

體育評估工具的信度和效度：  
技能及情意領域

**Assessment Tool: Reliability & Validity of  
Motor & Affective Domains**

---

香港浸會大學體育學系

周碧珠教授

Prof. Bik C. CHOW  
HKBU, Dept. of Physical Education

# 內容

---

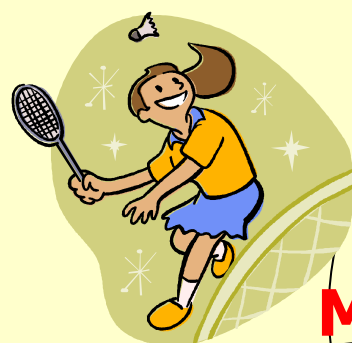
## 第一部份

- (甲) 傳統體育評估之信度及效度
- (乙) 非傳統體育評估之信度及效度

## 第二部份

- 體育成績等級

# 學習領域 LEARNING DOMAINS



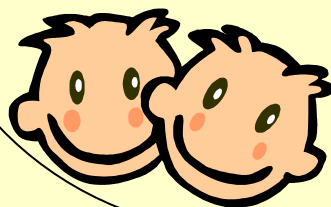
技能與體適能

**MOTOR & FITNESS**



情意

**AFFECTIVE**

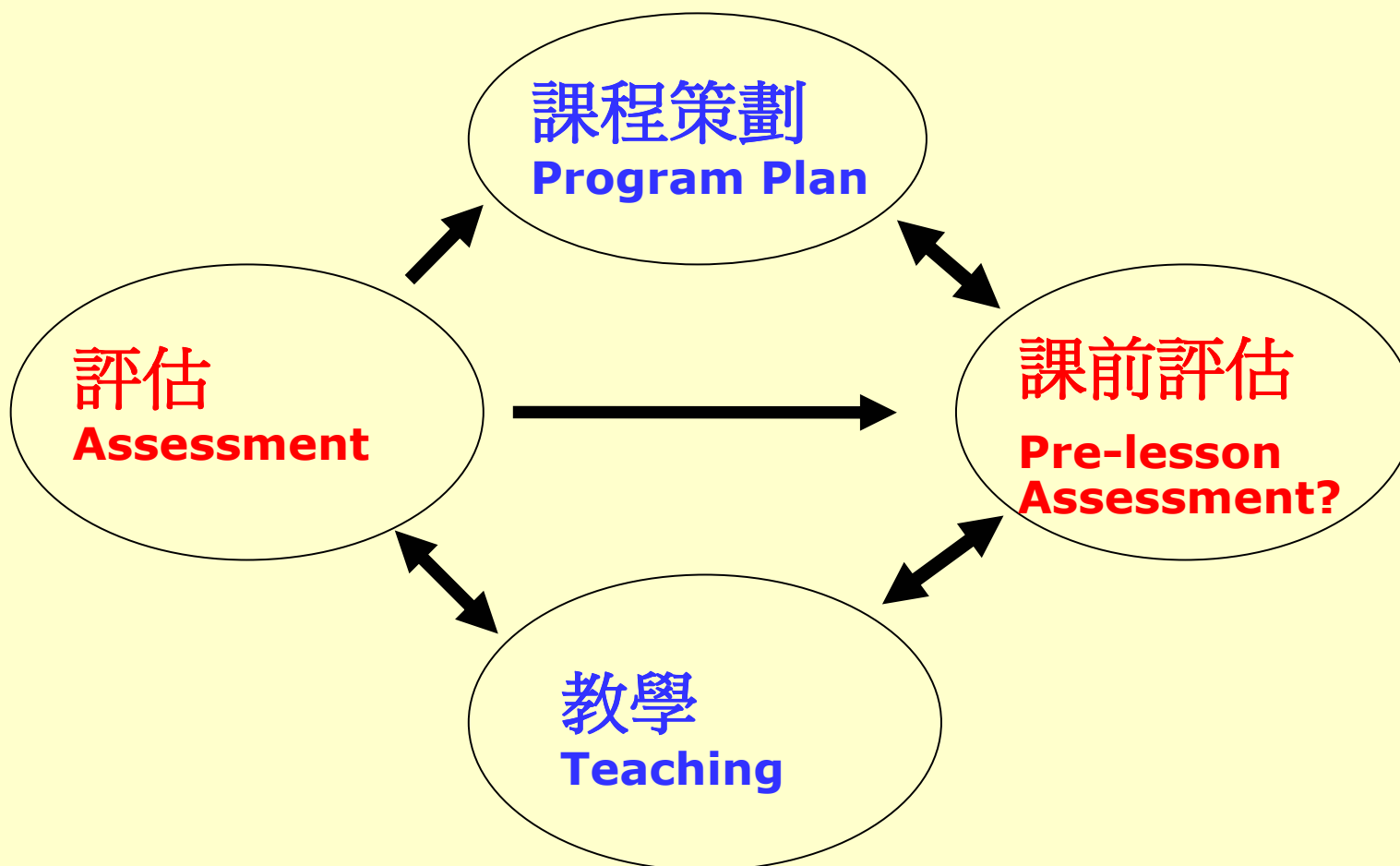


認知

**COGNITIVE**



# 評估模式



# 過去十年體育評估趨勢

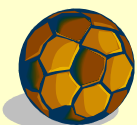
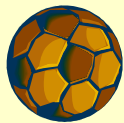
---

- 訂立國家學習目標與學習成果。
- 學習理念：照顧個人差異，培育思考等。
- 評估：加入非傳統評估(*alternative assessment*)或真實評估(*authentic assessment*)，評估內容在真實處境(如比賽)中進行或評估題目貼近現實(如專題研習)。
- 評估由常模為本(*norm-referenced*) 漸轉至評準為本(*criterion-referenced*) 。
- 亦強調表現為本(*performance-based*)，即評估技能外，亦需包含社交責任、價值觀等。

---

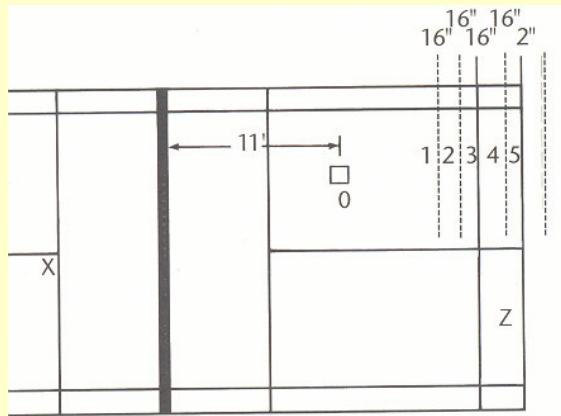
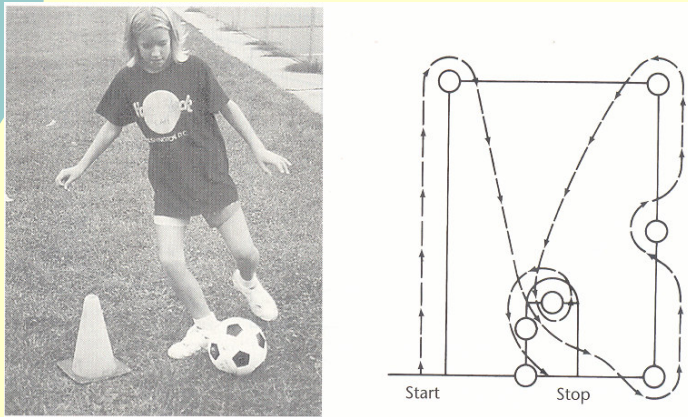
# 第一部份: 信度及效度

## (甲) 傳統體育評估



# 傳統評估：技能

## 成果評估



## 過程評估：評等量表 (rating scale)

羽毛球: 高遠球	經常	有時	沒有
1. 引拍: 手臂提高向後	3	2	1
2. 重心由後腳移至前腳	3	2	1
3. 最高點擊球	3	2	1
4. 擊球動作(串連上臂、 下臂及手腕)	3	2	1
5. 跟進動作(球拍揮至身體 另一下方)	3	2	1
總分: _____ / 15			

# 傳統評估：情意

- 利用已設計妥善之量表：評估體育態度，興趣，自我觀念，體育精神等。

## 體能活動之感受

(黎克型量表)

SD D U A SA

- |                   |   |   |   |   |   |
|-------------------|---|---|---|---|---|
| 1. 我期待參予體能活動      | 1 | 2 | 3 | 4 | 5 |
| 2. 缺席一天體能活動感到鬆了口氣 | 1 | 2 | 3 | 4 | 5 |
| 3. 體能活動是愉快的       | 1 | 2 | 3 | 4 | 5 |
| 4. 體能活動是一天最興奮的時刻  | 1 | 2 | 3 | 4 | 5 |

等共12題



# 評估工具：信度

---

信度 (**Reliability**): 評估量度是一致

## 信度類別

- 內在一致性信度 (*internal consistency*)
- 穩定信度 (*stability*)
- 評等者間信度 (*inter-rater reliability*)

確保信度之傳統評估指引

# 信度類別 1. 內在一致性信度 (a)

## 1a. <技能>於同一測試中有最少二次試驗分(trials)

1(a)	試驗一	試驗二	試驗三
甲	5	6	7
乙	3	3	4
丙	7	6	8

利用統計: 組內相關 (*Intraclass Correlation Coefficient, R*) 找出信度係數 (Reliability Coefficient)

- 係數愈接近1.0顯示高信度, 0.8 以上為可接受信度

{SPSS: Analyze --- Scale --- Reliability Analysis [Statistics: intraclass corr  
coeffic.]{SPSS same as test-retest} 例題答案:  $R = 0.97$

# 信度1. 內在一致性信度(b)

---

## 1b. <體適能>例題:

皮摺量度(毫米)	1.	2.	3.
三頭肌	8	10	16
小腿內側	14	14	12

# 信度1. 內在一致性信度(c)

---

## 1c. <情意評估>於評估表上有最少二條題目

### 體能活動之感受

	SD	D	U	A	SA
1. 我期待參予體能活動	1	2	3	4	5
2. 缺席一天體能活動感到鬆了口氣	1	2	3	4	5
3. 體能活動是愉快的	1	2	3	4	5
4. 體能活動是一天最興奮的時刻	1	2	3	4	5

等共12題

{SPSS-Analyze-Scale-ReliabilityAnalysis (Model: Alpha)}  
- Cronbach's Alpha: 信度係數， 0.8 以上為可接受信度

# 信度2. 穩定信度

---

## 2. 穩定信度可由覆測 (*test-retest*) 來檢定

	第一天	第二天
學生甲	9	9
乙	1	2
丙	8	7

利用統計: 組內相關, 等同**1a**方法

- 係數愈接近1.0顯示高信度, 0.8 以上為可接受信度

例題答案:  $R = 0.98$

(*inter-rater reliability*)

3. <技能評估> 評等量表 (*rating scale*):

動作姿勢過程評估

體操：分腿騰越木箱 (3=經常 2=間中 1=沒有)

	評分員1	評分員2
1. 助跑速度	3	3
2. 雙腳着踏板	3	2
3. 推手動作	2	2
4. 騰空動作	2	1
5. 直腳分腿動作	2	2
6. 穩定着地	1	1

{信度係數：組內相關-等同1a方法，評分員分替代試驗分}

例題答案:  $R = 0.87$

# 確保信度之傳統評估指引

---

- 提供清晰評估指示及步驟。
- 儀器：檢查儀器之準確性(血壓計、脂肪磅)。
- 給予學生足夠練習及熱身。
- 於覆測之間給予足夠恢復時間，如一至三天。
- 若評等量表有太多項目(如：動作有8項姿勢分)，可考慮刪減。
- 隨機抽選部份學生參予覆測，檢查信度。

# 評估工具：效度

---

**效度(Validity)：** 內容能夠量度其評估之目的  
(例如：仰臥起坐是測量腹肌之肌力及肌耐力)

## 效度類別

- 1. 內容效度 (*content validity*)
- 2. 評準效度 (*criterion validity*)
- 3. 構念效度 (*construct validity*)

確保效度之傳統評估指引



# 效度1. 內容效度

---

## 1. 專家檢定內容是否符合評估目的

- <技能評估> 向牆對打測試：是否能夠真正反映技術能力？
- <情意評估> 由不同教師判斷內容是否符合目的。

# 效度2: 評準效度

---

## 2. 以新設計評估對比另一評準評估

- <技能評估>

- **2(a)** 對比: 評等量表評估動作姿勢

- 新設計 (如足球射球分)

- {計算二組分之相關係數: *Pearson Correlation*}

效度係數超過0.80為可接受，0.90-0.99為高

# 效度2: 評準效度

---

## 2(b) 比賽等級

- 另一評準對比: 比賽排名
- 新評估: 個人/雙人運動技術成果評估
- {計算相關係數: *Spearman Correlation*}

## 2(c) 預設評準 (例)

- 另一評準對比: 20米來回跑(*pacers/beep test*)
- 新評估: 9分鐘跑
- {計算二組分之相關係數: *Pearson Correlation*}

# 效度3. 構念效度

---

## 3a. 高低組別能力差異

- <技能評估> 技能平均值：校隊應高於非校隊
- {統計：t 檢定 - *independent t-test*}

## 3b. <情意評估>

- 二份類似構念問卷有明顯高相關值
- 二份非類似構念問卷只有低相關值

# 提升效度之評估指引

---

- 要有效度，必先有信度。但確保信度並不足保證效度。
- 要確保內容效度，評估工具必須按清晰學習成果而設計。 內容 ↔ 目的
- 評估前應給予學生練習及準備。
- 評估項目較多能提升信度，亦確保效度。  
如籃球技術評估包括三項：運球、傳球、射球比單一項為較合適。



---

# 第一部份：信度及效度

## (乙) 非傳統體育評估

# 非傳統評估

---

- 或稱爲**真實評估** (*authentic assessment*)
- 特色：強調評估內容近似或符合真實環境。
- 專注於過程及成果，強調深層思考及學習。
- 包括
  - 自我、同儕、教師評估
  - 表演(設計自由體操或舞蹈串連動作)
  - 專題研習、習作資料夾、學習日誌等

# 非傳統評估: 技能領域

評量指引表(Rubrics): 用文字詳細地形容不同等級之表現。

## 技能應用 (*Application of Skill*)

級別4: 能一致性(75%時間)、合法及有效地運用合適技能，  
沒有或只有少量技巧差誤。[**熟練**]

級別3: 能正確、合法但較少慣常(50%時間)地運用出良好技巧[**使用**]

級別2: 使用較少合法及較多不一致的技巧(15-50%時間)，  
部份技術運用出現很多錯誤。[**尚能控制**]

級別1: 很少有效地運用技巧 (<15%時間)，技巧不正確。[**未能控制**]



# 非傳統評估：情意領域 評量指引表 (*Rubrics*)

## 個人及社交責任

級別4：能表現出自發之領袖才能，參與活動時積極投入，有自控及尊重公平競賽精神，為其他參與活動者提供正面經驗。

級別3：一貫積極地投入參與，有自控及尊重公平競賽精神。

級別2：需要別人提醒及上述行為並不一致。

級別1：情緒自控能力薄弱及不懂得尊重他人。

# 非傳統評估: 情意領域

## 評準為本

- **學習目的:** 學生於新設計球類活動中，能表現出團隊合作精神。
- **評準為本:** 觀察**10**分鐘比賽，學生能表現出理想及不理想行爲之比例-不少於**4:1**。

### 理想行爲

- 支持鼓勵隊友 \_\_\_\_\_
- 儘量傳球給予各隊友 \_\_\_\_\_
- 遵守規則 \_\_\_\_\_
- 接受對隊裁決 \_\_\_\_\_
- 其他: \_\_\_\_\_

### 不理想行爲

- 沒有鼓勵隊友 \_\_\_\_\_
- 故意放棄傳球 \_\_\_\_\_
- 不守規則 \_\_\_\_\_
- 質疑不利之裁決 \_\_\_\_\_
- 其他: \_\_\_\_\_

共(甲): \_\_\_\_\_ 共(乙): \_\_\_\_\_ 比例(甲/乙)= \_\_\_\_\_

# 非傳統評估指引：確保信度

---

**內在一致性信度：**評量指引表要確保能夠清晰形容所觀察行為，教師並需熟習指引。

- 先試行練習，重覆檢定。
- 可考慮減少評審表內的項目，務求能熟練地評估。
- 評估方式必須一致：例如評定傳接球，不是接發球。
- 觀察評估過程：應留意觀察位置(距離及角度)。
- 學生應清楚知道評估細則及要求。

**穩定信度：**抽選部份學生進行覆測。

**評等者間信度：**當採用同儕評估時，必須提供足夠之訓練。

# 非傳統評估指引：確保效度

---

- **內容效度**：清晰訂立評估內容目的及其定義，再按其定義設計評量等級表及工具。

# 非傳統評估：分析

- **評估**：學生之社交責任（願意對別人負責）
- **方式**：於體操課堂，觀察**15**分鐘練習時段，  
課堂已強調如何幫助同伴作保護及協助。
- **評審**：5=經常 4=常常 ... 1=沒有
- **內容**：觀察保護技巧；給予同伴保護；根據指示作保護；  
保護時作出修正並能協助同伴

- |   | 有 | 沒有 |
|---|---|----|
| ○ <b>信度</b> ：兩位不同教師能給予類似評分              | — | —  |
| ○ <b>效度</b> ：評估學生對別人負責                  | — | —  |
| ○ <b>量度</b> ：評準為本之分析標準是根據<br>行為領域的內容而釐定 | — | —  |

## 第二部份

---

### 體育成績等級之評估 (*Grading*)

# 體育成績等級之評估

---

- 構成體育成績等級之部份應能反映學習之目的及成果。

考慮因素：

- 評估過程？ 成果？ 進度？
- 學習領域之比重： 技能？體適能？情意？認知？
- 成績等級： 常模為本？ 評準為本？

# 體育成績評估：例

---

## 校內評估(例)： 中四

- 技能(一) ----- 20%
  - 射球 (10%)
  - 傳球 (10%)
- 技能(二) ----- 20%
  - 過程：姿勢(非傳統評估) (20%)
- 體能： 2400米跑 ----- 20%
- 情意 (非傳統： 個人及社交責任) ----- 10%
- 情意： 態度(出勤、主動學習) ----- 5%
- 情意： 校隊成員 ----- 5%
- 認知： 筆試- 技能(一)(二)規則及(二)裁判 -- 20%



# 總結

---

- **信度**：受教師知識及經驗所影響。
- **效度**：受評估工具之內容所影響。
- **非傳統評估**：一般要自行設計，熟練使用評量指引表可提升信度，而效度方面，主要可建立內容效度。
- **成績等級**：按學校宗旨、特色、各教員之理念及共識而訂立。



本節完結

# 參考

## 傳統評估

Baumgartner, T.A. & Jackson, A.S. (1995). *Measurement for evaluation*. Dubuque, IA: WCB.

---

Collins, D. & Hodges, P. (1978). *A comprehensive guide to sports skills test and measurement*. Springfield, IL: Charles C. Thomas.

Lacy, A.C. & Hastad, D.N. (2003). *Measurement & evaluation in physical education & exercise science*. San Francisco: Benjamin Cummings.

## 非傳統評估

Doolittle, S., & Fay, T. (2002). *Authentic assessment of physical activity for high school students*. Oxon Hill, MD: NASPE

Melograno, V.J. (1996). *Designing the physical education curriculum*. Champaign, IL: Human Kinetics. [chapter 6: pp.151-194]

Rink, J.E. (2006). *Teaching physical education for learning*. Boston: McGraw-Hill. [chapter 12: pp.266-295]

## 互聯網資料(評量指引表)- June 10, 2006

<http://www.pecentral.org/assessment/alternative/underh...>

<http://whs.dist214.k12.il.us/www/academics/pe/rubrics/advpe.htm>



---

互聯網資料(非傳統評估) - June 14, 2006

評量指引表

<http://www.pecentral.org/assessment/alternativeassess.asp>

<http://whs.dist214.k12.il.us/www/academics/pe/rubrics/advpe.htm>

「體育學習成果架構」2005，課程發展部體育組。

(項目: 球類、田徑、游泳、獨木舟、基礎活動)

<http://cd1.emb.hkedcity.net/cd/pe/tc/tep/LOF2005/LOF2005e.htm>