

## 第八講 體育課程設計之 教學評量

By  
Mei-Yao Huang, Ph.D

### 什麼是評量？

- ◆各位同學，您認為評量的定義是什麼？

### 測驗與評量溯源

- 漢朝發展出非常複雜和精緻的步驟來選取公職人員。惟漢代舉拔官員的「鄉舉里選」，其權力實操諸地方官吏之手，成就東漢末年以下所謂「門第」現象的產生。
- 魏晉時代雖然創設了所謂的「九品中正制度」，但其用意仍與漢朝之鄉舉里選相差不遠，而中正的九等表終不免以門第高下為標準。

### 測驗與評量溯源(續)

- 隋唐時代的科舉制度為一套完整的取士制度。
- 西方早在十三世紀大學畢業考試就開始使用口試。

### 測驗、測量與評量

- ◆ 測驗 (test) 乃是為了測量一組行為樣本的一種工具 (instrument)、或有系統的程序 (systematic procedure)。
- ◆ 測量 (measurement) 係獲致個體具有某種特質程度之數字描述的過程。
- ◆ 評量 (evaluation) 係蒐集、分析、與解釋資訊，以決定學生達成教學目標之程度的系統化過程。

### 教學評量的目的

- ◆ 讓「教學者」知道學生學會了多少
  - ◆ 告訴學生他們的已知與未知
  - ◆ 告訴老師學生的「弱點」
  - ◆ 引導學生設定實際且可達到的目標
  - ◆ 協助學生察覺自己的進步、熱衷學習
  - ◆ 協助「教學者」知道教學的效益
  - ◆ 協助「教學者」規劃教學的範圍與內容
- ↑ 目前普遍的想法  
↓ 可以努力的方向

## 教學評量的功能

- 瞭解學生潛能與學習成就，以判斷其努力程度。
- 瞭解學生學習的困難，做為補救教學及個別輔導之依據。
- 評估教師的教學成效，做為改進教材教法之依據。
- 獲悉學習進步情形，激發學生學習的動機。

## 教學評量的類型

- ◆ 安置評量 (placement evaluation)
- ◆ 形成性評量 (formative evaluation)
- ◆ 診斷性評量 (diagnostic evaluation)
- ◆ 總結性評量 (summative evaluation)

### 安置評量

- ◆ 係指一個新生初到學校新環境，要安置在適當班級中接受教學，或一位教師到一個新的班級，想瞭解這些學生對於新的課程已有多少的相關知能而做的教學前評量。

### 形成性評量

- 常以隨堂測驗、非正式考試或單元測驗，測試學生對某些概念原則或某個單元精通的程度，以便隨時發現問題，提供給學生和教師校正及回饋的機會。
- 一方面避免學生困難的累積而降低學習動機，一方面提供教師調整教學方法與教學進度。

### 診斷性評量

- ◆ 在形成性評量過程中發現學生學習困難，並診斷其學習困難的原因，以便提供補救教學或個別化教學。

### 總結性評量

- 一個單元教學結束或幾章課程結束或學期結束，為瞭解學生對教材內容精通的程度，以及學生的行為表現是否已經達成教學目標而進行的教學評量。

## 教學評量的類型

- 常模參照評量 (norm-referenced evaluation)
- 效標參照評量 (criterion-referenced evaluation)

## 常模參照評量

- 將評量結果獲得之個人分數拿來與參照點做比較，即可看出其分數在團體中的相對位置，而判定其成績的相對優劣，此種評量謂之常模參照評量。

## 效標參照評量

- 若教學目標訂得很明確，其終點行為亦訂得很具體，並設定一些標準；達到一定標準者代表學習「精熟」，未達既定標準者代表學習「未精熟」，此種評量謂之效標參照評量。

## 測驗應具備的特性

- 正確性 (validity)
- 可靠性 (reliability)
- 鑑別性 (discrimination)
- 客觀性 (objectivity)
- 比較性 (comparative)
- 實用性 (usability)

## 正確性 (validity)

- 係指一個測驗能夠測量到它所要測量的能力程度。
- ◆ 傳統定義 指測驗分數的正確性。指一個測驗能夠測量到他所想要測量的特質的程度。
  - 數學測驗的分數若主要是由語文能力來決定，這樣的測驗就無效(效度低)。
- ◆ 以上的定義說明一個測驗只有一個效度，事實不然，效度可依不同的方法得到不同效度。
- ◆ 較滿意的定義一個測驗在使用目的上的有效性。更具體的是指測驗能夠達到某種目的的程度。

## 可靠性 (reliability)

- 係指一個測驗所測得的分數的可靠性或穩定性。
- 亦即指同一群人在同一測驗上數次測量結果的一致性。
- 大凡一個測驗如無信度則無效度，但有信度未必有效度。
- 信度並不就是效度的保證，可是信度卻是效度的一個基本條件。
- 如果兩次測驗的間隔時間適度，則學生前後完成同一工作所需的時間大致相等，亦即前後兩次測驗的工作速度是一樣的。

- 測驗的鑑別性，目的在鑑別各個受試者不同等級的能力。
- 一個有鑑別力的測驗，內容包括各種難度的題目，從容易到繁雜的都有，同時是容易題目到繁雜題目依難易程度層次排列。

## 比較性 (comparative)

- 測驗所得結果必須與常模 (norm) 或是既定的行為目標比較。

## 客觀性 (objectivity)

- 係指測驗實施手續的客觀化而言。
- 測驗的客觀性與可靠性有密切的關係，測驗的實施手續如不客觀，測驗的結果一定不可靠。
- 理想的測驗是即使數人評閱同一測驗所得結果的成績是一樣的。
- 評量時盡可能不要讓個人的主觀因素滲入測驗結果。

## 實用性 (usability)

- 一個測驗所需的時間、人力和費用，必須為使用者的客觀環境所許可，這個測驗才有被施行的機會。採用個別化策略的測驗，往往缺乏實用性。

