

應用概化理論評估羽球短發球之最佳試做次數與合理計分方式

許 伯 陽
國立體育學院

摘 要

本研究的目的是利用概化理論推估羽球反手短發球的試做次數與計分方式，再由信度係數來選取羽球反手短發球「最佳」試做次數與「合理」計分方式。由國立體院89學年度下學期修習羽球課同學為受試者，共133位。每位受試者反手短發十球於設計好的目標區中，由GENOVA程式(Crick & Brennan,1983)進行各受試者每次發球得分的資料分析。研究結果發現，羽球反手短發球測驗在試做次數上以試做25次為最佳，在計分方式上以三分計分為最合理。由概化理論可評估羽球反手短發球測驗。

關鍵詞：概化理論、信度係數、羽球發球測驗

壹、緒論

一、前言

古典測驗理論將觀察分數的變異分成兩個部分，一是系統化的真分數變異，另一是隨機化的誤差變異(Shavelson,1989)，測驗在追求真分數不可得的情況下，轉而追求誤差分數，求其誤差變異量，進而以信度係數來代表測量結果的可靠性。古典測驗理論所提供的信度計有：1.重測信度(test-retest)：計算由時間上(day to day)所產生的誤差。2.內部一致性信度(internal-consistency)：計算由試題上所產生的誤差。3.複本信度(alternate-forms)：計算由時間上和試題上所產生的誤差。以上古典測驗理論的各種信度估計方式都是以整體的誤差為來源，沒有細究其內部的各種誤差，余民寧(民82)指出：「古典測驗理論最大缺失，乃是在它以一個測量誤差來函蓋所有的誤差來源，沒有說明其來源的測量情境」。另外，姚漢禱(民87)指出：「古典測驗理論在測驗時各個層面的水準即為母群的全部，稱為固定模式，無法做推論」。因此，整體性誤差以及無法做推論實為古典測驗理論的兩大缺失。

概化理論採用多種信度係數的概念，視誤差為多種情境的組合 (Suen,1990)，另外，概化理論至少會有一個層面是取自母群的隨機樣本，稱為隨機模式，可以去推論到其他的水準，因此概化理論可解決古典測驗理論的兩大缺失。

概化理論的理論基礎是運用變異數分析(ANOVA)的方式，將受試者間的變異數視為「真實變異數」，將評分者間、評分者內和試題間的變異數視為「誤差變異數」(Suen,1990)，概化理論的內涵即是在促使每種測驗情境中，真實變異數的部分越大越好，而誤差變異數的部分越小越好(Shavelson & Webb,1989)。在應用上，概化理論可分為兩個部分，一為概化研究(G-study)：估計有關各種測量層面的變異數大小。二為決斷研究(D-study)：運用G-study所提供的訊息，做特殊目的的有效決策(余民寧，民82)。

在運動技能測驗應用上，概化理論可以處理評分者、試做次數、工具、公式和測量者等變異來源，提供更多的訊息，了解測驗不良的原因。姚漢禱(民87)的研究指出：「概化理論可以改善運動技能測驗的四個主要問題：1.最佳的試作次數 2.合理的計分方式 3.適當的成功標準 4.多層面的整體性信度」。

二、研究目的

大學生經過一個學期的羽球課程後，於期末所做羽球反手短發球之測驗，測驗的方式是否合理？試做次數與計分方式是否適當？本研究的目的是將概化理論運用到羽球反手短發球的運動技能測驗上，嘗試探究下列子題：

- (一) 探究大學生羽球反手短發球技能測驗的最佳試作次數。
- (二) 探究大學生羽球反手短發球技能測驗的合理計分方式。

貳、研究方法

一、研究對象

國立體院89學年度下學期修習羽球課之同學，男生71位，女生62位，共133位。

二、測驗場地

將目標區分設為1.15cm×15cm，為四分區。2.向外延伸10cm，為三分區。3.再向外延伸5cm，為二分區。4.而三分區之外的目標區，為一分區。5.在目標區之外的對方場地、界外或掛網，為零分。參考圖一。

三、測驗方法

受試者依據羽球規則，將球發至目標區，每個人都有足夠的時間反手短發球10次。

四、計分方式

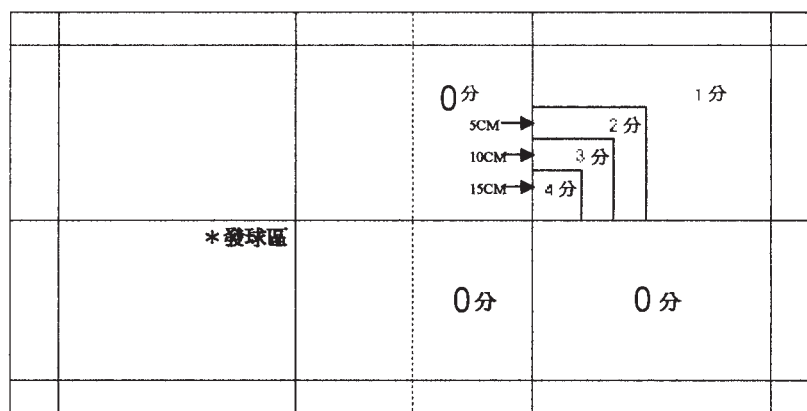
(一) 多重計分方式：依據球的落點給予0到4分。

(二) 四分計分：以4分界線為分界分數，球落入四分區，計為1分，之外為0分。

(三) 三分計分：以3分界線為分界分數，球落入三分區以上者，計為1分，之外為0分。

(四) 兩分計分：以2分界線為分界分數，球落入二分區以上者，計為1分，之外為0分。

(五) 一分計分：以目標區為分界分數，球落入目標區者，計為1分，之外為0分。



圖一 羽球場地規劃

五、實驗設計

以133位受試者，試做10次做為樣本，當做推估母群的不偏估計，受試者、試做次數

和計分方式為變異來源，受試者為固定模式，試做次數和計分方式為隨機模式，為單層面重複設計。概化研究(G-study)的部分包括受試者在各種計分方式下與試做次數以及其交互作用的變異數分析和變異數成分。決斷研究(D-study)即提供各種計分方式下的概化係數，作為選擇最佳試做次數與合理計分方式。

六、資料分析

所得資料使用GENOVA程式(Crick & Brennan,1983)進行分析，第一部份為概化研究(G-study)，第二部分為決斷研究(D-study)。

七、判斷標準

根據姚漢禱(民87)的建議：「測驗的概化係數達到0.6時稱為普通水準」，(Shavelson & Webb,1991)的建議：「測驗的概化係數達到0.8時，稱為合理的水準」。概化係數是信度係數的一種，信度係數大的並不一定就能表示測驗的信度好，與受試者、測驗題數的多寡都有關係。它是一種參考值，提供測驗編製者建立信度的方法或是施測者使用該測驗時的參考。在社會科學中，可接受的最小信度係數看法也未見相同，如 α 係數0.8 (Bryman & Cramer,1997)與0.7(DeVellis ,1991)都有學者認同；在概化係數上也是如此，0.6與0.8是前人研究的一項參考值，在本研究中列出作為未來測驗編製者進行修正或是施測者進行施測的依據。

參、結果與討論

一、性別上的差異

在表一中，顯著水準 值為0.05，計算出的p值為0.2，未達顯著水準，可見男女性別對於羽球反手短發球這項運動技能，在統計上是沒有差異的，所以本研究使用的受試者包括男性女性，排除性別的差異。

表一 羽球反手短發球之性別差異t考驗摘要表

組別	人數	平均數	標準差	t 值	p 值
男性	71	31.89	5.60	1.28	0.20
女性	62	30.53	6.61		

*p<.05

二、多重計分試做次數的概化研究

(一) 表二為試做次數的單層面重複設計變異數分析摘要表，P為受試者，T為試做次數，P×T為交互作用，受試者均方為4.25，試做次數均方為0.60，交互作用均方為1.57。

表二 試做次數變異數分析摘要表

變異來源	離均差平方和	自由度	均方	F 值
P	561.24	132	4.25	2.71*
T	5.44	9	0.60	0.38
P× T	1864.96	1188	1.57	
Total	2431.64	1329		

* $p < .05$

(二) 表三為試做次數變異成分摘要表，受試者的變異成分(指的是推估母群的變異成分)，也就是在概化理論中真分數的部分，在兩種計算方式下(迭代法和期望均方公式)其值為0.21，試做次數的變異成分為0，為0的原因可能在於試做次數的成分小於程式的預設值。在殘差部分(P×T的交互作用)，變異成分為1.5698，由變異百分比可看出殘差值比受試者和試做次數的變異成分皆高出很多(殘差值為88%，受試者為12%，試做次數為0)，在此可解釋為實驗沒有控制好受試者或是試做次數這兩個變異成分之外的因素，所導致的結果。

表三 試做次數變異成分摘要表

效果	自由度	迭代法	期望均方公式	標準誤	變異百分比
P	132	0.21	0.21	0.0473	11.80
T	9	0	0	0.0022	0
P× T	1188	1.57	1.57	0.0644	88.20

三、多重計分試做次數的決斷研究

表四呈現試做次數決斷研究的部分，研究設計固定受試者為133位，去決斷在試做次數方面的概化係數(信度係數)，在表四中可發現，試做一次時的母群全體分數(真分數)為0.21，期望觀察分數(誤差分數)為1.8256，下限差距(相對誤差變異數)為1.5646，上限差距(絕對誤差變異數)為1.5657，上下限差距是顯示推估到母群時的誤差變異數，因此，若是相同，表示推估的信度係數會相同，試做一次的概化係數0.1411為受試者133位時的信度

係數， Ψ 係數0.1411為推估於母群的信度係數，通常概化係數會比 Ψ 係數來得高一些，相同的原因在於試做次數的變異成分為0(試做期望均方)。理論上，受試者可以試做無限次，可得到1的信度係數，但在實際上時間、疲勞與學習都會有所影響，如何在有限的資源下能得到最高的信度係數是本研究的目的。概化理論在層面上可設定為隨機模式，本研究設定試做次數為隨機層面，可做推估，由普通水準的信度係數0.6與合理水準的信度係數0.8為標準，反回來推估試做次數到底需要幾次。由表四中可發現，試做10次時概化係數為0.6206，達普通水準，試做25次時概化係數可達0.8068，達合理水準。

表四 試做次數決斷研究摘要表

抽樣大小		變異數					信度係數	
受試者 訊息	試作 訊息	母群 分數	期望觀 察分數	下限 差距	上限 差距	平均數	概化 係數	Ψ 係數
133	1	0.21	1.8256	1.5646	1.5657	0.0115	0.1411	0.1411
133	10	0.21	0.4154	0.1544	0.1558	0.0022	0.6206	0.6206
133	12	0.21	0.3923	0.1305	0.1305	0.0015	0.6673	0.6673
133	20	0.21	0.3358	0.0738	0.0758	0.0028	0.7092	0.7092
133	25	0.21	0.2235	0.0615	0.0664	0.0026	0.8068	0.8068
133	30	0.21	0.3119	0.0554	0.0513	0.0026	0.8344	0.8344
133	40	0.21	0.2928	0.0337	0.0359	0.0028	0.8698	0.8698
133	50	0.21	0.2905	0.0311	0.0303	0.0021	0.8955	0.8955

四、計分方式的概化研究

(一) 表五為計分方式的單層面重複設計變異數分析摘要表，是計分方式分為一分計分、兩分計分、三分計分和四分計分等四種，P為受試者，T為試做次數， $P \times T$ 為交互作用，其值呈現如表五。

表五 計分方式變異數分析摘要表

計分方式	變異來源	離均差平方和	自由度	均方	F 值
一分計分	P	32.92	132	0.25	1.38
	T	0.72	9	0.08	0.44
	$P \times T$	215.48	1188	0.18	
兩分計分	P	49.72	132	0.38	1.78
	T	0.77	9	0.09	0.40
	$P \times T$	251.33	1188	0.21	
三分計分	P	63.50	132	0.48	2.13*
	T	0.43	9	0.05	0.21
	$P \times T$	268.17	1188	0.23	

四分計分	P	38.72	132	0.29	1.67
	T	1.10	9	0.12	0.70
	P× T	208.80	1188	0.18	

* $p < .05$

(二) 表六為計分方式的變異成分摘要表，在四種計分方式下，兩種計算方法(迭代法和期望均方公式)的結果均相同，受試者層面的變異成分皆很小(一分計分為0.0068，兩分計分為0.0165，三分計分為0.02555，四分計分為0.0118)，而試做次數層面的變異成分皆為0，原因可能在於試做次數的成分小於程式的預設值。在殘差部分(P×T的交互作用)，比受試者和試做次數的變異成分皆高出很多(一分計分為96%，兩分計分為93%，三分計分為90%，四分計分為94%)。這可解釋為受試者或是試做次數這兩個變異成分之外的因素影響很大。

表六 計分方式變異成分摘要表

計分方式	效果	自由度	迭代法	期望均方公式	標準誤	變異百分比
一分計分	P	132	0.0068	0.0068	0.0031	3.61
	T	9	0	0	0.0003	0
	P× T	1188	0.1814	0.1814	0.0074	96.39
兩分計分	P	132	0.0165	0.0165	0.0047	7.23
	T	9	0	0	0.0003	0
	P× T	1188	0.2116	0.2116	0.0087	92.77
三分計分	P	132	0.02555	0.0256	0.0059	10.19
	T	9	0	0	0.0002	0
	P× T	1188	0.2257	0.2257	0.0093	89.81
四分計分	P	132	0.0118	0.0118	0.0037	6.29
	T	9	0	0	0.0004	0
	P× T	1188	0.1758	0.1758	0.0072	93.71

五、計分方式的決斷研究

表七上半部是概化係數接近0.6，下半部是接近0.8的各種計分方式下的結果，概化係數接近0.6時，也就是可達普通水準，在四種的計分方式下其試做次數分別為40次、20次、15次和25次，概化係數接近0.8時，也就是可達合理水準，試做次數分別為110次、55次、40次和60次，其中以三分計分方式達到普通水準和合理水準的試做次數為最少。

表七 計分方式決斷研究摘要表

計分方式	試做次數	母群分數	期望觀察分數	下限差距	上限差距	概化係數
一分	40	0.0068	0.0113	0.0045	0.0045	0.6000
兩分	20	0.0165	0.0271	0.0106	0.0106	0.6095
三分	15	0.0255	0.0406	0.0150	0.0150	0.6092
四分	25	0.0118	0.0188	0.0070	0.0070	0.6059
計分方式	試做次數	母群分數	期望觀察分數	下限差距	上限差距	概化係數
一分	110	0.0068	0.0085	0.0017	0.0017	0.8050
兩分	55	0.0165	0.0200	0.0035	0.0035	0.8040
三分	40	0.0255	0.0312	0.0056	0.0056	0.8090
四分	60	0.0118	0.0147	0.0029	0.0029	0.8006

參、結論

試做次數是由多重計分模式來測驗其最佳的次數，多重計分能提供較多的訊息，較精確的區分受試者的能力。增加試做次數可以提高測驗的信度，若是沒有學習效果上所造成的假性精熟，或是疲勞效果等的不適當成績表現，試做愈多次，愈能代表受試者的真實能力，也就是在古典測驗理論中的真分數，但由於上述的外在因素或是受限於施測時的時間限制、或是測驗工具經濟上的考量，不得不選擇一項測驗試做次數的標準，概化理論在此可以提供一個客觀的數據，作為施測者評量決斷的有利依據。研究結果得知在多重計分的方式下，試做10次能達到普通水準，試做25次即能達到合理水準。

另外，在計分方式的選取上，本研究比較四種的二分計分方式(四分、三分、兩分、一分計分)，二分計分較單純，容易符合模式要求，但分界分數的設定會影響到測驗的難易，如何選取適合受試者的分界分數是重要的。概化理論推估上述四種不同的計分方式，由所提供的信度係數比較不同計分方式下的試做次數，以此來決斷合理的計分方式。研究結果發現，三分計分的計分方式，其達普通水準(0.6092)和合理水準(0.809)的試做次數(15, 40)，比較容易測驗(一分計分、兩分計分)的試做次數為低，另外也比困難測驗(四分計分)的試做次數為低，也就是說，較難或較易的測驗需要更多的試做次數才能達到普通或合理的水準，要達到同樣的概化係數(0.8)，三分計分為最經濟有效的計分方式，因此，三分計分為合理的計分方式。

本研究嘗試將概化理論運用到羽球反手短發球的運動技能測驗上，評估最佳的試做次數和合理的計分方式，結果發現，單層面重複設計的概化理論是適合於重複動作的運動技能測驗。本研究的結論為，達到0.8的合理水準(Shavelson & Webb,1991)時，最佳的試做次數為25次，合理的計分方式為三分計分(25cm×25cm)，可做為編製羽球反手短發球測驗的理論依據。

引用文獻

- 余民寧。(1993)。測量理論的發展趨勢。載於中國測驗學會編：心理測驗的發展與應用 (34-39頁)。台北：心理出版社。
- 姚漢禱。(1992)。運動技術測驗中設定試做的適宜次數。體育學報，14，339-350。
- 姚漢禱。(1997)。評分量尺模式和RASCH二分模式應用在運動技能測驗上的比較。體育學報，22，95-106。
- 姚漢禱。(1998)。概化理論在運動技能測驗上的專題研究。國科會專案計畫。
- Crick, J. E., & Brennan, R. L. (1983). Manual for GENOVA: A generalized analysis of variance system. Iowa City, IA : ACT.
- Suen, H. K. (1990). Principles of test theories. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sanders, P. F. (1992). Alternative solutions for optimization problems in generalizability theory. Psychometrika,57(3),351-356.
- Shavelson, R. J., Webb, N. M., & Rowley, G. L.(1989). Generalizability theory.American Psychologist,44(6),922-932.
- Shavelson, R. J., & Webb, N. M.(1991). Generalizability theory a primer. Newbury Park,Inc:SAGE.
- Godbout, P.,& Schutz,R.W.(1983).Generalizability of ratings of motor performances with reference to various observational designs.Research Quarterly for Exercise and Sport,54(1),20-27.
- Stamm, C. L.,& Moore, J. E.(1980).Application of generalizability theory in estimating the reliability of a motor performance test. Research Quarterly for Exercise and Sport,51(2),382-388.

投稿日期：90年12月

審稿日期：91年 1月

接受日期：91年 3月

The Application of Generalizability Theory for the Evaluation of the Best Trials and the Fitness of Scaling in Badminton Short Serve Test

Hsu Po-Yang
National College of Physical Education and Sports

Abstract

The purpose of this study was to evaluate the best trials and the fitness of scaling in the badminton backhand serve test by Generalizability Theory. Subjects were 133 college students who learned badminton course in half academic year. Subjects were requested to serve 10 shuttles over the net and to land it close to the front corners of the target zone. GENOVA (Crick & Brennan,1983) was used to analyze the data. We could use reliability coefficients to estimate the best trials and the fitness of scaling from analysis of GENOVA . The results were as follows: the best trials was 25 times and the fitness of scaling was third . Generalizability Theory could evaluate the badminton backhand serve test.

keywords : Generalizability Theory , reliability coefficients , badminton serve test