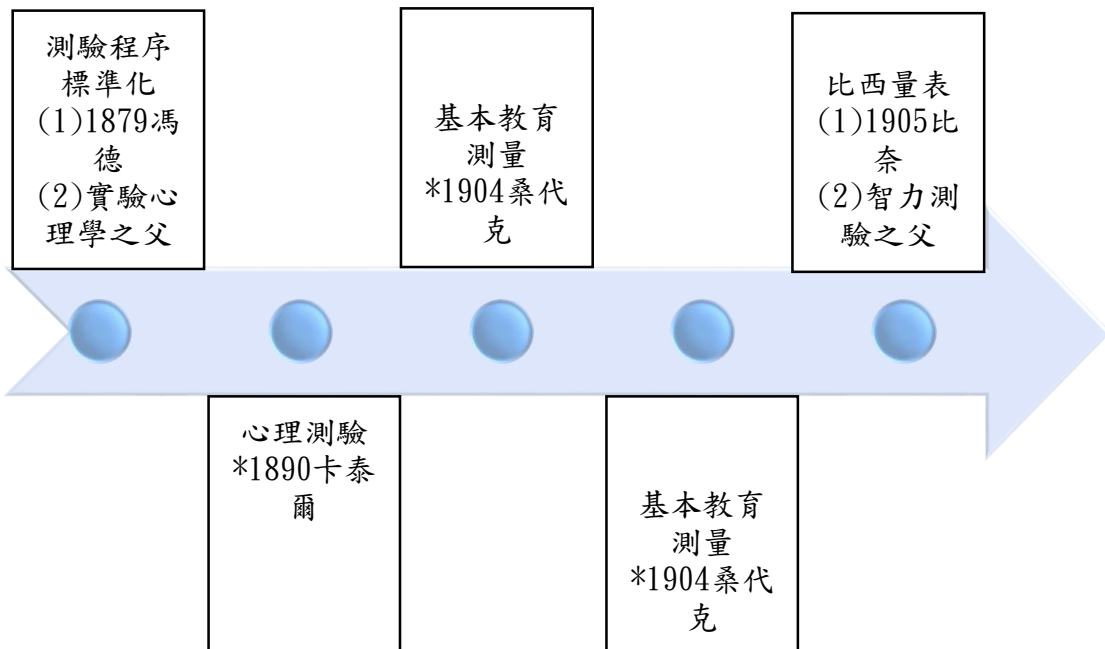


中等學校課程與教學：測驗與評量

(一) 基本概念

	意義	特徵	實施方式
測量	以測量工具來描述所測事物的屬性	不涉及價值判斷	檢核表、問卷 身高測量
測驗(考試)	考試；評量工具	強調客觀性	智力測驗、性向測驗等
評量	教師蒐集學生學習資料的過程	需要價值判斷	實作、真實評量 變通性評量 學習單、作業、日記等...
評鑑	鑑定、判斷的過程	需要價值判斷	形成、總結、安置、 診斷性評量

(二) 測驗的發展



(三)理論背景

	學習理論	評量理念
實證(行為)主義	強調 知識與經驗 的一致性	<ul style="list-style-type: none"> ◇ 重視信、效度 ◇ 有效公平客觀的評量 ◇ 標準化的紙筆化評量
理性(認知)主義	強調 概念間 的連貫性	<ul style="list-style-type: none"> ◇ 看出學生的心智能力成長 ◇ 評量實作、長程表現 ◇ 以多元評量衡量學生的進展 ◇ 實作評量、卷宗評量
社會(歷史)主義	知識與 人與人，人與物 互動 中建構 ◇ 情境主義	<ul style="list-style-type: none"> ◇ 評量學生參與學習的品質 ◇ 評量學習歷程(社會互動) ◇ 學生自省自評 ◇ 卷宗評量

(四)測驗類型

分類標準	測驗形式	意義	測驗內容
根據不同教育目標	認知測驗	測量 個人思考、知識、問題解決 能力	智力測驗 性向測驗 成就測驗
	情意測驗	測量 個人動機、情緒、價值觀、自我觀念 特質	社會計量測驗 人格測驗 投射測驗
	技能(動作)測驗	測量 手、腳、腦的協調反應 多半使用 實作測驗 方式進行	檢核表 評定量表

分類標準	測驗形式	意義	測驗內容
根據不同編製過程的標準化程度	標準化測驗	測驗專家根據 專業程序 設計出的測驗 有一定的編制程序(試題取樣，信效度等)	智力測驗 性向測驗
	非標準化測驗	又稱 教師自編測驗 ，教師用非正式方式，依自己教學需要所設計的測驗(個別差異)	教師自編測驗

分類標準	測驗形式	意義	測驗內容
根據測量分數的解釋方式	常模參照測驗 NRT	區分學生間彼此間的成就水準差異	學科成就競試 標準化成就測驗
	標準(效標)參照測驗 CRT	<u>達到標準就過關</u> 找出學生已學會及尚未學會的原因及困難	平時考 診斷性教師自編測驗

分類標準	測驗形式	意義	測驗內容
測驗上的反應型態	最大表現測驗	假定個人以最佳表現應試分數越高，能力越佳	智力測驗 成就測驗 性向測驗
	典型表現測驗	假定受試者誠實回答問題沒有好壞分數，有興趣就是有興趣，沒興趣也亦可	人格測驗 興趣測驗 態度測驗

分類標準	測驗形式	意義	測驗內容
根據測驗所給的時限	速度測驗	測驗有限時間內，學生反應的速度。 難度通常不高，因此不常出現於教育測驗中。	不常出現於教育測驗中
	難度測驗	評量學生獲得知識與技能的 <u>程度</u> ，非速度快慢	成就測驗

	個人智力測驗	團體智力測驗
人數	每次施測一人	可團體施測
功能	檢測資賦優異或是心智功能低下的兒童之工具	提供篩選和分類所需之一般性資料
形式	大多為綜合性的成套測驗或測驗組合，可同時測量各種不同的能力，並評鑑受試者各種能力發展上之優劣	多為紙筆測驗和機器評分，具有經濟、方便之優點

	性向測驗	成就測驗
測量能力	偏向學習新事物的能力，亦即重視未來表現的潛能	偏向已經習得的能力，強調目前已獲得的知識技能水準
測量目的	預測未來的表現水準	確定個人從過去到現在的學習所獲得的成就水準或熟練程度

測量內容	取材範圍較 廣泛 ，較普遍性	內容僅限於 學校教學的特定內容或學習經驗
測量時機	教學 前 測量	教學 後 測量

人格測驗	
自陳量表	
意義	受試者根據個人的感覺或實際狀況回答，根據答案衡量受試者此種特性所表現的程度
特點	*易受 測試對象 和 測試形式 的影響·通常採用一定的措施控制這種影響 *多採用 紙筆 的形式 *可以測試 個體 亦可測試 團體
優點	可 團體 施測，比其它方法省時省力
缺點	填答者需要具備閱讀能力，不適用於 幼童 及 視障者 或容易受社會期望的影響

(五)常模

意義	特定參照團體在測驗上獲得的 平均分數 ，也可視為平均數 學生考了85分，不知道是好還是不好，透過常模參照，就可以知道分數在團體中的 相對地位 。
功用	分數在團體中的相對位置 提供比較的 量數 ，比較個人在不同測驗上的分數
類型	全國性常模：根據 全國性代表樣本 建立的常模。
	地區性常模：根據 同地區學生樣本 建立的常模
	特殊團體常模：根據 特殊族群或團體 來建立的常模
	學校平均數常模：同 學校學生 建立的常模
適切性	新近性：常模須為 最新
	適切性：常模樣本是否跟 施測對象 相類似
	代表性：樣本的 代表性、抽樣方法

*適切性：在選擇測驗與解釋測驗分數時，依三個標準評定常模適當性

常模的建立方法

(一)發展性常模：根據個人所獲得的發展水準所得的常模

測驗形式	適用範圍	注意事項
年齡常模		
依不同年齡在測驗上所得的平均數 (*參照年齡常模表) *如測驗得分為 8 分，閱讀年齡則為 9 歲 2 個月	會隨著年齡增長的特質 如：身高、體重、智力、閱讀能力等	心理和教育特徵的成長，各年齡階段 <u>不一定一致</u> ，因此 <u>沒有一致單位</u>
年級常模		
依不同年齡在測驗上所得的平均數 (*參照年級常模表) *測驗得分為 22 分，年齡等值為 3 年級 7 個月	隨著年級而逐漸增加的特質 如：小學各種基本技能的學習	*年級常模單位不相等 *不同學科的年級常模有不同意義

年齡常模表			
原始分數	閱讀年齡	原始分數	閱讀年齡
0	7-8	13	10-6
1	7-10	14	10-8
2	7-11	15	11-2
3	8-1	16	11-8
4	8-5	17	12-1
5	8-7	18	12-6
6	8-9	19	12-8
7	9-0	20	12-11
8	9-2	21	13-5
9	9-6	22	13-11
10	9-9	23	14-5
11	10-2	24	15-1
12	10-4		

年級常模表(取自桑代克, 1977)			
原始分數	年級等值	原始分數	年級等值
19	3.2	31	4.7
20	3.4	32	4.8
21	3.6	33	5.0
22	3.7	34	5.2
23	3.8	35	5.3
24	3.9	36	5.4
25	4.0	37	5.6
26	4.1	38	5.7
27	4.2	39	5.9
28	4.3	40	6.0
29	4.4	41	6.1
30	4.6	42	6.2

(二)團體/組內常模：依個人於**特定團體內相對位置**建立的常模

(1)百分等級：

百分等級(PR 值)	
意義	將一個團體分為 100 等分，每一等分點稱作 百分點
	百分等級代表的是 等第 ，指 各百分點以下 包含多少人數
求法	樣本數不多時，可先依人數給 名次 ，再使用公式求百分等級 PR 值： $100 - \frac{(100R - 50)}{N}$ N：全體人數；R：名次
	例題 五位學生成績 <u>93</u> , <u>84</u> , <u>75</u> , <u>62</u> , <u>46</u> ，一二名的百分等級為？ PR _{1st} ： $100 - \frac{(100 * 1 - 50)}{5} = 90$ ；PR _{2nd} ： $100 - \frac{(100 * 2 - 50)}{5} = 70$
優點	易於了解
	適用各種測驗、各種對象(成人及兒童)
	無論原始分數為何，都可以正確解釋分數
注意事項	百分等級與百分比不同
	百分等級的 單位不相等 ，屬於 順序量尺 而非等距量尺
	分布於兩端與分布於中央的差異顯著，故順序量尺 無法計算
	百分等級分布為 1-99

(2)標準分數：

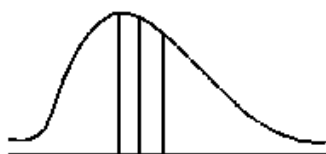
標準分數	直線標準分數	Z分數	意義	Z分數可以表明原始分數在平均數之上或之下幾個標準差	
			計算方式	公式： $Z = \frac{X - M}{S}$ X：原始分數；M：平均數；S：標準差	
			缺點	有負數與小數點，使用不便。	
		T分數	意義	強制使平均數為50，標準差為10，使T分數為正整數（修正Z分數有負數及小數的問題）	
			計算方式	公式： $10Z + 50 = 10 \frac{X - M}{S} + 50$	
			意義	最普遍使用的常態標準分數。與T分數相同是以平均數為50，標準差為10，但T量表分數必為常態分配。	
	常態化標準分數	T量表分數	計算步驟	將原始分數轉換成百分等級 依常態分配表，找出相對的Z分數 利用 $T = 10Z + 50$ 方式轉換為T量表分數	
			標準九常模	計算方式	將原始分數分為九等分 平均分數為5；標準差接近2
				用途	性向測驗與成就測驗
		特點	特點	只用個位數表示分數，資料處理方便 意義容易被了解 屬於常態標準分數，可用於不同測驗的分數	

(三)測驗分數的解釋

*正偏態與負偏態

(1)正偏態：通常出現於主科(國、英、數等)

(2)負偏態：通常出現於副科(社會、生科等)



(A)正偏態時



(A)負偏態時

(六)信度與效度

(1)信度

信度	意義	為一個測驗結果的一致性，信度越高測驗的結果越一致	
		信度高的測驗會有 良好的穩定性	
		測驗分數差異越大，異質性越高，信度通常也會比較高	
		在原有條件不變(學生人數、難度、鑑別度等)的情況下， 測驗的題目越多 ，信度會提高。	
	性質	指評量工具獲得「 結果 」的可靠性，非工具本身	
		信度是效度的 必要條件 ，但不是充分條件	
	功能	代表測驗分數具有 一致性、穩定性及可預測性	
		測驗要能夠為同一位學生提供 前後一致 的測驗結果	
		測驗是否 優良的指標	
	係數	題目越多 ，信度越高	
		難度中等 比難度高或低的成就測驗有較高的信度	
		愈能區分 高、低成就者的測驗，題目信度 越高	
		r值範圍在0.00~1.00之間，沒有負值， r=0表示無信度；r=1表示信度最高	
	指數	是個人測驗分數與真實分數的相關，而 信度指數的平方 即為信度係數	
	測量誤差	系統誤差/偏誤： 學生的學習、訓練、遺忘、生長等 ，是指一種以固定、一致的方式影響測驗分數高低的測量誤差。不同情境中，對不同學生影響都是一致的。	
		非系統誤差/隨機誤差： 學生的身心狀況 (動機、情緒、態度、意願)、 施測情境 (噪音、溫度、照明)、 測驗試題 (編擬、抽樣、計分、解釋)等，是以隨機、無規則、不可預測的方式，在不同情境中，隨時影響不同學生的測驗分數	
影響因素	測驗的長度	測驗越 長 ，內容愈具有 代表性 ，分數受到猜測因素的影響越小，所以信度越高	
	分數的分布情形	分數的 分布範圍越大 ，第二次測量時，分數相對位置改變的可能性就愈小，亦即測量誤差對相對位置的影響愈小	
	測驗的難度	測驗難度適中 ，能使分數變異程度加大，信度變高	
	客觀性	具 客觀性的測驗 其評分結果較不會受到評分者的判斷與意見的影響，所以相對而言其穩定性高，信度高	
補救方法	增加試題或刪除不	增加試題 可以提高測驗的信度，但需考慮學生是否會疲勞	

	良試題		
	校正相關係數的萎縮	對於低信度的測驗最好不要使用，若不慎使用了，則務必進行 相關係數萎縮的校正 ，才能獲取接近真實測量下的真正相關。如斯布校正公式使信度提高	
教學評量的應用	幫助教師評鑑已發行評量工具的信度		
	幫助教師增進自行編制之評量工具的信度		
類型	重測信度	同一份測驗於不同時間對相同學生前後重複測量兩次	
		再測信度的高低和兩次間隔時間的長短有關	
		間隔大多以一至兩週最為恰當	
		前測與後測之間，測驗分數常有提高的現象，是因「練習效應」和「累積性的成長」	
	複本信度	意義	相同編制藍本使兩份題數、難度、指導說明、施測時限與例題舉隅等都相當的測驗拿給同一批學生施測
		優點	不受練習和記憶的影響，為最好的信度
		缺點	編製不易，費時且成本高 易受練習影響的測量行為，即使採用複本，影響只能減少，不能避免 重測與複本信度都必須進行兩次施測或使用兩份測驗，使用不方便
	評分者信度	意義	採用不同評分者評閱測驗卷
		採用時機	對一些無法進行完全客觀記分的測量工具，例如：人格測驗、創造性思考測驗及申論題、作文題等必然涉及評分者主觀的判斷，故須對評分者間的變異加以衡量。
	內部一致性信度	折半信度	意義
使用			*一般教師自編測驗大多用折半法 *不必施測兩次，因此學生也不會因兩次考試而感到厭煩，影響信度。
庫李信度 KR20/21		意義	依學生對試題的反應，分析試題間的一致性，以確定測驗中的試題是否都測量到相同特質

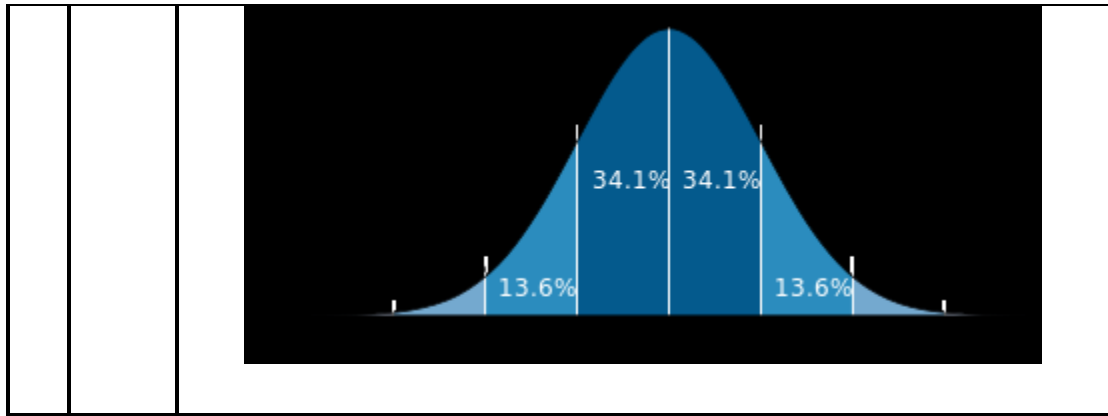
				基本假設	計分使用「對或錯」的二元計分方式
					不受作答速度的影響
					試題都是同質的，亦即都測量到一個相同的因素

*信度誤差來源

再測信度	時間抽樣	
複本信度	時間抽樣、內容抽樣	
評分者間/內信度	評分者誤差	
內部一致性信度	折半信度	內容抽樣
	庫李信度	內容抽樣、內容異質
	α 係數	內容抽樣、內容異質

*測量標準誤

測量標準誤	意義	為表示信度的方法，指個人測驗上實得分數和真正分數的差，稱為測量誤差		
	使用	*適合於解釋個人的分數 (信度係數較適合於比較不同測驗的信度) *測量標準誤與信度呈現反比關係。信度愈高，測量標準誤愈低		
	公式	$SE = Sx\sqrt{1-r}$ SE=測量標準誤；Sx=測驗標準差；r=測驗信度係數		
	標準差	統計分布的程度		
	變異數	意義	所有資料到平均數的距離。是「標準差」的平方	
		計算方式	離差分數	意義 每個分數與平均數的差
				公式 $X - \bar{X}$
			離差分數平方	$ X - \bar{X} ^2$
			變異數	意義 將離差分數平方除以總數
			公式 $SD^2 = \frac{\sum X - \bar{X} ^2}{N}$	
信賴區間	68%的信賴區間：真實分數介於實得分數，正負 1 個標準誤 95%的信賴區間：真實分數介於實得分數，正負 1.96 個標準誤 99%的信賴區間：真實分數介於實得分數，正負 2.58 個標準誤			

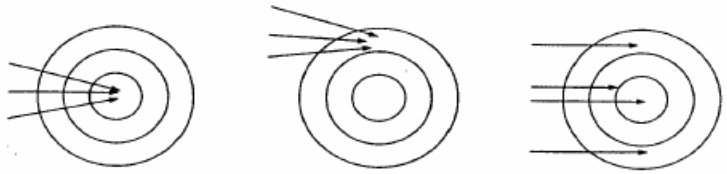


(2)效度

效度	意義	概念定義		測驗能確實測量到它所欲測量的 特質或功能 的程度，效度越高表示該測驗 越能達到它所要測量的目標
				一個心理測驗在評量上最重要的特徵
				若其他條件均相等，樣本的 異質性越大 ，則 效度會越大
	性質	指「 測驗結果 」 正確性 的程度，而非指工具本身		
		不是「全有」或「全無」，而是「 程度的差別 」。		
		是針對 某種特殊的用途 而言，而不可認為具有普遍性		
	影響因素	測驗本身		
		教學		
		測驗實施與計分		
		受試者		
群體和效標的性質				
測驗與效標兩者實施時間的間隔				
信度的影響： 效度係數的最大值與信度高低有直接的關聯				
預測測驗之題數				
補救方法	放棄不用 *重新編擬試題	以內容效度而言	藉助 雙向細目表 ，若發現試題不適合，即將其刪除	
			若 太多不良題目 則應放棄整份測驗	
	以效標關聯效度而言	外在效標 選取不當		
		理論構念 不周全		
		試題不佳 ，進行修改或重編		
校正相關係數的萎縮： 效度係數 有時會被 不可靠的測驗分數 給萎縮，而造成 效度係數降低 的情形				
類型	內容效度	意義	測驗試題樣本內容具有 代表性 ，能 反映出教學目標或教材內容	

		目的	確定受試者目前對於測驗中所呈現的代表性樣本之表現程度			
		考驗方法	利用雙向細目表，判斷每個試題是否與教材內容所涵蓋的範圍及教學目標相符			
		考量因素	適當性(適合的年齡水準&內容)及完整性			
		表面效度	*指測驗給人的印象「好像」是在測量某種特質的指標，而不是指測驗事實上能測量到什麼特質 *一份適當的測驗不但要具備表面效度，更重要的是要有內容效度			
		專家效度	由專家以雙向細目表之方式來得到效度			
	效標關聯效度	意義	研究測驗分數與外在效標間關聯性的一種指標。預測學生未來的表現或估計學生目前在某些效標表現上的未知狀況			
		特徵	適切性	可靠性	客觀性	可用性
		效標種類	學業成就	學校成績、升留級記錄、特殊榮譽、教師的評分等		
			特殊化的訓練成績	音樂系和美術系的術科成績、商學院的打字、簿記、會計等		
			實際工作表現			
			對照團體、相關人員的評定			
			精神病學的診斷			
		預測效度	測驗能預測考試者未來表現的程度			
	同時效度	指測驗能評估考試者目前表現的程度				
	概念(建構)效度	意義	以心理學的理论概念來說明並分析測驗分數的意義			
		成分	信度	量表本身可信的程度		
			聚斂/符合效度	使用不同方法測量相同特質		
			區辨效度	使用相同方法測量不同特質		
		例子	機械性向的測量分數和在校機械科目的成績算相關，如果是高相關，表示有符合效度。而和語文測驗的成績是低相關，則有區辨效度			
	內部	意義	內部品質是否測量相同的心理特質			

		一致性分析法	相關分析	分析個別試題與總分間的相關。凡相關係數經考驗達到顯著水準者，即可保留下來	
			團體對照	依據學生測驗總分分為高低分組，比較兩組學生在每個試題的答對百分比，該題的高分組其答對百分比顯著的高於低分組的話，即可保留下來	
		外在效標分析法	意義	以一個有效度的測驗為效標，計算此測驗與新測驗的相關係數	
			相關分析	建構出一個測驗，並將此測驗施測後與一個適當的外在效標做相關，統計後若達顯著，即表示效標關聯的構念效度獲得驗證	
			團體對照	實驗組與控制組(相對族群)的學生在某個測驗的表現上有明顯的差異，此皆可作為支持該測驗其建構效度的證據	
		因素分析法	從一堆試題中抽出少數幾個共同因素，用以反應或代表這堆試題的共同結構。分析大量變項間的關係，常用在人格量表上		
		多因素/特質分析法： 以聚斂效度和區辨效度來作為構念效度的證據			
		內在效度	*指實驗研究者所操縱的實驗變項對依變項所造成的影響真正程度。 *亦即一個實驗研究能夠有效的實驗出所要實驗的因果關係，若實驗的干擾愈多，正確性愈差，則該實驗的內在效度便愈低，反之干擾變項能完全控制，其正確性愈高，則該實驗的內在效度便愈高		
		外部效度	*指實驗結果的可推論性之程序的大小 *實驗結果的可推論性愈大，亦即其適用性，代表性愈大，實驗的外在效度則愈高。易言之，就是指實驗研究結果是否可推論到研究對象以外的其他受試者，或研究情境以外的其他情境		

信度與效度的關係	意義	信度	測驗結果的一致性和穩定性		
		效度	測驗能夠測量其所欲測量的特質之程度		
	考驗方法	信度	邏輯的分析與統計的應用		
		效度	完全採用統計方法		
	使用	信度是效度的必要條件而非充分條件 測驗要有效度必須要有信度 但有信度並不一定有效度			
		 <p>有信度和效度 有信度但無效度 無信度又無效度</p> <p>射箭圖的信度與效度關係</p>			
		高	信度	效度	不一定高
低	一定低				
一定高	高				
效度係數一定不會大於信度係數的平方根					

(七) 試題分析

試題分析	高/低分組	高分組	最高分的 27% 學生	
		低分組	最低分的 27% 學生	
	難度分析	意義	計算全體受試者答對每題的人數百分比	
		公式	* $P = (R/N) * 100\%$ (數值越大, 難度越低)	
			* $P = P_H + P_L / 2$	
	鑑別度分析	意義	了解各個試題的功能是否和整個測驗的功能一致	
			高分組答對的比例應顯著大於低分組的比例	
		算式	$D = P_H - P_L$	
		公式分析	鑑別度值介於 -1.00 至 +1.00 之間	
			指數愈高, 鑑別力愈大	
若指數若接近 0, 代表該題沒有鑑別度 (題目太難或太簡單, 大家都對或都錯) (題目不夠清楚, 以致猜測)				

難度與鑑別度的關係	意義	難度適中，鑑別力最大
		試題平均鑑別度愈高，測驗信度也會愈高
	試題選擇標準	鑑別度在.25以上，愈高愈好
		難度最好接近.50，以.40-.70為範圍
		考量測驗的「雙向細目表」以符合內容效度
	選項有效分析	依高、低分組在各個選項上選答的次數進行判斷
		每項不正確的選項，至少應有一個低分組的受試者選它
		低分組選擇不正確選項的人數應多於高分組
		修改選項之前，要仔細考量是題目本身的缺陷，或是學生反應上的錯誤
	試題分析解釋	鑑別指數不等於題目效度
		鑑別指數低未必表示試題有缺點
		試題分析資料是假設性的

*例題分析

試題分析(一)							
組別	選項					D	P
	A	B	C	D	未答		
高分組	14	0	0	6	0	.10	.65
低分組	12	0	8	0	0		
分析	<p>*正確答案為(A)</p> <p>*<u>(C)是有效的選項</u>，因為它吸引更多低分組受試選答</p> <p>*<u>(B)則是沒有誘答力的選項</u>，因為兩組均無人選他，故須要修改</p> <p>*<u>(D)也是不佳選項</u>，因為它吸引更多高分組受試選答，是否意義不清楚?或有兩個正確答案?或受試者作答不小心?</p> <p>*如果(B)與(D)選項能加以修改，則本題鑑別力可能會因之提高</p>						

試題分析(二)							
組別	選項					D	P
	A	B	C	D	未答		
高分組	9	0	10	1	0	.20	.40
低分組	5	4	6	5	0		
分析	<p>*正確答案為(C)</p> <p>*<u>高分組選答(A)與(C)的人數接近</u>，顯見這兩個選目在正確性上不夠清楚，應修改使之只有一個答案或最佳答案</p> <p>*<u>低分組對每個選答都非常接近</u>，這表示每個選目對他們均有誘答力，可能是由於他們缺乏知識所致</p>						

(八)教學評量

(1)意義：欲瞭解學生的行為改變是否如預期的教學結果，就必須在教學後，做客觀而正確的評量

(2)種類：

依評量的目的				
	預備性評量	診斷性評量	形成性評量	總結性評量
功能	*測量學習前必要的基本技巧 *確認已達成課程目標的程度	確認學習困難的原因	提供學習進步的回饋給教師和學生	在教學結束時給予成績等等，或確認精熟程度
時間	教學之初	學習發生困難	教學進行中	教學末尾
工具	*學前測驗 *教師自編測驗	*診斷測驗 *標準化成就測驗或檢核表	教學須要而特別設計的： *評定量表、 *作業或口頭考問 *實際演示 *問題研討	學期末或教學單元結束時的考試，大多為教師自編的測驗
取樣考慮	*包含每項必要的起點行為 *選擇能代表課程目標的樣本	包含共同學習錯誤的樣本	所有單元目標(或最主要的)	選擇能代表課程目標的樣本
難度	大部分試題是基本的、容易的、難度P值在0.65以上	大部分試題是基本的、容易的、難度P值在0.65以上	試題配合單元目標的難度	大部分的試題難度P值在0.3~0.7間，較難與較易的試題也有
評分方式	標準參照	標準參照或常模參照	標準參照、精熟或非精熟	通常是常模參照，也可能是標準參照

依如何解釋評量結果而分		
	常模參照評量	標準參照評量
目的	區分學生成就的 個別差異 (分班編組)	確定學生學習的 精熟度 (補救教學)
內容	涵蓋學習課題的廣大領域，但每一獨特的課題僅以少數題目測量	涵蓋學習課題的有限領域，但每一獨特課題有較多題目測量
難度	中等難度，排除極難和極易題目	題目難度和學習內容相配合，不試圖排除極易題目或變更其難度
結果解釋	依分數在 團體中相對地位 解釋	依據分數達到 預定標準 解釋
分數種類	百分等級、標準分數、年級常模	內容參照分數、目標參照分數、精熟分數
適用評量	安置性評量、總結性評量	預備性評量、形成性評量、診斷性評量

變通/替代性評量			
	真實評量	實作評量	卷宗評量 (學習檔案)
特性	強調 應用 、重視 直接評量 、使用 真實的問題 、鼓勵 開放思考	能同時評量 認知與技能 方面的教學目標 評量 情境接近現實生活 ，有利學生的學習遷移	有系統的蒐集學生作品 ，透過長時間 連續性的蒐集資料 ，來觀察學生在一個或多個學習領域內的 努力、進步與成就
目的	強調在使兒童於班級教學的 真實活動中 ， 直接去評量 學習者在教室活動中直行或操作一項工作，以建構有意義的學習經驗	*彌補紙筆測驗的不足 *對學生的學習成就做更正確的推論 *對教師的教學活動產生正向的引導	培養學生 自我評鑑 、 自我改進 的能力，養成學生學習責任感
形式	<u>紙筆工作</u> 、表演、作品集 (歷程集&成果集)	紙筆實作、(實物)辨認測驗、模擬實作	展示卷宗(最佳作品)、描述性卷宗、評估性卷宗、進步卷宗