

運動相關測驗效度的發展與檢定的趨勢

黃芳銘

國立體育學院

陳文銓 鄭桂玫

大仁技術學院

摘 要

運用測驗理論來解決運動相關測驗的效度問題，近年來在國內已漸漸受到重視。早期運動相關測驗的編製，外部效度（效標關聯效度）的處理居多，隨著這種處理的侷限性，運動相關測驗研究者改採內部效度的處理，其中，在處理建構效度時，相當倚賴內部一致性分析及探測性因素分析，這兩種技術往往只能夠處理個別項目的鑑別性、因素負荷量，以及因素間的全無或全有關係，卻無法更彈性的處理因素部分相關及整體理論因素結構的檢定。

隨著統計技術與軟體的提升，國內在檢定運動相關測驗的建構效度上，已有採用結構方程模式中的肯證式因素分析。肯證式因素分析不僅能解決內部一致性分析與探測性因素分析所遭遇的困難之外，其能夠以事先建立的理論因素模式來檢定，更是符合建構效度的定義與要求。本研究乃希望透過這種新趨勢的介紹，以便引起共鳴，促使更多學者的投入，讓運動相關測驗的發展更上一層。

關鍵詞：測驗，效度，結構方程模式，探測性因素分析，肯證式因素分析，內部一致性分析。

壹、緒論

發展良好的運動相關測驗，對體育或運動訓練都是相當重要。因為測驗可以瞭解運動員是否達到既定的訓練目標，發現進步的情形，指出其運動技能的優缺點，提供訓練處方回饋訓練計畫，以及作為選才之用。基於這些好處，在國內，運動漸漸受到重視的同時，運動測驗也受到重視。因而，運用測驗理論解決運動測驗的相關問題，近年來是愈來愈受到重視。

測驗理論的主要研究內容，涉及兩種層面：測量(measurement)與分析 (analysis)，前者主要是探討抽樣程序、評量工具（測驗或量表）的發展、信效度等相關議題，後者則牽涉統計技術的運用與實際測驗內容的檢驗等主題 (Pedhazur & Schmelkin, 1991)。就測驗的內涵上而言，效度的問題是最難解決的，而一個測驗若沒有效度，即使再實用、可靠、客觀皆是白費的。

效度的意義與理論的處理，隨著測驗的不斷發展而愈來愈健全，也拜統計技術發展之賜，特別是結構方程模式 (structural equation modeling) 統計的產生，使得測驗效度在分析層面上是愈來愈容易處理且更加簡便。透過對新的測驗趨勢在運動測驗的探討，本文主要的目的是對此種測驗理論趨勢在我國運動測驗發展做一介紹，並且針對各種效度的問題提出討論，最後以結構方程模式的統計方式之應用來說明檢定運動測驗建構效度的基本程序，作為運動測驗建構效度檢定的典型例證。

貳、測驗效度意義的發展

在測驗的歷史發展過程中，效度的意義隨著發展的成熟性而不斷的改變。過去，效度經常定義為：測驗所欲測量的特質或行為的正確性，或者是指衡量測驗是否能夠測量到其所要測量特質程度之統計指標。因此，效度係數愈高，表示愈能夠測量到一個測驗所想要測量的特質。這種定義盛行於 50、60 年代，使得測驗效度檢定的絕大部分責任落在測驗發展者的身上，因發展者必須先宣稱這是一個什麼樣的測驗。

美國國家教育測量諮詢會 (NCME)、美國教育研究協會 (AERA) 以及美國心理學會 (APA) 共同出版的測驗標準手冊，在一九八五年時，將此一手冊命名為《教育與心理測驗歷程標準》(Standards of Educational and Psychological Testing)。在此一手冊中「效度」一詞被定義為 (1985, p.9): 「對於測驗分數之特定推論的適當性 (appropriateness)、意義性 (meaningfulness) 及其有用性 (usefulness)。效度就是一種不斷累積證據以支持上述推論的過程。」這種效度的定義將過去強調測驗結果的定義加以擴大，也就是對效度的思考已包含整個測驗發展的完全歷程，一種包括發展、施測、到結果解釋的動態歷程 (王文中、呂金燮、吳毓瑩、張郁雯、張淑慧，民 88)。這種改變主要有兩個方向，一是概念結構的改變，二是探究重點的改變 (吳毓瑩，民 85)。

如何證明一份測驗具有良好的效度是心理與教育測驗學者在發展測驗時相當關心的課題，也是發展運動相關測驗學者所關心的事。不過，測驗的目標不同，所需提供資料的型態就不同，

其效度的種類也就會不同。依據美國國家教育測量諮詢會、美國教育研究協會以及美國心理學會的看法，認為效度有三種形式：效標關聯的（criterion-related）效度、內容關聯的（content-related）效度，以及理論建構關聯的（construct-related）效度。

參、效標關聯效度在運動相關測驗的發展及檢定的趨勢

效標關聯效度是一種以經驗性的方法研究測驗分數與一些外在效標間的關係，所以又可以稱為經驗效度或統計效度。也就是說，是根據一些外在的效標所決定的效度。如果測驗分數與外在效標間的關係愈高，表示效標效度愈高。反之，則不然。效標效度愈高，測驗分數愈是能夠有效預測外在效標。因此，使用適當的效標是決定效標效度的關鍵。

效標關聯效度是國內早期編製運動相關測驗時，常用的一種效度檢定方法。經常使用的效標包括比賽名次或評分、國際或國內排名、裁判評分、專家評比以及教練對其選手能力之評比等。其中，當效標收集的時間與測驗本身大致同時，所獲得的效度證據稱為同時效度（concurrent validity）。積差相關係數的統計方式，最常被使用於同時效度的檢定。當然，當效標與測驗分數都不是連續變項的形式時，那就必需使用其他類的相關統計技術。如果獲得效標的時間與測驗晚一段時間，且此效標具有預測某種未來成就的效果，則所獲得的效標證據稱為預測效度（predict validity）。

適當的效標應具備四項特徵：適切性、可靠性、客觀性、可用性。適切性乃是指效標資料應該能反映出測驗所欲測量的重要特徵。可靠性是指效標資料本身必須是可靠的，否則，再可靠的測驗也無法預測不可靠的效標。客觀性是指良好的效標必須能避免偏差和效標混淆。可用性則是在選擇效標時，效標資料的取得是否容易和方便。一些學者將效度區分為內部效度（inside validity）與外部效度（outside validity），效標關聯效度就是一種外部效度，從前面所描述的效標標準來看，要獲得一個良好的效標是相當不容易的事。其中以教練評分、專家評分、裁判評分作為效標，其客觀性較弱。而國際與國內排名，因測驗需要的樣本數相當大，獲排名人數無法符合樣本所需之數量，且收集需要相當長的時間。使用比賽名次做效標不僅需花費鉅資，且收集時間也是相當的長。例如，孫宜芬（民 73）對師大選修羽球男生 40 名來編製羽球技能測驗，以比賽名次為效標，所需的比賽場次達 239 場。基於這些理由，一些運動相關測驗編製者，則傾向於採用內部效度的方法，包括內容效度與建構效度。

肆、內容效度在運動相關測驗的發展及檢定的趨勢

內容效度乃是指一個測驗本身所能包含的概念意義範圍或程度。亦即測驗的內容是否針對所欲測量的目的，且具有代表性與適當性。就運動訓練的角度來看，乃是檢驗每一個題目是否符合該項運動的訓練內容與訓練的目標。如果測驗項目很能代表訓練內容的樣本及所預期的行

為改變，而沒有其他無關因素的影響，則表示測驗具有良好的內容效度。因此，我們知道一個測驗是否有較理想的內容效度，其關鍵在於測驗項目的取樣是否適當，而非僅在於測驗的外表。

分析內容效度的方式大致有兩類：邏輯的分析方式以及實證的分析方式。對於運動相關測驗而言，使用邏輯分析的方式乃是邀請專家及學者們對於所要編製的運動相關測驗之內容領域，加以做一致性的界定。例如，在編製羽球基本體能測驗時，可以先參考國內文獻與理論，將國內羽球相關運動體能的要素加以統整，如許樹淵（民 86）將一般體能的要素界定為：肌力、速度、瞬發力、柔韌性、耐力、穩定平衡、韻律性、正確性、敏捷性、感覺，以及協調性。林正常（民 78）在其《運動科學與訓練—運動教練手冊》一書中將體能的內容界定為：肌力、瞬發力、耐力、柔軟性、協調性、敏捷性、時機、平衡、放鬆以及集中力。張永文（民 86）認為羽球運動體能的要素應包括：肌力、速度、瞬發力、柔韌性、耐力、平衡、正確性、韻律性、敏捷性以及協調性。然後，請羽球教練或相關研究專家，共同討論出在這些項目中哪些最能代表羽球基本運動體能？獲得的結論是瞬發力、速度、耐力、柔韌性、敏捷性、協調性、平衡性等七項可以做為羽球基本運動體能。

實證性的分析方式則是採用量化的方式來處理內容效度，其目的是能夠客觀地呈現內容效度，提倡此種方式的學者包括 Aiken（1980，1985），Mehrens & Lehmann（1991）以及余民寧（民 82）等。其中 Aiken 的內容效度係數適用於此一方法，其公式如下：

$$V = (\sum V_j / n) \quad V_j = \left[\frac{S_j}{n(c-1)} \right] \quad S_j = \sum d_{ij} \quad j = 1, \dots, n \quad i = 1, \dots, n$$

V 即是整份測驗的內容效度係數， V_j 則是第 j 個試題之內容係數。 S_j 表示 n 位專家在 c 個評定等級之評量，表示其對試題 j 之評分的離差分數之總和； d_{ij} 是專家 i 在試題 j 之評分與最小評分差的絕對值。 V 之數值將介於 0 與 1 之間，數值愈大表示內容效度愈高。

國內運動相關測驗較少處理內容效度，因此，此二種方法較少被使用。而過去有關運動相關測驗的編製，相當重視建構關聯效度，以下介紹建構效度的發展與檢定的趨勢。

伍、建構效度在運動相關測驗的發展與檢定的趨勢

建構效度是指研究的任何發現關聯到是否測量工具所得到的結果會一如理論對該建構所做的預測。建構概念是指任何一個理論的概念，而建構效度的意義僅限於測量工具產生符合理論預測結果的實證證據。所以張至滿（民 84）將建構效度定義為「利用推理和統計結果來判斷某測驗是否具備應有的構想。」

許多的測驗書籍將建構效度視為是效度概念的核心。Loevinger（1957）更是強調任何用以證明效度的證據，包括內容效度證據、效標關聯效度證據等皆是建構效度的部分證據。因為，這些證據乃是被收集來了解測驗結果是否能夠適切地，且有意義地詮釋出想要評量的能力，其實，此即是一種理論建構的意涵。

建構效度如何驗證呢？Gronlund & Linn (1990) 等人提出四個步驟：

- 1.先針對所欲測量的構念做理論的分析與建構，然後，據以發展出一套評量工具和策略；也就是說，提出有關理論建構的說明，來設計測驗所需的項目。
- 2.對所建構的理論提出預測或假設的說明。
- 3.蒐集資料，以實證分析的方式驗證上述的預測或假設；也就是說，採用各種方法收集實際的資料，考驗上述所提出的預測或假設的正確性。
- 4.收集其他型態的輔助證據，修正理論或是淘汰與理論建構相反的項目，然後，重覆第二和第三步驟，一直到上述的預測或假設得到驗證，且測驗的建構效度獲得支持為止。否則，即表示該測驗的效度是有問題的，或則是該理論建構有問題，或者兩者皆是。此時，必須重覆上述步驟，直到理論建構被驗證或決定放棄驗證工作為止。

一般而言，檢定測驗的建構效度有以下數種方法：

一、內部一致性法

此法與效標關聯效度的分析法相當類似，其將測驗本身的總分視為是「內在效標」，以分析個別測驗題目與整個測驗總分間的一致性。這種方法可以採用下列二種方法來進行：

(一)關係數法：包括校正項目總分相關係數 (corrected item-total correlation)、「 ϕ 相關」(phi coefficient)、「點二系列相關」(point-biserial correlation)、以及「二系列相關」(biserial correlation)，凡相關係數經考驗後達到顯著水準者，即可被保留下來；否則，即刪除淘汰。至於要使用這些統計技術的哪一類，則需視測驗項目的性質而定。校正項目總分相關係數是國內運動相關測驗經常使用的方式，其目的是用來做項目分析，其方式是計算每一個項目與總分之間積差相關的程度，一般的要求在 0.3 以上，且需達到顯著水準。很顯然，項目分析在測驗上必須被檢定，乃是因其會影響一個測驗的效度。 ϕ 相關也曾被使用過，姚漢禱(民 88)在「編製羽球效標參照組合測驗」的研究時獲得效標參照測驗效度係數 (ϕ 相關) 0.48。

其所使用的編製羽球效標參照組合測驗雙向度類別表，及 ϕ 相關之計算公式如公式一。

表一 編製羽球效標參照組合測驗雙向度類別表

	次數	預測評價		
		乙組	甲組	合計
實際評價	乙組	142	6	148
	甲組	31	22	53
	合計	173	28	201

公式一：效標參照測驗效度 (ϕ 相關) $\phi = (142 \times 22 - 6 \times 31) / \{ (148 \times 53 \times 28 \times 173) ^{1/2} \} = 0.477$

(二)團體對照法：即依據受試者的測驗總分高低，將其分成高分組與低分組兩組，然後比較這

兩組受試者在每個測驗項目上的答對百分比值。經過統計考驗後，如果高分組的答對百分比值顯著地高於低分組的答對百分比值的話，即表示測驗項目具有較高的內部一致性；否則，即表示測驗項目是無效的，應該刪除或修改。這種方法經常在國內運動相關測驗的編製上出現，往往擺在項目分析裡，當研究者使用項目分析的內部一致性效標法（*criterion of internal consistency*），就是此類的建構效度之檢定方式。

很顯然地，國內在發展運動相關測驗時，在處理項目分析時，經常使用此內部一致性分析法來印證其測驗項目是否可以有效地區別出受試者的能力。所以說，項目分析也是證明效度的一種方式。但是，對項目分析而言，其能夠證明該份測驗的每一個項目是具有鑑別力，也能夠清楚的辨別某一題目與其他題目之間的關係。但是如果牽扯到項目背後的主要成分或因素時，項目分析就顯得不夠，所以，許多的運動相關測驗在編製過程中，皆會運用探測性因素分析（*exploratory factor analysis, EFA*）來配合項目的診斷。

二、探測性因素分析法

探測性因素分析是目前運動測驗編製者在處理建構效度時最常使用的檢定方法之一。EFA的目的，是用來確定測驗所建構的潛在特質，藉著共同因素的發現，確定這些潛在特質是由哪些有效的測驗項目所構成。EFA利用所蒐集的測驗資料，形成相關矩陣，以統計的方式，例如主因素抽取法（*principal component method*），從一堆測驗項目中抽出少數幾個共同因素（*common factors*），用以代表這一堆項目的共同結構；並且，從中獲得每個項目和每個共同因素間的因素負荷量（*factor loading*），用以代表測驗項目測量共同因素的重要性指標。

接著，再計算每個測驗項目在所有共同因素上的因素負荷量之平方和，即稱作共同性（*communality*），用以代表每個測驗項目的總變異數中被所有共同因素解釋到的百分比。若再用總變異數(1.00)減去共同性，即得到特殊因素（*specific factor*）和誤差因素（*error factor*）所造成的變異數。針對這些共同因素進行命名和解釋，如果它們符合事前所提出的理論建構，此即提供檢定建構效度存在的最佳證據。

EFA可以讓測驗編製者從因素負荷量來判斷個別項目與相對因素的關係，更可以從因素轉軸中所獲得的指標來評量項目的優劣。由於EFA並未在分析之前就對觀察變項的結構做事前的假設預定，因此，一些學者認為EFA的方法只能作為測驗剛建立時結構探討及初步理論形成之用，而無法作為檢定理論因素建構之用。其次，EFA也有下列之限制：

- 1.測驗的個別項目只能分派到一個因素，並只有一個因素負荷量。如果一個測驗項目與兩個或以上的因素構念有關，探測式因素分析就無法處理。
- 2.因素與因素之間的關係必須是全有或全無，及因素必須是完全相關或完全無相關。
- 3.因素分析假定測驗項目與測驗項目之間的誤差是不相關的，但事實上，許多測驗的項目與項目之間的誤差來源是相似的。

所以，當結構方程模式（*structural equation modeling, SEM*）的統計出現後，這些問題便

漸漸地獲得解決。

三、結構方程模式法

SEM 可以說是現行心理、教育與社會學界用以處理建構效度的一個主流。近幾年來，運動相關測驗的研究者已漸漸引用此一統計檢定方法。SEM 主要在處理兩類型的模式：測量模式 (measurement model) 及結構模式 (structural model)。其中測量模式是處理觀察變項如何建構潛在因素的關係，這種關係的分析稱為肯證性因素分析 (confirmatory factor analysis, CFA)。檢定 CFA 的方式是：先以理論來建構測驗項目間的關係，再以實證資料來考驗理論建構與資料間的差距。差距愈小，適合度愈高，效度愈高。

CFA 在 SEM 的發展過程中，一直佔有相當重要的地位。由於 EFA 的限制，早期的 SEM 的先驅，如 Joreskog 等人，便積極的企圖突破這些限制，擴大因素分析模式的使用範圍，最後促成 SEM 的出現。

CFA 具有以下的分析功能：

- (一)可作項目分析，精確的估計個別項目，且將項目分析的概念融合於因素結構的檢測中。
- (二)可檢定個別項目的測量誤差，並且將測量誤差從項目中的變異量中抽離出來，使得因素負荷值具有較高的精確度。
- (三)可依據理論，預先設定項目放置於哪一個因素中，或哪幾個因素中。也就是說，一個項目可以同時分屬於不同因素，並可設定一個固定因素負荷量，或設定任何幾個項目的因素負荷量相等。
- (四)可依據理論，設定某些因素之間具有或不具有相關，甚至於將這些相關設定為相等的關係。
- (五)可對整體因素模式做統計的評估，來瞭解理論所建構的因素模式與所收集資料間的符合程度。所以說，EFA 是一種理論模式檢定 (theory-testing) 的統計方法。

綜合這些功能，一些學者認為 EFA 是檢測建構效度的一種統計利器 (Byrne, 1998; Chiou & Hocevar, 1995; Marsh & David, 1995)。也就是說，EFA 是依據理論的知識及經驗研究來設立一種事先 (a priori) 的連結類型 (因素結構)，以統計的方式檢定這些連結類型的假設，如果假設獲得肯定，則表示測驗具有量好的建構效度。

近年來許多學者已發展了一些統計軟體來處理此一分析。常見的軟體有 Joreskog & Sorbom (1993)所發展的線性結構關係模式(linear structure relationship model, LISREL)，以及 Bentler, P. M. (1989)所發展的 EQS。

四、多特質--多方法分析

要徹底分析建構效度，多特質--多方法分析 (multitrait—multimethod approach, MTMM) 是目前最理想方法，也是心理與教育測驗界流行相當廣泛的方法。MTMM 是由兩位測驗學者 (Campbell & Fiske, 1959)所提出。他們認為一份測驗要具有良好的建構效度，所獲得的資料必

須呈現以下的證據：

- (一)聚斂效度(convergent validity)即一份測驗分數要能夠和其他測量相同理論建構或潛在特質的測驗分數間具有高相關。
- (二)區別效度(discriminant validity)即一份測驗分數也要能夠和其他測量不同理論建構或潛在特質的測驗分數間具有低相關。
- (三)方法效果(method effects)即使用相同方法測量相同特質與使用不同方法測量相同特質所得的相關須比使用相同方法測量不同特質與使用不同方法測量不同特質所得的相關還高。

要使用 MTMM，其前提是：必需要具有兩種以上的測量方法，以及兩種以上的潛在特質，並且，這些潛在特質必須是一樣的，但所用的測量方法必須是不相同的。

Campbell & Fiske (1959) 提出多重特質多重方法矩陣來測量測驗中的聚斂效度、區別效度與方法效果。以下我們提出一個假想式的例子來說明這種矩陣的應用：

首先，在方法上採羽球運動測驗方法和教練評定方法來共同測量羽球選手的三種構念的面向：技術、體能與心理。

表二 假想的多重特質多重方法矩陣

特質		方法一：羽球基本測驗法			方法二：教練評定法		
		技術	體能	心理	技術	體能	心理
方法一： 羽球基本 測驗法	技術	(.92)					
	體能	.52	(.97)				
	心理	.70	.65	(.95)			
方法二： 教練評定 法	技術	(.75)	.25	.20	(.85)		
	體能	.30	(.90)	.20	.48	(.92)	
	心理	.25	.22	(.86)	.65	.60	(.90)

- 表二中的資料包含四類：1.相同方法測量相同體能的特質：【.92, .97, .95；.85, .92, .90】。
 2.相同方法測量不同體能的特質：【.52, .70, .65；.48, .65, .60】。
 3.不同方法測量相同體能的特質：【.75, .90, .86】。
 4.不同方法測量不同體能的特質：【.30, .25, .22；.25, .20, .20】。

一份測驗具有良好的建構效度，須具備以下的檢定結果：使用相同方法測量相同特質與使用不同方法測量相同特質所得的相關（即 1 和 3 的係數）應該是高相關係數（聚斂效度），且須比使用相同方法測量不同特質與使用不同方法測量不同特質（即 2 和 4 的係數）所得的相關還高（方法效果）。而使用不同方法測量不同特質的相關係數是低相關或無相關（區別效度）。由表二的結果顯示符合理論的要求，這些證據證明了羽球運動測驗具有建構效度。

在 SEM 尚未成熟發展時，Campbell & Fiske 所用的方法是相當實用，漸漸地，以 MTMM 矩陣來檢定建構效度也受到相當的批判（Marsh, 1988, 1989; Schmitt & Stults, 1986）。最嚴厲的批判是認為這種檢測方式缺乏客觀的指標，研究者只能依據主觀的方式來判定其所獲得的結果是否具有聚斂效度與區別效度。另外，就犯錯率而言，MTMM 矩陣的每一個相關是個別獨立

計算出來，愈多次的比較，犯錯率（第一類型誤差）就愈高。其計算公式如下：總犯錯率 = $[1 - (1 - \alpha)^n]$ ，其中 n 為檢定的次數， α 則研究者在獨立計算相關時設定的第一類型誤差。此外，一些學者認為 MTMM 矩陣的檢定並未觸及潛在因素構念的檢定，因此無法用以說明潛在因素特質的測量能力 (Lowe & Ryan-Wenger, 1992; Byrne & Goffin, 1993)。

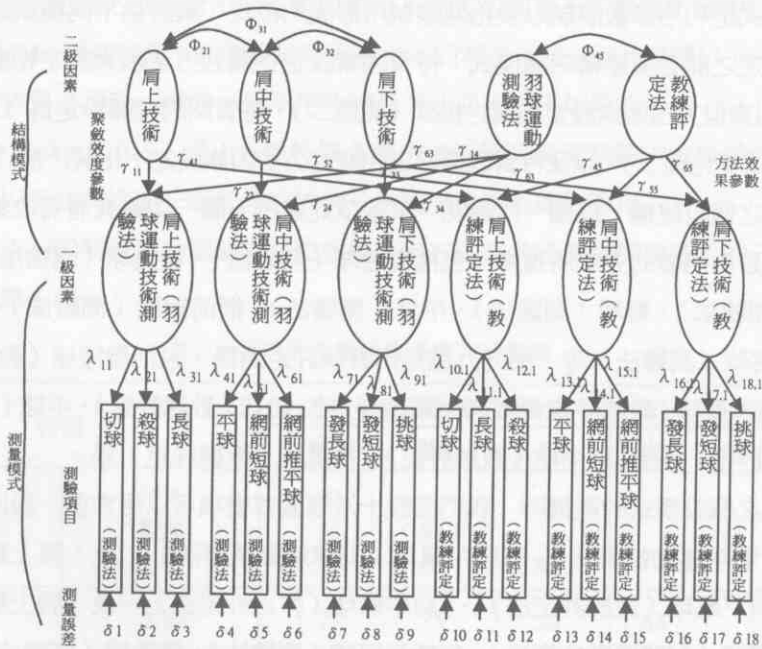
雖然許多替代的方法曾經被提出（要回顧這些方法可以參見 Schmitt & Stults, 1986），來解決這些問題，其中，以 CFA 的檢定方式最受歡迎。對 CFA 而言，其在處理聚斂效度、區別效度、及方法效果的方式是利用參數限制以及相關限制所形成的模式，來評估不同模式之下卡方值變化的結果。在檢定之前必須建構四種模式：特質因素設定相關且方法因素設定相關之模式（見圖一），無特質因素但方法因素設定相關之模式（見圖二），特質間的相關設定為 1.00 且方法因素設定相關之模式（見圖三），以及特質因素設定相關且方法因素設定不相關之模式（見圖四）。

關於四個圖之假設建構，以圖一作說明，其餘依此類推。圖一中研究者需收集十八個觀察變項，其中九個是由測驗的方式所獲得，包括發短球（測驗法）、發長球（測驗法）、長球（測驗法）、切球（測驗法）、殺球（測驗法）、平球（測驗法）、網前短球（測驗法）、網前推平球（測驗法）、及挑球（測驗法）等。另外九個是由教練評定而得，包括發短球（教練評定）、發長球（教練評定）、長球（教練評定）、切球（教練評定）、殺球（教練評定）、平球（教練評定）、網前短球（教練評定）、網前推平球（教練評定）、及挑球（教練評定）等。

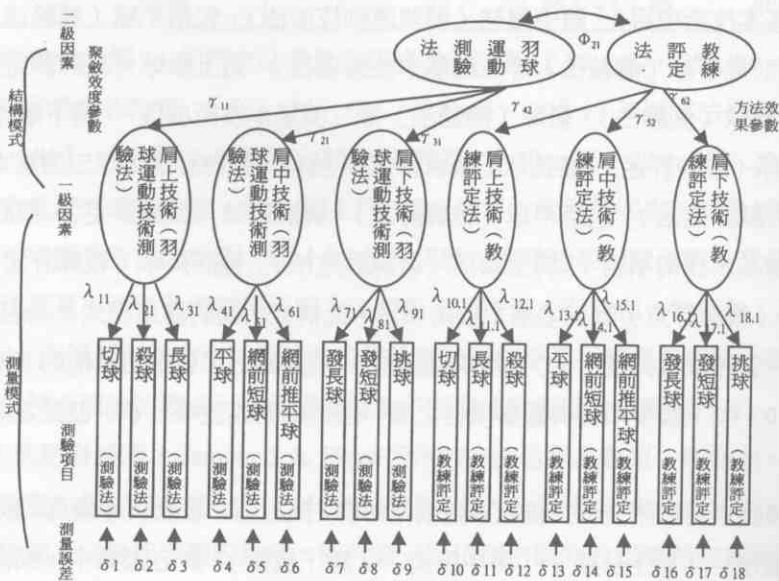
在因素結構之假設模式的建構時，我們假設十八項觀察變項可以用六個一級因素來解釋，即「肩下擊球（羽球運動技術法）」、「肩中擊球（羽球運動技術法）」、及「肩上擊球（羽球運動技術法）」、「肩下擊球（教練評定法）」、「肩中擊球（教練評定法）」、及「肩上擊球（教練評定法）」。「肩下擊球（羽球運動技術法）」包括發短球（測驗法）、發長球（測驗法）、挑球（測驗法）等三個基本技術項目；「肩中擊球（羽球運動技術法）」包括平球（測驗法）、網前短球（測驗法）、網前推平球（測驗法）等三個基本技術項目；「肩上擊球（教練評定法）」包括長球（測驗法）、切球（測驗法）、殺球（測驗法）等三項基本技術項目。「肩下擊球（教練評定法）」包括發短球（教練評定）、發長球（教練評定）、挑球（教練評定）等三個基本技術項目；「肩中擊球（教練評定法）」包括平球（教練評定）、網前短球（教練評定）、網前推平球（教練評定）等三個基本技術項目；「肩上擊球（教練評定法）」包括長球（教練評定）、切球（教練評定）、殺球（教練評定）等三項基本技術項目。此構成所謂的測量模式，其規定為（1）每一觀察項目有一個因素負荷量（ λ ）在其所測量的一級因素上，但是對其他的五個一級因素之因素負荷量是 0。（2）觀察項目與觀察項目之間，所連結的誤差項（ δ ）彼此之間沒有任何關聯存在。

此六個一級因素可以用五個二級因素建構來解釋，此五個二級因素建構為「教練評定法」、「羽球運動測驗法」、「肩上技術」、「肩中技術」、「肩下技術」。此五因素中，前兩個因素為方法因素，後三個因素則為特質因素。「教練評定法」在「肩下擊球（教練評定法）」、「肩中擊球

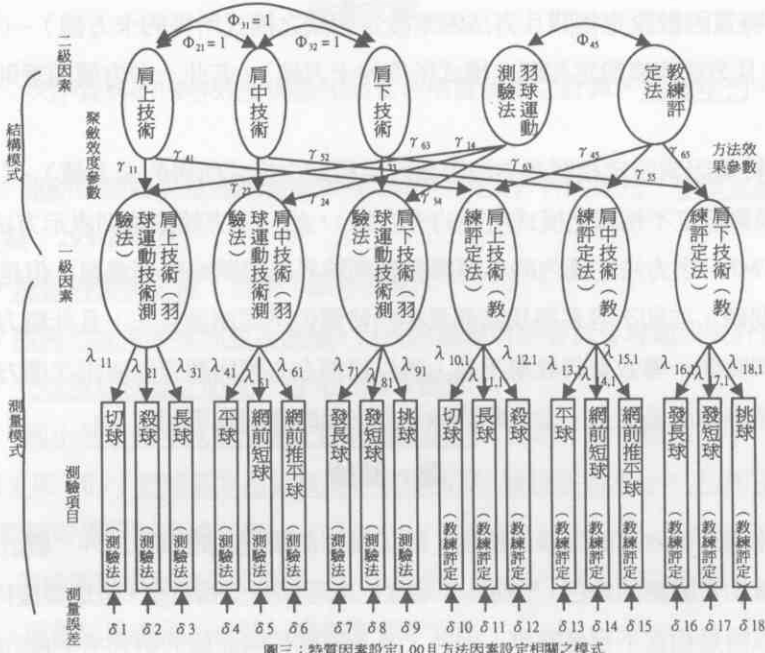
(教練評定法)」及「肩上擊球(教練評定法)」之上各有一個因素負荷量(γ)。「羽球運動技術法」在「肩下擊球(羽球運動技術法)」、「肩中擊球(羽球運動技術法)」及「肩上擊球(羽球運動技術法)」之上各有一個因素負荷量。「肩下技術」在「肩下擊球(教練評定法)」及「肩下擊球(羽球運動技術法)」之上各有一個因素負荷量。「肩中技術」在「肩中擊球(教練評定法)」及「肩中擊球(羽球運動技術法)」之上各有一個因素負荷量。「肩上技術」在「肩上擊球(教練評定法)」及「肩上擊球(羽球運動技術法)」之上各有一個因素負荷量。五個二級因



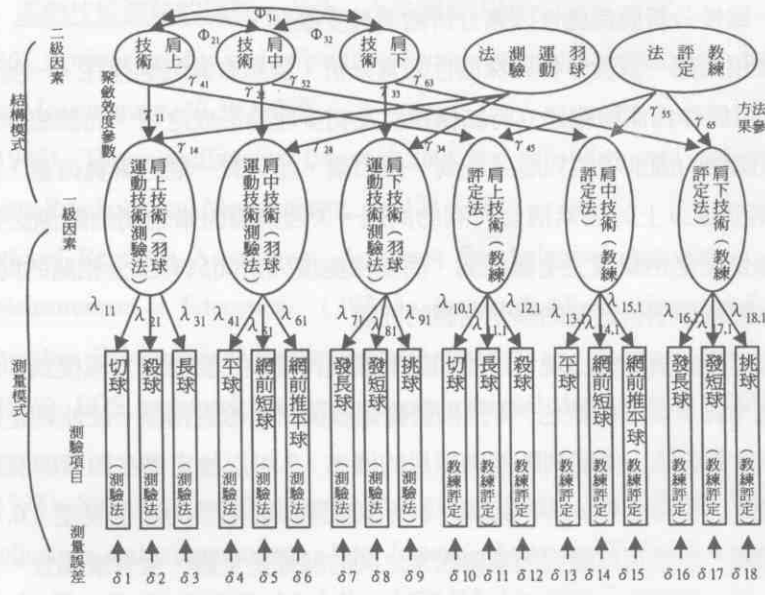
圖一：羽球運動技術測驗之MTMM圖(特質因素設定相關且方法因素設定相關之模式)



圖二：無特質因素但方法因素設定相關之模式



圖三：特質因素設定1.00且方法因素設定相關之模式



圖四：特質因素設定相關但方法因素設定不相關之模式

素之間，也有假設，其中「教練評定法」與「羽球運動測驗法」有相關 (Φ)，「肩上技術」與「肩中技術」有相關、「肩上技術」與「肩下技術」有相關，「肩中技術」與「肩下技術」有相關。其餘皆無相關存在。這些潛在變項與潛在變項間的關係構成所謂結構模式。

聚斂效度、區別效度，以及方法效果的檢定方式如下：

- 1.聚斂效度：(特質因素設定相關且方法因素設定相關之模式所得的卡方值) - (無特質因素但方法因素設定相關之模式所得的卡方值)。若相減的卡方值達到水準，則聚斂效度成立)。

- 區別效度：(特質因數設定相關且方法因素設定相關之模式所得的卡方值) - (特質間的相關設定為 1.00 且方法因素設定相關之模式所得的卡方值)。若此一卡方值顯著則表示區別效度成立。
- 方法效果：(特質因素設定相關且方法因素設定相關之模式所得的卡方值) - (特質因素設定相關且方法因素設定不相關之模式所得的卡方值)，此一卡方值顯著則表示方法效果成立。

雖然，MTMM 之方法在國內的心理與教育測驗界已經開始受到重視，但是此一方法存在一個相當大的問題，亦即不容易尋找到測量相同特質但不同測量方法，且此種方法的樣本收集相當耗費成本與時間，導致資料收集不易。所以目前在心理與教育界使用次種方法來檢定建構效度的研究，依然是鳳毛麟角。而運動測驗在此方面的應用幾乎闕如¹。

陸、結論

國內運動相關測驗效度的發展與檢定，除了受到測驗理論的影響之外，統計技術與軟體的不斷提升扮演著相當關鍵的角色。早期外部效度的處理相當受到重視，但是要獲得適切、可靠、客觀、可用的效標是相當不容易的事。因此，許多運動相關測驗的研究者則傾向於採用內部效度，其中內部一致性分析與探測性因素分析兩者最受歡迎。

當然，在使用內部一致性分析與探測性因素分析，依然有其限制存在，一些學者認為探測性因素分析：1.只能作為測驗剛建立時結構探討及初步理論形成之用，而無法作為檢定理論因素建構之用。2.測驗的個別項目只能分派到一個因素，並只有一個因素負荷量，而無法處理一個測驗項目與兩個或以上的因素構念有關的問題。3.因素與因素之間的關係必須是全有或全無，及因素必須是完全相關或完全無相關，而無法處理因素間只有部分相關的問題。4.其亦無法處理測驗項目與測驗項目之間的誤差相關的問題。

隨著結構方程模式統計的出現，這些問題漸漸獲得解決。對結構方程模式中的肯證性因素分析在效度的應用具有以下特色：1.提供整體假設模式之檢定指標，使能判定建構效度存在與否。2.提供因素負荷值，使能判斷個別項目的優劣。3.可以檢定潛在因素與測驗項目之間的關係。4.可以讓同一項目分屬於不同因素。5.可以偵測測驗項目的測量誤差。6.可以檢定因素與因素之間的相關，並且可以設立因素與因素之間的關係是全無、全有或部分。

這些優點使得國內運動相關測驗的研究者漸漸的注意到此一統計方式。可以說，以肯證性因素分析來處理建構效度的問題是當今運動測驗發展的一個新趨勢。本篇文章的發表，正是希望透過對此一技術的介紹，能夠引起更多的共鳴，讓運動相關測驗的發展更上一層，嘉惠體育教育與運動訓練。

參考文獻

- 余民寧(民82): 次序資料的內容效度係數和同質性信度係數之計算。測驗年刊, 40 輯, 199-214 頁。
- 季力康(民83): 運動目標取向量表的建構效度-驗證性因素分析的應用。中華民國體育學會體育學報, 18 輯, 299-310 頁。
- 林正常(民78): 運動科學與訓練。台北: 銀禾文化。
- 姚漢禱(民88): 編製羽球效標參照組合測驗。行政院國家科學委員會專題研究計劃成果報告。
- 孫宜芬(民73): 羽球技能測驗之研究。台北: 師大體育。
- 紀世清(民88): 國小羽球基本技術測驗之建構效度研究。台北: 師大書苑。
- 紀世清、黃芳銘(民88): 羽球基本運動技術之建構效度--驗證性因素分析之應用。1999 年國際大專運動教練科學研討會論文集。
- 張永文(民86): 羽球運動選材的基本要素。中華體育, 11 卷, 3 期, 72-80 頁。
- 許樹淵(民86): 運動科學導論。台北: 偉彬體育研究社。
- 張至滿(民75): 運動技能測驗的編製。台北: 水牛圖書出版。
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. Educational and Psychological Measurement, 40, 955-959.
- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. Educational and Psychological Measurement, 45, 131-142.
- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1985). Standards of educational and psychological testing. Washington, DC: American Psychological Association.
- Bentler, P. M. (1989). EQS structural equations program manual. Los Angeles: BMDP statistical Software.
- Byrne, B. M. (1998). Structural equation modeling with LISREL, PRELIS & SIMPLIS: Basic concepts, application, and programming. New Jersey: Lawrence Erlbaum Association.
- Byrne, B. M. & Goffin, R. D. (1993). Modeling MTMM data from additive and multiplicative covariance structures: An audit of construct validity. Multivariate Behavioral Research, 28, 67-96.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.
- Chiou, H. J., & Hocevar, D. (1995). Examination of population-invariant construct validity in the multipopulation-multitrait-multimethod design. Paper presented at the 1995 Annual Convention

- of the American Psychological Association, New York, NY.
- Gronlund, N. E., & Linn, R. L. (1990). Measurement and evaluation in teaching (6th ed.) . NY : Macmillan.
- Joreskog, K. G., & Sorbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Chicago, IL: Scientific Software International.
- Loevinger, J.(1957). Objective tests as instruments of psychological theory. Psychological Reports, 3, 635-694.
- Lowe, N. K., & Ryan-Wenger, N. M. (1992). Beyond Campbell and Fiske: Assessment of convergent and discriminant validity. Research in Nursing and Health, 15, 67-75.
- Marsh, H. W., & David G. (1995). Latent variables models of multitrait-multimethod data. In Hoyle, R. H. (Eds), Structural equation modeling: Concepts, issues, and application. (pp.177-198). Thousand Oaks : Sage Publication.
- Mehrens, W. A., & Lehmann, I. J. (1991). Measurement and evaluation in education and psychology (4th ed.). New York: Holt, Rinehart & Winston.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Schmitt, N., & Stults, D. M. (1986). Methodological review: Analysis of multitrait-multimethod matrices. Applied Psychological Measurement, 10, 1-22.

THE STUDY OF DEVELOPMENT OF VALIDITY IN SPORTS-RELATED MEASUREMENT AND TESTING TREND

Fang-Ming Hwang

National College of Physical Education and Sports

Wen-Chuan Chen & Kuei-Mei Cheng

Tajen Institute of Technology

ABSTRACT

Applying measurement theory to solve the validity issues in sports-related measurement has gained much emphasis in academic field in Taiwan during the recent years. For years, external validity (criterion-related validity) has been used to verify the testing in sport-related measurements. Nevertheless, this method is not without limitations. Construct validity, for example, depends on internal consistency analysis and exploratory factor analysis. These analyses are able to handle the issues such as individual item specification, factor loading, and non-or-all relationship between factors. However, issues such as testing of factor-related partial and whole theoretical structure may not be suitable under those circumstances. Due to its limitations, many researchers have turned to adopt internal validity instead.

As the advance of statistic technique and computer software, confirmatory factor analysis of structural equation modeling has already been used to verify construct validity in sport-related measurement academically in Taiwan. This method is able not only to correct the deficiency when apply internal consistency analysis and exploratory factor analysis to the measurement, but also to test constructed theory-factor modeling as well. These are the essence of construct validity by definition and its requirement. Hopefully, this introductory study of the new trend in sport-related measurement will get much more attention and scholars' involvement in order to further advance of development in this statistic technique.

Key words: measurement, validity, structural equation modeling, exploratory factor analysis, confirmatory factor analysis, internal consistency analysis.