b^* + \theta^*) \quad (\text{公式十})$$

公式十即是 Rasch 模式(1960)對成功的機率所下的原始定義。

正確反應機率既如上述定義在 θ^* 量尺上的 $P(\theta^*)$ 所示，不正確反應機率則為 $Q(\theta^*) = 1 - P(\theta^*)$ ，亦即是

$$Q(\theta^*) = b^* / (\theta^* + b^*) \quad (\text{公式十一})$$

因此，成功的勝算(odds for success) O 可以定義成

$$\begin{aligned} O &= P(\theta^*) / Q(\theta^*) \\ &= \theta^* / b^* \quad (\text{公式十二}) \end{aligned}$$

假設有兩位考生在某一試題上的能力各為 θ_1^* 和 θ_2^* ，且其成功的勝算各為 O_1 和 O_2 ，則他們的成功的勝算比為

$$\frac{O_1}{O_2} = \frac{\theta_1^* / b^*}{\theta_2^* / b^*} = \frac{\theta_1^*}{\theta_2^*} \quad (\text{公式十三})$$

公式十三意謂著，在 θ^* 量尺上，若某考生的能力是另一考生能力的兩倍，則他答對某一試題的機率也是另一考生的兩倍。同理，若同一考生在兩題不同難度值（如 b_1^* 和 b_2^* ）的試題上成功的勝算各為 O_1 和 O_2 ，則該考生答對該二試題的勝算比為

$$\frac{O_1}{O_2} = \frac{\theta^* / b_1^*}{\theta^* / b_2^*} = \frac{b_2^*}{b_1^*} \quad (\text{公式十四})$$

由公式十四可以知道，假設第二題試題的難度是第一題難度的兩倍（如： $b_2^* = 2b_1^*$ ），則該考生答對第一題較簡單的試題的機率是他答對第二題較困難試題的兩倍。

上述 θ^* 和 b^* 量尺所具有的比率量尺的特性，僅適用在一個參數模式裡。關於二個參數和三個參數模式，則量尺又必須另外定義，有興趣的讀者可自行參考 Hambleton & Swaminatlian(1985)的詳細說明。

一個參數模式中另外一種較有意義的非直線轉換，便是採「對數勝算」(log-odds)的轉換。例如，兩位考生對同一試題的成功的勝算比為

$$\frac{O_1}{O_2} = \frac{\theta_1^*}{\theta_2^*} = \frac{e^{\theta_1}}{e^{\theta_2}} = e^{(\theta_1 - \theta_2)} \quad (\text{公式十五})$$

公式十五取自然對數後，則變成

$$\ln\left(\frac{O_1}{O_2}\right) = (\theta_1 - \theta_2) \quad (\text{公式十六})$$

如果兩位考生的能力相差一個單位，即

$$\ln(O_1 / O_2) = 1$$

則

$$O_1 / O_2 = e^1 = 2.718$$

亦即，在能力量尺上相差一個單位，則相當於在 θ 量尺上的成功的勝算相差約 2.72 的量。同樣的道理，如果同一考生回答兩個不同難度的試題，則

$$\ln\left(\frac{O_1}{O_2}\right) = (b_2 - b_1) \quad (\text{公式十七})$$

亦即，在試題難度上相差一個單位，即相當於在成功的勝算上相差約 2.72 的量。

在對數勝算量尺上的單位，即稱作「洛基」(logits)。洛基單位可以由下列程序直接求得，亦即

$$P(\theta) / Q(\theta) = e^{(\theta - b)} \quad (\text{公式十八})$$

則取自然對數後，公式十八的單位即是洛基：

$$\ln\left(\frac{P(\theta)}{Q(\theta)}\right) = \theta - b \quad (\text{公式十九})$$

轉換成真實分數量尺

其實， θ 量尺最主要的轉換用途是將它轉換成真實分數量尺(true-score scale)；因為真實分數量尺的範圍是由 0 到 n ， n 為測驗的題數，而 θ 量尺的範圍卻是介於正負無窮大之間（亦即 $-\infty < \theta < \infty$ ），若將 θ 量尺轉換成真實分數量尺，不僅有助於我們陳報考生的能力高低，更有助於我們解釋測驗分數和作為對換測驗(test equating)之用。

前面曾經說過，真實分數是答對題數和分數之期望值，以數學公式表示如下：

$$\tau = E(X) = E\left(\sum_{j=1}^n U_j\right) \quad (\text{公式二十})$$

其中， τ 為真實分數， X 為答對題數和分數， U_j 代表第 j 個試題上的反應分數（即答對者給 1 分，答錯者給 0 分）， E 代表求期望值的運算符號。若根據期望值的運算方法，公式二十可以展開如下：

$$\begin{aligned} \tau &= E\left(\sum_{j=1}^n U_j\right) = \sum_{j=1}^n E(U_j) \\ &= \sum_{j=1}^n [1 \cdot P_j(\theta) + 0 \cdot Q_j(\theta)] \\ &= \sum_{j=1}^n P_j(\theta) \quad (\text{公式二十一}) \end{aligned}$$

亦即，真實分數即是能力為 θ 的考生在一堆試題上的試題特徵曲線 (item characteristic curves) 之和。由此看來，真實分數其實就是考生在某一測驗上的測驗特徵曲線 (test characteristic curves)，當然，這種說法也僅有在試題反應模式適用於該資料的條件下才成立。

真實分數可以被看成是 θ 的一種非直線轉換，因為 θ 與 τ 間具有一種依序遞增的函數關係。另一種常用的轉換，便是將 τ 轉換成真實比例正確分數 (true proportion correct score) 或內容範圍分數 (domain score) 如下：

$$\pi = \tau / n = \sum_{j=1}^n P_j(\theta) / n \quad (\text{公式二十二})$$

可見 π 的值介於 0 與 1 之間，如同百分比一般介於 0% 到 100% 之間。在一個參數和二個參數模式下， π 的下限值為 0；而在三個參數模式時，由於 θ 趨近於 $-\infty$ ，所以 $P_j(\theta)$ 趨近於最低的漸近線 C_j ，故 π 的下限值為 $\sum C_j / n$ ，與之相對的 τ 的下限值則為 $\sum C_j$ 。

將 θ 轉換成真實分數或內容範圍分數有許多好處：第一，負的分數可以被消除，便利於大眾的理解能力；第二，新量尺的範圍介於 0 與 n 之間（或 0% 到 100% 之間），分數本身即具有解釋涵義在裡頭；第三，內容範圍分數比 θ 量尺更好決定區別精熟與否的切割分數 (cut-off score)，便利於精熟測驗 (mastery testing) 的實施；第四，將真實分數對照其相對應的 θ 值，畫成一個雙向度的分佈圖，有助於判定切割分數的位置。

為了說明起見，茲舉表一的五個試題的基本資料為例，分別計算 $\theta = -3, -2, -1, 0, 1, 2, 3$ 時，三個參數模式的正確反應機率，並合併這些機率值為真實分數，及算出其相對應的內容範圍分數，並畫出內容範圍分數與 θ 值的分佈圖，如表二及圖一所示。

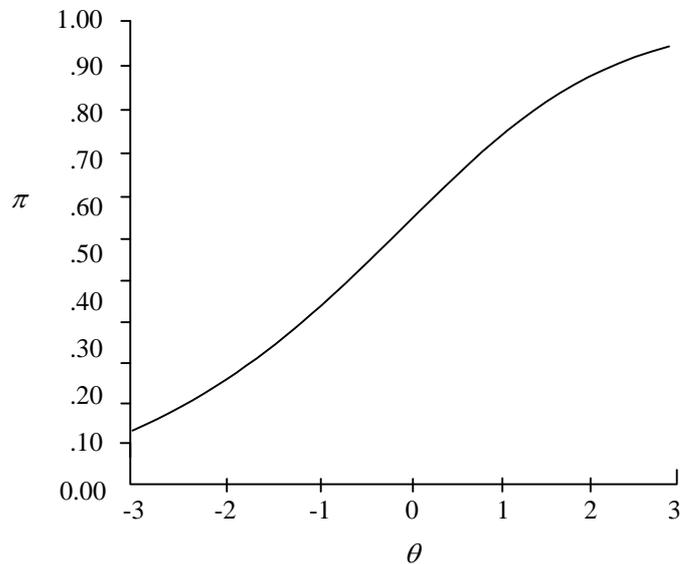
表一 五個試題的試題參數值

試題	難度 (b)	鑑別度 (a)	猜測度 (c)
1	-2.00	0.80	0.00
2	-1.00	1.00	0.00
3	0.00	1.20	0.10
4	1.00	1.50	0.15
5	2.00	2.00	0.20

表二 θ 、 τ 與 π 之間的關係

θ	$P_1(\theta)$	$P_2(\theta)$	$P_3(\theta)$	$P_4(\theta)$	$P_5(\theta)$	$\tau = \sum P_i(\theta)$	$\pi = \tau / n$
-3	.20	.03	.10	.15	.20	.69	.14
-2	.50	.15	.11	.15	.20	1.12	.22
-1	.80	.50	.20	.16	.20	1.85	.37
0	.94	.85	.55	.21	.20	2.75	.55

1	.98	.97	.90	.58	.22	3.65	.73
2	.99	.99	.99	.94	.60	4.51	.90
3	1.00	1.00	1.00	1.00	.96	4.96	.99



圖一 θ 與 π 關係之分佈圖

由圖一可以看出 θ 與 π 之間具有依序遞增的關係 (monotonically increasing relationship)，而與 π 相對應的 θ 量尺上的分數，即可作為判斷精熟與否的切割分數 (*Hambleton & deGruiter, 1983)。

參考書目

- Hambleton, R. K., & deGruiter, D. N. M. (1983). Application of item response models to criterion-referenced test item selection. Journal of Educational Measurement, 20, 355-367.
- Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston, MA: kluwer.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: SAGE.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.
- Woodcock, R. W. (1978). Development and standardization of the Woodcock-Johnson Psycho-Educational Battery. Hingham, MA: Teaching Resources Corporation.
- Wright, B. D. (1977). Solving measurement problems with the Rasch model. Journal of Educational Measurement, 14, 97-116.