用試題反應理論估計運動項目的成績表現排名

— 用定錨法處理運動項目循環賽排名估計的失序 —

姚漢禱

國立體育學院

摘要

本研究的目的是用定錨法處理運動項目循環賽排名估計的失序。Linacre 指出:「 FACETS 電腦程式 (試題反應理論分析質的觀察值) 可將原始測量的順序量尺經過校準後, 全部都是相同的線性構造, 可以推論至一般的等距量尺。」姚漢禱研究用試題反應理論來估計運動項目的成績表現排名, 得到結論:「試題反應理論將排名 (成績表現) 順序資料轉換 (對數轉換) 爲近似等距數線 (連續性數線) 的量尺, 它可以提供更多測驗訊息和更精確的成績表現。」但是, 當積分相同時, 估計潛能和比賽規則判定名次不一致, 產生失序現象。本研究以桌球循環賽相同積分爲研究對象, 運用 FACETS 電腦程式估計受試者能力。本研究的結論是定錨法依規則判定的名次, 將順序量尺的排名, 完成精確的估計。

關鍵詞: 定錨、試題反應理論、循環賽、失序、Rasch 模式。

第一節 緒論

有鑑於體育運動科學的研究通常都以成績表現爲最佳的效標, 在對抗性的運動項目, 如: 桌球、羽球和籃球等等, 其成績表現爲比賽結果的排名; 就測驗的量尺而言, 排名屬於順序量尺, 如果當做多元迴歸分析的效標, 那是錯誤的。在不得以的情況下, 有很多體育運動科學的研究以比賽的勝率或勝分爲效標,「勝率或勝分」只是該受試者本身整體的評估, 缺乏受試者間的比較, 也就是說, 受試者間沒有共同的立足點。或是, 聘請專家評分, 利用總分作爲效標變項, 專家評分不但涉及主觀評分的問題, 而且此變項實際上仍屬於順序量尺。甚至有些研究, 將預測變項的 T 分數總加, 作爲效標變項, 這樣不但違反統計的獨立原則, 更因自身相關, 高估迴歸分析的資料, 也是大有問題。事實上, 吾人可以利用試題反應理論提升量尺的精確性, 在試題反應理論的基礎上, 將量尺以 logit 爲單位, 建立在單向度線性等距量尺上, 具有可加性, 可以進一步做計量的處理。

姚漢禱 (2001) 研究用試題反應理論來估計運動項目的成績表現排名, 結果成功的依據試題反應理論, 利用 FACETS 電腦程式, 估計多層面 Rasch 分析成績表現, 將運動項目循環賽的成績表現排名 (順序資料), 透過 Rasch 模式轉換 (對數轉換) 爲近似等距數線 (連續性數線) 的量尺, 它可以提供更多測驗訊息和更精確的成績表現。研究的估計模式包括: 勝負場模式、勝負局模式和聯合模式三種, 雖然順利的完成估計工作, 但是, 當積分相同時, 比賽規則的判定依據是「相關對手的勝負」, 而試題反應理論的依據爲「全部的反應」, 三種模式都無法滿足「規則判定的名次」和「精確估計」的要求。

應用試題反應理論量化運動項目循環賽的成績表現排名, 在正常情況下, 能夠獲得豐富的測驗訊息, 以及更精確估計成績表現的潛能; 但是, 當比賽勝負總數相同時, 計算的積分是相等, 姚漢禱 2001 年的研究無法滿足「規則判定的名次」和「精確估計」一致的要求, 因此, 本研究的目的在於: 用定錨法處理運動項目循環賽排名估計的失序。

第二節 文獻探討

首先探討對抗形式的運動競賽制度, 它主要分爲「循環賽」和「淘汰賽」兩大賽制, 其餘的競賽制度都是由前述兩者演變而成。本研究的重點是循環賽制的單循環比賽, 單循環賽制是指參加比賽的各單位都需相互對抗比賽壹次, 相較於單淘汰賽制, 其優點是較符合公平原則, 缺點則是比賽場次較多, 且需要充足的場地、時間和裁判。比賽場次=N(N-1)／2[N 爲參加比賽的單位數目]。比賽場次順序的編排如表一 (以五隊爲例,

引自姚漢禱 2001 年體育行政講義):

表一　五隊單循環賽的比賽場次順序編排表

第一輪	第二輪	第三輪	第四輪	第五輪
× —— A	× —— E	× —— D	× —— C	× —— B
E —— B	D —— A	C —— E	B —— D	A —— C
D —— C	C —— B	B —— A	A —— E	E —— D

註：1. ×代表輪空，英文字母代表參賽單位。
2.五隊時每輪有兩場比賽
3.上例第一輪比賽為；第一場E對B比賽，第二場D對C比賽，以下類推。
4.輪空不動，其餘字母依順時鐘方向輪轉。（亦可逆順時鐘轉）
5.奇數比賽單位，每一輪有一次輪空；偶數比賽單位，則第一號取代輪空位置。

循環賽制記錄通常採用雙向表格或圖示法, 參賽隊數較多時, 使用雙向表格較簡便; 參賽隊數較少時, 使用圖示法較清晰明確。雙向表格因本身對本身不必比賽, 形成對角線劃斜線刪除, 通常用下半三角矩陣代表比賽順序, 上半三角矩陣等待填寫比賽結果 (參考表二); 或由橫列爲主, 在相對位置, 填寫本身與他隊的比賽結果。右側邊有積分和名次兩欄, 積分算法是: 勝一場得兩分, 敗一場得一分, 棄權以零分計。名次判定爲: 積分最高者排第一名, 依序往下排名; 兩隊積分相同, 以兩隊比賽結果決定名次; 三隊積分相同, 以相關隊比賽結果的勝率決定名次。

表二　五隊單循環賽時的雙向表格記錄表

結果 / 場次	A	B	C	D	E	積分	名次
A							
B	6						
C	9	4					
D	3	7	2				
E	8	1	5	10			

接著探討國內外有關配對比較量化之研究, 早在 1927 年 Thurstone 針對社會評價提出配對比較法, 到了 1952 年 Bradley & Terry 針對不完全區間設計的排名分析, 建

立配對比較法的理論公式。Davidson 和 Beaver(1977) 的進一步研究:「將 Bradley-Terry 模式擴充, 允許配對比賽結果爲平手 (和局)。」試題反應理論提昇量化的水準後, 1995 年 Linacre 更進一步指出: 配對比賽的 Bradley-Terry 模式, 能夠用 Rasch 測量模式寫成:

$$\log\left(\frac{P_{n>m}}{P_{n<m}}\right) = B_n - B_m.$$

在 Bradley-Terry 模式中 n 和 m 不能相等 (和局), 則方程式不能成立, Davidson 和 Beaver (1977) 在模式中加上參數 ν, Davidson 和 Beaver 的模式成爲:

$$P_{n=m} = \sqrt[\nu]{P_{n>m}P_{n<m}}.$$

荷蘭人 Matthews 和 Morris (1995) 應用 Davidson-Beaver 模式配對比較研究藥品減輕痛苦的效果。Linacre (1995) 用 Rasch 模式表示相等和排名的效果, 用下列兩個公式呈現:

$$\log\left(\frac{P_{n>m}}{P_{n=m}}\right) = B_n + F - B_m - T,$$

$$\log\left(\frac{P_{n=m}}{P_{n<m}}\right) = B_n + F - B_m + T.$$

上述的參數 F 是指獲得利益的層面, 參數 T 是指相等層面。由這些理論架構的研究, 可以了解到『配對比較』能夠應用 Rasch 測量模式表示, 也就是說, 對抗形式運動項目的競賽制度, 透過層面 (效果、或要素) 參數建立線性的對數模式, 因此, 本研究的排名賽制也是『配對比較』的比賽, 所以理論上, 也能夠利用線性對數的 Rasch 模式來估計排名賽制的成績表現。

Linacre (1997) 利用 Winsteps (Rasch 模式估計專用電腦軟體) 實際驗證前述『配對比較』的測量估計, 他是以 1971 年威尼斯西洋棋錦標賽 (循環賽) 爲例 (參考圖一), 共有 Browne 等十一名選手, 比賽結果勝者記『1』、敗者記『0』、平手記『D』。從測量觀點來看: 資料是獨立樣本 (每位受試者在每一個試題, 只有一次連結), 因此, 完全的全部資料 (重複模式) 應該有 726 個資料點, 而配對比較只有 110 個資料點, 缺失值高達 85 %(616 個資料點)。就測量而言, 循環賽是估計潛能的最低限度設計, 每一個試題只有一次連結。Rasch 模式爲 $B_n - B_m = \log(R_n／R_m)$; 相同比賽的難度 Do, 估計 n 選手的能力爲 B_n' 爲 $B_n' - D_o = \log(R_n／R_m)$, 同理估計 m 選手的能力爲 $B'm$, $B'm - Do = \log(Rm／Rn)$。因此 $B_n' - B_m' = 2\log(R_n／R_m) = 2B_n - B_m$, 由此可知, 循環賽因主客隊重複計分, 宜用 0.5 加權調整之。

選手姓名	1971 年威尼斯西洋棋錦標賽
Browne	1D.0..1...1....1.....1......D.......D........1.........1..........
Mariotti	0.1.D..0...1....1.....1......D.......1........D.........1.........
Tatai	.D0..0..1...D....D.....1......1.......1........1.........D........
Hort	...1D1...D...D....1.....D......D.......D........1.........0.......
Kavalek	......010D....D....D.....1......D.......1........1.........D......
Damjanovic	.........00DDD.....D.....D......D.......1........D.........1.....
Gligoric	...............00D0DD......D......1.......1........1.........0....
Radulov	....................000D0DD.......D.......1........D.........1...
Bobotsov	..........................DD0DDD0D........0........0.........1..
Cosulich	..................................D00D00001.........1.........1.
Westerinen	....0D000D0D10..........1
Zichichi	....

引自 Linacre (1997) Paired Comparisons with Standard Rasch Software.

圖一　1971 年威尼斯西洋棋錦標賽成績

Linacre(1999) 和 Linacre(2001 b) 研究配對比較的測量極端分數, 指出:「配對比較是一種多變且強韌的方法, 對編製測量時: 資料收集簡單、容易分析。」研究結果發現:「仍有很大的缺點, 在收集資料的初步期無法預測, 分析 NCAA 籃球隊對第一場到第六場 (或更多場) 都不適當。使用貝氏 (Bayesian) 程序調整極端分數,

說明『以合理的 0.33 來調整極端分數 (勝負的加權), 導出球隊能力爲』: 上述公式其意爲:

用 Rasch 測量估計球隊的能力時, 在初步階段因資料不足, 導致估計的偏差; 使用 0.33 的加權, 能增加勝負的預測。

$$B_T \approx M_T + X_T \log\left(\frac{W-0.33}{0.33}\right) = M_T + X_T \log(3W-1)$$

B_T 是球隊能力。

M_T 是對方球隊平均能力。

X_T 是根據對方球隊能力分布的發展因子。」

Bode (2000) 研究配對比較的選擇力, 指出:「使用傳統的評分量尺形式, 無法研究醫院服務的反應, 利用護士和醫師配對比較調查, 能明顯的估算其異同。」2001 年 Linacre(2001b) 再用配對比較法測量團隊成績表現, 研究兩年的美國大學籃球賽和美式足球賽歸納出十三項要點。摘譯相關的重點如下:「●選擇合理的測量單位, 每個 logit

分為十個單位, 它能估算數週到整季最好和最差的球隊。●選擇合理的賽季前段排名、賽季最後的排名或專家意見, 指派合理的加權指數。●隨時累積比賽結果, 包含勝負、主客場或中立訊息。● 排除示範賽、邀請賽等非正式比賽。● 設定分析程式: 三層面, 第一層面出現兩次, 第二層面主場優勢。主場優勢加到主隊。模式=主, 客, 局部測量調整, 評分量尺「 Model=?,-?,?,R: 」表示程式有三個程層面, 用逗號將三個問號隔開, 問號代表層面, 負號表示配對賽的對手,R 表示評分量尺。最後執行測量估計和預測未來結果, 如果「主隊能力」加上「主隊優勢」, 大於或等於「客隊能力」, 則預測主隊勝; 其他的預測主隊負。實際預測結果有 90 % 正確。

由國外配對比較法的研究可知: 從 Rasch 測量分析介入開始, 克服「平手」的問題, 成功的估計循環賽的潛能, 但 Rasch 多重計分模式僅能估計循環賽, 而後導向多層面分析, 提出合理的設計、加權、重要層面、測量單位和模式, 使預測正確的結果高達 90 %。可是到目前為止仍局限於循環賽的估計, 而國內研究始自筆者 1999 年提出申請的國科會專題研究計畫, 2001 年完成「用試題反應理論估計運動項目的成績表現排名」(姚漢禱, 2001), 發展勝負場模式、勝負局模式和聯合模式三種估計模式, 成功的估計運動項目成績表現的潛能; 但是, 當積分相同時, 比賽規則的判定依據是「相關對手的勝負」, 而試題反應理論的依據為「全部的反應」, 產生估計潛能失序的現象 (與規則判定名次不同), 三種模式都無法滿足「規則判定的名次」和「精確估計」的要求。2002 年接續進一步研究:「以 Rasch 測量有效的等化分組循環賽的成績表現」, 將估計循環賽延伸至分組循環賽, 研究得到的結論是: 「 Rasch 測量可以提供分組循環賽所有參賽者的精確校準能力。」 仍然無法解決「積分相同時, 估計潛能失序」的問題。

在研究估計循環賽制的潛能中, 從建立配對比較法的理論公式開始, 然後擴充模式包含平手 (和局) 的配對比賽, 進而以 Rasch 測量模式取代, 最近更進一步, 採用多層面 Rasch 測量模式分析, 除了融入相對比賽的反應, 也涵蓋整體比賽的成績。在運動競技的賽制中, 使用 Rasch 測量模式量化循環賽, 已有初步的成果, 對於估計積分相同潛能時的失序, 有待繼續研究。

第三節　研究方法

壹、研究對象

本研究以大學甲組桌球隊的循環賽測驗爲研究對象, 共有十五位選手, 選手編號由一至十五號。

貳、研究問題

十五位選手循環賽測驗的結果 (參考表三, 資料來源: 姚漢禱, 2001b), 根據比賽規則判定選手的名次: 第一位選手 (一號) 十二勝二負、排名第一, 第二位選手 (二號) 十一勝三負、排名第二 (因爲積分相同的相關選手比賽結果, 第二位選手勝第三位選手), 第三位選手 (三號)) 十一勝三負、排名第三 (因爲積分相同的相關選手比賽結果, 第二位選手勝第三位選手), 第四位選手 (四號) 九勝五負、排名第四 (因爲積分相同的相關選手比賽結果, 四勝), 第五位選手 (五號) 九勝五負、排名第六 (因爲積分相同的相關選手比賽結果, 二勝二負, 第七位選手勝第五位選手), 第六位選手 (六號) 九勝五負、排名第八 (因爲積分相同的相關選手比賽結果, 一勝三負, 第八位選手勝第六位選手), 第七位選手 (七號) 九勝五負、排名第五 (因爲積分相同的相關選手比賽結果, 二勝二負, 第七位選手勝第五位選手), 第八位選手 (八號) 九勝五負、排名第七 (因爲積分相同的相關選手比賽結果, 一勝三負, 第八位選手勝第六位選手), 第九位選手 (九號) 八勝六負、排名第九, 第十位選手 (十號) 六勝八負、排名第十, 第十一位選手 (十一號) 五勝九負、排名第十一, 第十二位選手 (十二號) 三勝十一負、排名第十二, 第十三位選手 (十三號) 二勝十二負、排名第十三 (因爲積分相同的相關選手比賽結果, 第十三位選手勝第十四位選手), 第十四位選手 (十四號) 二勝十二負、排名第十四 (因爲積分相同的相關選手比賽結果, 第十三位選手勝第十四位選手), 第十五位選手 (十五號) 零勝十四負、排名第十五。

表三　十五位選手循環賽測驗的結果表

選手	勝場	負場	積分	規則判定原因	規則名次
1	12	2	26		1
2	11	3	25	二號選手勝三號選手	2
3	11	3	25	三號選手負二號選手	3
4	9	5	23	積分相同選手的相關比賽結果，四勝	5
5	9	5	23	積分相同的相關比賽二勝二負，第七位選手勝第五位選手	7
6	9	5	23	積分相同的相關比賽一勝三負，第八位選手勝第六位選手	4
7	9	5	23	積分相同的相關比賽二勝二負，第七位選手勝第五位選手	6
8	9	5	23	積分相同的相關比賽一勝三負，第八位選手勝第六位選手	9
9	8	6	22		8
10	6	8	20		10
11	5	9	19		11
12	3	11	17		12
13	2	12	16	13號選手勝 14號選手	14
14	2	12	16	14號選手負 13號選手	13
15	0	14	14		15

據比賽規則判定選手的名次, 判斷三種模式估計的合理性和精確度, 在明確的勝負不同時, 試題反應理論估計的量尺和規則判定的名次一致, 但估計的結果將量尺由順序性, 提昇到等距性, 可用於其他統計的運算; 但是循環賽積分相同時就出現分歧 (參考表四), 勝負場模式不違反規則判定選手的名次, 但無法分辨選手的優劣。勝負局模式違反規則判定選手的名次, 但選手優劣的鑑別力最好, 且具有較精確的量尺。聯合模式剛好介於兩者之間, 違反規則判定選手的名次變少, 選手優劣的鑑別力中等, 且具有較精確的量尺。

表四　比賽規則和三種模式估計的排名比較表

選手	規則名次	勝負場模式		勝負局模式		聯合模式	
		潛能	名次	潛能	名次	潛能	名次
1	1	0.75	1	1.1	1	0.94	1
2	2	0.59	2	1.06	2	0.85	2
3	3	0.59	2	1.01	3	0.82	3
4	4	0.29	4	0.63	5	0.48	5
5	6	0.29	4	0.53	7	0.42	7
6	8	0.29	4	0.7	4	0.53	4
7	5	0.29	4	0.6	6	0.46	6
8	7	0.29	4	0.19	9	0.21	9
9	9	0.15	9	0.44	8	0.31	8
10	10	-0.14	10	-0.17	10	-0.17	10
11	11	-0.28	11	-0.51	11	-0.42	11
12	12	-0.58	12	-0.65	12	-0.62	12
13	13	-0.74	13	-1.87	15	-1.32	14
14	14	-0.74	13	-1.43	13	-1.12	13
15	15	-1.05	15	-1.62	14	-1.36	15

上述比較產生差異最主要的原因，在於所依據的法則不同，當積分相同時，比賽規則的判定依據「相關對手的勝負」，而試題反應理論的依據是「全部的反應」。如何才能消弭試題反應理論和比賽規則的出入？須要運用測驗計分的技巧，合理有效的估計選手的潛能。

參、研究方法

本研究以循環賽的規則判定爲最高原則，多層面 Rasch 測量分析爲估計方法，利用定錨技巧逐步調整合理的估計值，直到估計選手的潛能完全符合規則的判定。選擇以勝負場模式的估計潛能爲主要定錨指標，利用 FACETS(Linacre, 2001a) 電腦程式執行運算。

第四節 研究結果

壹、第二、三名定錨估計潛能

首先說明比賽結果的情況: 二號選手和三號選手都是十一勝三負, 本節重點在於估計第二、三名的潛能, 在此排名之前的第一名已確定是一號選手, 一號選手估計的潛能爲 0.75, 在此排名之後的第四名, 尚未確定是那一位選手, 而且積分相同的選手有五位, 但就規則判定相關比賽結果, 四號選手應該是第四名。

定錨估計的第一種方法: 除了二號選手和三號選手外, 其餘十三位選手定錨。估計的結果: 二號選手和三號選手潛能都是 0.60 (較原有估計的 0.59 略高, 但仍無法分辨兩位選手潛能的高低), 平均估計測驗標準誤 0.40, 非期望反應的標準化殘差大於、等於二者僅一筆 (佔 0.5%), 表示整體測驗的估計良好。

第二種估計方法: 保留前後排名的選手, 刪除第五位 (五號) 至第十五位 (十五號) 等十一名選手, 剩下第一位 (一號) 至第四位 (四號) 選手, 只有四名選手的相關比賽資料; 定錨潛能一號選手 0.75、四號選手 0.29。估計潛能結果: 二號選手和一號選手都是 0.75, 三號選手 0.24、四號選手 0.29, 平均估計測驗標準誤 0.83, 無非期望反應的標準化殘差大於、等於二者, 表示整體測驗的估計良好。雖然分辨二號選手和三號選手潛能的高低, 但產生失序, 第一名和第二名估計潛能的相同, 第三名低於第四名; 因爲局部資料的估計, 四位選手之間的差距被放大, 但上下限被定錨, 導致估計的潛能和名次不一致。

第三種方法是改進第二種估計方法, 只保留排名前面的選手定錨, 因此只定錨潛能一號選手 0.75 。估計潛能結果: 一號選手是 0.75, 二號選手是 0.19、三號選手 -0.37、四號選手 -0.96, 平均估計測驗標準誤 0.92, 無非期望反應的標準化殘差大於、等於二者, 表示整體測驗的估計良好。此種方法依第一名定錨潛能 0.75, 反應四位選手之間的差距; 但第四名估計的潛能與原來的 0.29 不一致, 因此, 後續還要調整估計的潛能。

本研究選取較佳的第三種估計方法, 將估計四位選手量尺的結果, 調整插入原先勝負場模式的估計量尺中, 一號選手估計潛能 0.75、和四號選手估計潛能 0.29 爲上下限定錨指標。依比率調整插入

得到: 二號選手是 $0.75-(0.75-0.19)\times\left(\frac{0.75-0.29}{0.75-(-0.96)}\right)=0.60$

三號選手是 $0.75-(0.75-0.19)\times\left(\frac{0.75-0.37}{0.75-(-0.96)}\right)=0.45$

貳、第十三、十四名定錨估計潛能

第十三、十四名定錨估計潛能採用前面的第三種方法, 因此定錨第十二名選手勝負場模式估計的潛能 -0.58, 以第十二位 (十二號) 至第十五位 (十五號) 選手相關的比賽結果進行估計。估計潛能結果: 十二號選手是 -0.58、十三號選手 -1.13、十四號選手 -1.16, 十五號選手是 -2.27, 平均估計測驗標準誤 0.92, 也無非期望反應的標準化殘差大於、等於二者, 表示整體測驗的估計良好。

接著, 將估計四位選手量尺的結果, 調整插入原先勝負場模式的估計量尺中, 十二號選手估計潛能 -0.58、和十五號選手估計潛能 -1.05 爲上下限定錨指標。依比率調整插入得到:

十三號選手是 $-0.58-[-0.58-(-1.13)]\times\left(\frac{-0.58-(-1.05)}{-0.58-(-2.27)}\right)=-0.73$

十四號選手是 $-0.58-[-0.58-(-1.16)]\times\left(\frac{-0.58-(-1.05)}{-0.58-(-2.27)}\right)=-0.74$

參、第四名至第八名定錨估計潛能

第四名至第八名定錨估計潛能仍採用前面的第三種方法, 因此定錨第三名選手勝負場模式估計的潛能 0.45, 以第三位 (三號) 至第九位 (九號) 選手相關的比賽結果進行估計。結果得到估計潛能: 三號選手 0.45、四號選手 0.19, 五號和六號選手是 -0.11, 七號、八號和九號選手是 -0.41, 平均估計測驗標準誤 0.59, 無非期望反應的標準化殘差大於、等於二者, 表示整體測驗的估計良好。

這裡四號選手估計潛能 0.19, 在規則判定排名爲第四名, 兩者吻合; 其餘四位選手尚無法正確判定, 因此, 僅調整四號選手的估計量尺, 其餘四位選手爾後再進一步處理。以已完成估計勝負場模式的結果: 第三名 (三號) 估計潛能 0.45、和第九名 (九號) 估計潛能 0.15 爲上下限定錨指標。依比率調整插入估計潛能, 得到: 四號選手 (第四名) 是 $0.45-(0.45-0.19)\times\left(\frac{0.45-0.15}{0.45-(-0.41)}\right)=0.36$。

第二步, 定錨第四名選手 (四號) 勝負場模式估計的潛能 0.36, 以第四位 (四號) 至第九位 (九號) 選手相關的比賽結果進行估計。結果得到估計潛能: 四號選手 0.36, 七號選手是 0.05, 五號、六號、八號和九號選手是 -0.29, 平均估計測驗標準誤 0.64, 無非期望反應的標準化殘差大於、等於二者, 表示整體測驗的估計良好。

這裡七號選手估計潛能 0.05, 在規則判定排名爲第五名, 兩者吻合; 其餘三位選手尚無法正確判定, 因此, 僅調整七號選手的估計量尺, 其餘三位選手稍後再處理。以已完成估計勝負場模式的結果: 第四名 (四號) 估計潛能 0.36、和第九名 (九號) 估計潛能 0.15 爲上下限定錨指標。依比率調整插入估計潛能, 得到: 七號選手 (第五名) 是 $0.36-(0.36-0.05)\times\left(\frac{0.36-0.15}{0.36-(-0.29)}\right)=0.26$。

第三步, 定錨第五名選手 (七號) 勝負場模式估計的潛能 0.26, 以第五位 (五號) 至第九位 (九號) 選手相關的比賽結果進行估計。結果得到估計潛能: 七號選手是 0.26, 五號、六號和八號選手都是 -0.14, 九號選手是 -0.55, 平均估計測驗標準誤 0.72, 也都是無非期望反應的標準化殘差大於、等於二者, 表示整體測驗的估計良好。

這裡五號、六號和八號三位選手估計潛能都是 -0.14, 因此三位取排名中間的八號爲代表 (在規則判定排名爲第七名)。因此, 調整八號選手的估計量尺, 以已完成調整估計勝負場模式的結果: 第五名 (七號) 估計潛能 0.26、和第九名 (九號) 估計潛能 0.15 爲上下限定錨指標。依比率調整插入估計潛能, 得到: 八號選手 (第七名) 是 $0.26-[(0.26-0.14)]\times\left(\frac{0.26-0.15}{0.26-(-0.55)}\right)=0.21$。

第四步, 估計第六名選手 (五號), 以第五名 (七號) 估計潛能 0.26、和第七名 (八號) 估計潛能 0.21 爲上下限定錨指標, 估計選手的相關比賽結果。得到估計潛能: 五號選手是 0.22、七號選手是 0.26、八號選手是 0.21, 平均估計測驗標準誤 1.00, 無非期望反應的標準化殘差大於、等於二者, 表示整體測驗的估計良好。

第五步, 估計第八名選手 (六號), 以第七名 (八號) 估計潛能 0.21、和第九名 (九號) 估計潛能 0.15 爲上下限定錨指標, 估計選手的相關比賽結果。得到估計潛能: 六號選手是 0.16、八號選手是 0.21、九號選手是 0.15, 平均估計測驗標準誤 1.00, 無非期望反應的標準化殘差大於、等於二者, 表示整體測驗的估計良好。

肆、定錨調整估計潛能結果彙集

彙集前面三部分定錨調整估計潛能的結果, 整個十五位選手循環賽的結果, 根據比賽規則判定名次的估計潛能: 一號 0.75、二號 0.60、三號 0.45、四號 0.36、五號 0.22、六號 0.16、七號 0.26、八號0.21、九號 0.15、十號 -0.14、十一號 -0.28、十二號 -0.58、十三號 -0.73、十四號 -0.74、十五號 -1.05(參考表五), 潛能都可鑑別。

表五 比賽規則、勝負場模式和定錨調整估計潛能結果的彙集表

選手	規則名次	勝負場模式		定錨調整	
		潛能	名次	潛能	名次
1	1	0.75	1	0.75	1
2	2	0.59	2	0.6	2
3	3	0.59	2	0.45	3
4	4	0.29	4	0.36	4
5	6	0.29	4	0.22	6
6	8	0.29	4	0.16	8
7	5	0.29	4	0.26	5
8	7	0.29	4	0.21	7
9	9	0.15	9	0.15	9
10	10	-0.14	10	-0.14	10
11	11	-0.28	11	-0.28	11
12	12	-0.58	12	-0.58	12
13	13	-0.74	13	-0.73	13
14	14	-0.74	13	-0.74	14
15	15	-1.05	15	-1.05	15

逐步定錨調整估計, 平均估計測驗標準誤從 0.40 至 1.00, 即平均訊息量從 1.00 至 6.25, 相當於古典測驗理論 Alpha 平均信度從 0.50 至 0.86。非期望反應的標準化殘差大於、等於二者, 僅出現一筆, 逐步縮小定錨範圍時, 會使平均估計測驗標準誤升高, 但無非期望反應出現; 1982 年 Wright 和 Masters 說明測驗的效度指出: 受試者適合表示測量有效。因此, 本研究的受試者測量是有效的。

伍、定錨調整估計潛能檢驗

用定錨調整後的估計潛能, 再進行 FACETS 執行運算, 以探討定錨調整後的估計潛能適合理論的情形; 並透過與原始資料估計的結果做比較, 同時說明測驗的品質。

估計方法採用調整後的潛能爲定錨, 進行分析, 結果顯示分析資料共有 210 筆, 與原始資料的分析相同。估計結果: 一次常態近似迭代 (PROX, normal approximation) 和一次非條件最大可能率法 (UCON, unconditional maximum likelihood) 即達到收

斂標準 (因爲定錨的潛能已經過估計); 原始資料的分析則是一次常態近似迭代得到初步的估計基礎, 接著五次的非條件最大可能率法達到收斂標準的要求。由估計的過程來看, 調整後的潛能適合多層面 Rasch 測量模式。

非期望反應分析, 得到標準化殘差大於或等於二者有一筆, 顯示第十四名的十四號選手勝過第五名的七號選手, 能力低 (-0.74) 的選手勝過能力高 (0.26) 的選手, 嚴重違反常理。這部分的結果, 也和原始資料的分析一致。

定錨調整後估計的各測量層面的對照圖 (參考圖二), 程式設定估計三個層面: 選手、比賽 (虛擬層面, 所有元素皆定錨爲零) 和比賽模式 (資料僅限定在所選擇的模式, 並定錨爲零)。圖中層面以垂直的欄位表示, 層面前的"+"表示選手層面爲「 正向值 」, "−"符號是代表比賽模式層面爲「 負向值 」。

```
------------------------------------------------------------
|測量值 |+選手(編號)              |-比賽模式          |計分    |
------------------------------------------------------------
+    1  +                         +                   +(2)    +
|       |                         |                   |       |
|       |  1                      |                   |       |
|       |                         |                   |       |
|       |  2                      |                   |       |
|       |  3                      |                   |       |
|       |  4                      |                   |       |
|       |  7                      |                   |       |
|       |  5    6    8    9       |                   |       |
|       |                         |                   |       |
*    0  *                         *  2 games          *  1    *
|       |  10                     |                   |       |
|       |                         |                   |       |
|       |  11                     |                   |       |
|       |                         |                   |       |
|       |                         |                   |       |
|       |  12                     |                   |       |
|       |  13   14                |                   |       |
|       |                         |                   |       |
|       |                         |                   |       |
+   -1  +  15                     +                   +       +
|       |                         |                   |  ---  |
|       |                         |                   |       |
|       |                         |                   |       |
|       |                         |                   |       |
|       |                         |                   |       |
|       |                         |                   |       |
|       |                         |                   |       |
|       |                         |                   |       |
|       |                         |                   |       |
+   -2  +                         +                   +(0)    +
------------------------------------------------------------
```

圖二　　定錨調整後估計的各測量層面的對照圖

圖二的垂直軸提供變項線性的定義量尺, 各個層面中的元素, 根據其估計值, 出現在量尺相對的位置, 亦即各層面有共同『測量值』的量尺 (單位 logit)。最右側是計分的期望反應量尺, 以整數呈現, 符號"- - -"是半分的位置, 其極端值加 ( ) 表示, 圖中有 (2) 和 (0) 兩個。將各層面建立在共同的量尺上, 這是試題反應理論的特色, 它使得各

層面能夠直接比較, 讓我們了解試題難度是否和受試者能力相適配, 圖二顯示模式和計分能夠鑑別選手能力。也就是說: 試題難度能夠和受試者能力相適配。本研究的重點是估計選手能力, 和原始資料的分析比較, 已經清楚的分辨 2 號、3 號、4 號和 7 號的潛能, 但「 5 號、6號 、8 號、9 號」、以及「 13 號、14 號」, 尚非常接近。

表六　定錨調整潛能和勝負場模式的選手測量估計比較表

選手（編號）	規則名次	勝負場模式		定錨調整	
		潛能	誤差	潛能	誤差
1	1	0.75	0.41	0.75	0.42
2	2	0.59	0.41	0.6	0.41
3	3	0.59	0.41	0.45	0.40
4	4	0.29	0.40	0.36	0.40
5	6	0.29	0.40	0.22	0.39
6	8	0.29	0.40	0.16	0.39
7	5	0.29	0.40	0.26	0.39
8	7	0.29	0.40	0.21	0.39
9	9	0.15	0.39	0.15	0.39
10	10	-0.14	0.39	-0.14	0.39
11	11	-0.28	0.40	-0.28	0.39
12	12	-0.58	0.41	-0.58	0.41
13	13	-0.74	0.42	-0.73	0.41
14	14	-0.74	0.42	-0.74	0.42
15	15	-1.05	0.44	-1.05	0.44
平均		0.00	0.40	-0.02	0.40
標準差		0.54	0.01	0.52	0.01

循環賽選手測量報導, 比較定錨調整潛能和勝負場模式的選手測量估計 (參考表六)。整體而言:①定錨調整估計潛能的平均 −0.02、標準差 0.52, 估計誤差的平均 0.40、標準差 0.01。② 勝負場模式估計潛能的平均 0.00、標準差 0.54, 估計誤差的平均 0.40、標準差 0.01。兩者結果都很接近, 原始資料的勝負場模式估計是以受試者為集中層面, 即程式設定估計潛能的平均為 0.00, 而定錨調整潛能是根據勝負場模式估計結果調整, 估計潛能的平均 −0.02, 相差甚微; 且估計誤差值相同, 因此, 定錨調整潛能和勝負場模式一樣, 具有相同的測量品質。

定錨調整潛能和勝負場模式的類別統計表, 參考表七。模式界定: 勝得兩分、敗得壹分, 因為十五位選手循環賽, 共有 105 場配對比賽 (對抗賽), 所以產生 105 個勝隊、

105 個負隊, 程式界定模式有 0.5 加權 (因比賽配對, 所有資料皆計算兩次), 即表中累計兩個類別各 52.5 次。量尺只有兩個類別, 屬於二項式試做模式 (Binomial trials model), 鑑別度定錨在 1.00 位置。

表七　定錨調整 潛能和勝負場模式的類別統計表

項目	資料				控制品質			觀察值與期望值的診斷殘差
	類別		百分比	累計	平均	期望值	偏離反應	
	分數	次數	%	%	測量值	測量值	均方	
定錨調整潛能	1	52.5	50%	50%	-.54	.00	1.4	6.2
	2	52.5	50%	100%	.54	.26	1.5	23.1
勝負場模式	1	52.5	50%	50%	-.56	.00	1.4	6.6
	2	52.5	50%	100%	.56	.28	1.5	23.0

註：二項式試做模式：定錨為1.00。

定錨調整潛能的類別校準估計品質爲: 平均潛能 (測量值, 單位 logit) 類別「1」(負隊) 爲 −0.54、類別「2」(勝隊) 爲 0.54, 期望潛能類別「1」(負隊) 爲 0.00、類別「2」(勝隊) 爲 0.26; 和勝負場模式的: 平均潛能類別「1」−0.56、類別「2」0.56, 期望潛能類別「1」爲 0.00、類別「2」0.28, 幾乎完成一致。顯示: 定錨調整潛能維持原來的類別校準品質。

程式設定選手潛能估計爲正向, 所以類別「1」的潛能低於類別「2」, 合乎勝隊潛能高於負隊潛能的構念。偏離反應均方是檢定類別的適合度, 兩者都是潛能類別「1」爲 1.4、類別「2」爲 1.5; 皆稍高於期望值 1.0, 表示此類別有非期望的觀察值, 前面分析非期望反應, 得到一筆資料標準化殘差大於或等於二, 僅佔全體的百分之一, 所以, 可以接受。

觀察值與期望值的診斷殘差 (Obsd-Expd Diagnostic Residual) 是指觀察的反應數與期望次數間的差距。定錨調整潛能診斷殘差的類別「1」爲 6.2、類別「2」爲 23.1, 與勝負場模式診斷殘差的類別「1」爲 6.6、類別「2」爲 23.0 相近。造成觀察值與期望值的診斷殘差的原因有三:①無法收斂。②資料的定錨值不相容。③界定的量尺結構發生反應無法配對 (例如資料的失誤)。本研究的問題應該是「資料的定錨值不相容」, 因爲不管比賽兩隊的潛能差距有多大, 結果仍只有「勝負」二分計分, 唯有透過

整個賽程的評估較合理, 如果只用單場的勝負判斷潛能, 會有較大的偏差。

總結整個逐步定錨調整估計過程, 其整體測驗適合多層面 Rasch 測量模式; 印證本研究結果符合試題反應理論, 定錨調整估計的潛能可以被接受的 (受試者測量有效)。利用循環賽結果, 估計出來的潛能是近似等距數線 (連續性數線) 的 logit 量尺, 即測驗屬於單向度 (建構關連效度); 類別校準中選手潛能估計爲正向, 結果類別「 1 」潛能 -0.54 低於類別「 2 」潛能 0.54, 合乎勝隊潛能高於負隊潛能的構念 (類別效度)。依據 Bond 和 Fox (2001) 提出的觀念, 本研究的測驗具有「受試者測量有效」、「建構效度」和「類別校準影響效度」三項效度證明。就效度而言, 效度是單一構念, 本研究以三項效度說明; 這和 AERA、APA 和 NCME(1999) 測驗標準的「效度論證 (validity argument)」:「一個整體的效度概念, 強調許多證據來支撐」相吻合。

第五節　結論

本研究的目的是用定錨法處理運動項目循環賽排名估計的失序。姚漢禱 (2001) 研究用試題反應理論來估計運動項目的成績表現排名, 結果雖成功的將運動項目循環賽的成績表現排名 (順序資料), 透過 Rasch 模式轉換 (對數轉換) 爲近似等距數線 (連續性數線) 的量尺, 但是, 當積分相同時, 估計潛能和比賽規則判定名次不一致, 產生失序現象。爲消除試題反應理論和比賽規則的差異, 本研究以循環賽的規則判定爲最高原則, 採用 FACETS 電腦程式執行多層面 Rasch 測量分析, 選擇以勝負場模式的估計潛能爲主要定錨指標, 利用定錨技巧逐步調整合理的估計值, 直到估計選手的潛能完全符合規則的判定。得到結果: 完成十五位選手潛能的線性量尺估計, 具有較佳的鑑別力; 最重要的測驗品質保持不變, 具有良好的信度和效度。最後, 本研究的結論是定錨法依規則判定的名次, 將順序量尺的排名, 完成精確的估計。

參考文獻

一、中文部分

姚漢禱。(2001a)。**姚漢禱 2001 年體育行政講義**, 國立體育學院, 運動技術學系, 共 102 頁。

姚漢禱。(2001b)。**用試題反應理論估計運動項目的成績表現排名**, 國科會專題研究計畫成果報告 (專題計畫編號:NSC 89-2413 -H- 179 - 013), 中華民國九十年七月三十一日出版, 共 90 頁。

姚漢禱。(2002)。**以 Rasch 測量有效的等化分組循環賽的成績表現** , 國科會專題研究計畫成果報告 (專題計畫編號:NSC 90 - 2413 -H- 179 - 003), 中華民國九十一年七月三十一日出版, 共 71 頁。

二、英文部分

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *The Standards for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

Bode, R. (2000). Paired Comparisons for Forced Choice. *Rasch Measurement Transactions*, 14 (3), 769.

Bond, T. G., & Fox, C. M (2001). *Applying The Rasch Model: Fundamental Measurement in the Human Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bradley, R.A. and Terry, M E. (1952). Rank analysis of incomplete block designs I: The method of paired comparisons. Biometrika, 39, 324 - 345.

Davidson, R.R.,& Beaver, R.J. (1977). On extending the Bradley-Terry model to incorporate within-pair order effects. Biometrics, 33, 693 - 702.

Fischer, G.H. (1995). The linear logistic test model. In G.H. Fischer & I.W. Molenaar (Eds.), Rasch model: *Foundations, recent developments, and applications* (pp. 131 - 155). New York: Springer-Verlag.

Linacre, J.M. (1995). Paired comparisons with ties: Bradley-Terry and Rasch. *Rasch Measurement Transactions*, 9 (2), 425.

Linacre J.M. (1997). Paired Comparisons with Standard Rasch Software. *Rasch Measurement Transactions*, 11 (3), 584 - 585.

Linacre J.M. (1999). Paired comparison measurement with extreme scores. *Rasch Measurement Transactions*, 12 (3), 646 - 647.

Linacre, J.M. (2001a). FACETS *computer program.* Facets for Windows Version No. 3. 36. 2. 1987 - 2001. Chicago: MESA.

Linacre J.M. (2001b). Paired Comparisons for Measuring Team Performance. *Rasch Measurement Transactions*, 15 (1), 812.

Matthews, J.N.S., Morris, K.P. (1995). An application of Bradley-Terry- type models to the measurement of pain. *Applied Statistics*, 44 (2), 243 - 255.

Thurstone, L.L. (1927). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, 21, 384 - 400.

Wright, B.D., & Masters, G.N. (1982). *Rating scale analysis: Rasch measurement.* Chicago: Mesa Press.

Assessment of Ranking of Performance in Sports form Item Response Theory II: Adjustment Disorder of Round Matches Rankings for using Anchor Method

Han-Dau Yau

Abstract

The purpose of the study was to adjust the disorder of Round Matches rankings for using anchor method. Linacre (1997) pointed out: the FACETS (Rasch measurement computer program)analyzes qualitative observations and it estimates a quantitative measure.

The measures for the elements obtained from one analysis are all in the same linear frame of reference on one common interval scale.

Yau (2001) assessment of ranking of performance in sports form item response theory.

The conclusion of this study was that the item response theory analysis computer program convert (a logistic transformation) ranking of ordinal data (performance) into an approximately equal-interval number line (linear continuum or scale representing the variable). It provided more than even test information and precision estimation in the scale of performance. When the equivalent of total scores, that had loss agreement both the estimation ability and the rank from rules. The person abilities were disorder. The subjects were the equivalent of total scores with round matches in table tennis. We were using the computer program FACETS to estimate person abilities. The conclusion of this study was that the anchor method analysis ranking of ordinal data into a precision estimation by the rank from rules.

Keywords: anchor, item response theory, round matches, disorder, Rasch model.