

國立中山大學卓越教學小組

97 學年度計畫成果報告

「教育測驗與評量研究」

讀書會



執行單位：教育研究所

計畫主持人：洪瑞兒 副教授

計畫聯絡人：陳香廷




執行期間：98年2月16日至98年6月30日






目 錄

一、成果報告表.....	1
二、計畫申請書.....	3
三、各週討論內容紀錄.....	7
四、討論照片舉隅.....	87
五、教育專題講座資料.....	89
六、教育專題講座照片.....	96

國立中山大學卓越教學小組 97 學年度計畫成果報告表

執行單位	教育研究所		計畫類別	提升學生學習意願- 讀書會	
計畫主持人	洪瑞兒副教授				
計畫聯絡人	陳香廷 (分機: 5885, mail: cht0913@hotmail.com)				
計畫名稱	97 學年度卓越教學小組學生學習意願提升子計畫之 97 學年度「教育測驗與評量研究」讀書會				
執行情形	<input type="checkbox"/> 計畫尚未完成，繳交期中成果報告 <input checked="" type="checkbox"/> 計畫已完成，繳交期末成果報告				
執行期間	98 年 2 月 16 日 至 98 年 6 月 30 日				
計畫成果摘要	量化成果	辦理活動 (課程) 名稱	辦理場次	參加人次	
		導讀暨討論會	16	11 人	
		專題講座- 科技與人文的橋樑生物科技教育	1	50	
質化成果 (1000 字以內)	<p style="text-align: justify;">進行一學期的教育測驗與評量研究讀書會，透過成員間對於各週主題相關期刊論文的內容討論，分享研讀心得以及問題交流，在相互督促中，提升研究生自主學習、批判思考、練習期刊論文寫作的知能，對教育測驗與評量有更深入的了解。除能對課程有進一步的認知與了解外，亦能藉此與讀書會的夥伴建立深厚 Study Partner 關係，相互切磋。本活動增進讀書會成員間的互動與交流，落實卓越教學制度，增進課業內容之瞭解。</p>				
成果照片與說明 (請放置辦理活動及購置設備之					
	讀書會討論情形		讀書會討論情形		讀書會討論情形

照片，至少四張)																							
	讀書會討論情形	教育專題講座情形	教育專題講座情形																				
成果自評與建議 (200 字以內)	<p>藉由本學期的讀書交流，成員不僅對於教育測驗與評量能有更深入的了解，且透過重點摘要統整，能供讀書會成員日後查詢及閱讀，唯以下幾點不足之處尚須改進：</p> <ol style="list-style-type: none"> 1.文章導讀以中文居多，建議往後多以外文為主，提昇研究深度。 2.本校圖書館的教育相關書籍較少，且書局訂購圖書又須等待一段時日，因此在安排讀書進度時，關於閱讀書籍與參考資料的購買可提早進行。 3.各週讀書會參與人數不固定，影響討論品質，建議事先訂定出席相關規則，以使讀書會進展更加順利。 																						
經費執行																							
<table border="1"> <thead> <tr> <th>申請經費項目</th> <th>核定補助金額 (元)</th> <th>實際執行經費 (元)</th> <th>經費執行率 (%)</th> </tr> </thead> <tbody> <tr> <td>1.資本門</td> <td>5,000</td> <td>5,000</td> <td>100</td> </tr> <tr> <td>2.經常門 (業務費及人事費)</td> <td>22,765</td> <td>13,365</td> <td>58.71</td> </tr> <tr> <td>3.國際差旅費</td> <td>-</td> <td>-</td> <td>-</td> </tr> <tr> <td>合計</td> <td>27,765</td> <td>18,365</td> <td>66.14%</td> </tr> </tbody> </table>				申請經費項目	核定補助金額 (元)	實際執行經費 (元)	經費執行率 (%)	1.資本門	5,000	5,000	100	2.經常門 (業務費及人事費)	22,765	13,365	58.71	3.國際差旅費	-	-	-	合計	27,765	18,365	66.14%
申請經費項目	核定補助金額 (元)	實際執行經費 (元)	經費執行率 (%)																				
1.資本門	5,000	5,000	100																				
2.經常門 (業務費及人事費)	22,765	13,365	58.71																				
3.國際差旅費	-	-	-																				
合計	27,765	18,365	66.14%																				
設備購置及設施改善執行情形：																							
<table border="1"> <thead> <tr> <th>設備(施)名稱</th> <th>購置(工程)期間</th> <th>放置(工程)地點</th> <th>購置進度</th> <th>備註</th> </tr> </thead> <tbody> <tr> <td>圖書一批</td> <td>98.4~98.7</td> <td>圖書館</td> <td>已完成</td> <td></td> </tr> </tbody> </table>				設備(施)名稱	購置(工程)期間	放置(工程)地點	購置進度	備註	圖書一批	98.4~98.7	圖書館	已完成											
設備(施)名稱	購置(工程)期間	放置(工程)地點	購置進度	備註																			
圖書一批	98.4~98.7	圖書館	已完成																				
計畫成果補充電子檔	電子檔如附件所示																						
繳交(上傳)時間	98.9.3																						
備註																							

97 學年度教育測驗與評量研究讀書會

一、計畫名稱：97 學年度教育測驗與評量研究讀書會

二、計畫目的：

本活動主要目的為透過成員間的討論、交流，以期提升研究生自主學習、批判思考、練習期刊論文寫作的知能，並藉此交流研讀心得及問題。除能對課程有進一步的認知與了解外，亦能藉此與讀書會的夥伴建立深厚 Study Partner 關係，相互切磋。本活動預期達到以下目的：

- (一) 增進讀書會成員之間的互動與交流。
- (二) 落實卓越教學制度。
- (三) 增進對課業內容之瞭解。

三、計畫內容：

讀書會計畫主要分為兩部份：

(一) 座談會

外聘教育相關領域學者進行 3 次各 3 小時的座談會。

(二) 導讀暨討論會

每次討論會由一位同學領讀教師指定閱讀內容，經成員討論與提問後，由該週導讀紀錄者負責統整；每次活動進行三小時，每週一次。因此導讀暨討論會部分，可再細分為討論日期、導讀書目、導讀者、及該週導讀之紀錄者。詳細實施方式如下表：

導讀暨討論會安排				
週數	討論日期	導讀內容	導讀人	紀錄人
1	2009/02/20	統計基本概念	黃怡菁	陳怡安
2	2009/02/27	統計分析及 APA 格式介紹	陳香廷	黃怡菁
3	2009/03/06	常模、信度與相關期刊論文	王薪惠	陳香廷
4	2009/03/13	效度與相關期刊論文	陳怡安	王薪惠
5	2009/03/20	試題分析與相關期刊論文	曾韻琴	陳怡安
6	2009/03/27	行為檢核表編製與使用與相關期刊論文	林曉瑩	曾韻琴
7	2009/04/03	訪談與焦點團體訪談相關期刊論文	張珮蓉	林曉瑩
8	2009/04/10	態度測驗編製與使用相關期刊論文	許筑鈞	張珮蓉
9	2009/04/17	智力測驗編製與使用及其相關期刊論文	鄭媯穗	許筑鈞
10	2009/04/24	性向測驗編製與使用及其相關期刊論文	周秀玉	鄭媯穗

11	2009/05/01	興趣測驗編製與使用及其相關期刊論文	林慧雅	周秀玉
12	2009/05/08	成就測驗編製與使用及其相關期刊論文	黃怡菁	林慧雅
13	2009/05/15	人格測驗編製與使用及其相關期刊論文	陳香廷	黃怡菁
14	2009/05/22	IRT 介紹及其相關期刊論文	王薪惠	陳香廷
15	2009/05/29	質性研究相關期刊論文	陳怡安	王薪惠
16	2009/06/05	總檢討	--	陳怡安

備註：參考書 / 教科書編號

參考書 / 教科書編號	
編號	參考書 / 教科書
①	Arthur Aron & Elaine N. Aron 著，黃瓊蓉譯(2000)。心理與教育統計學。台北：學富。
②	林重新編著(2001)。心理與教育測驗。台北：揚智。
③	Robert L. Linn & M. David Miller 著，王振世等譯(2009)。教育測驗與評量。台北：雙葉。
④	Kenneth D. Hopkins 著，李茂興譯(2002)。教育測驗與評量。台北：學富。
⑤	陳英豪、吳裕益(2001)。測驗與評量。高雄：麗文。
⑥	謝臥龍（主編）(2004)。質性研究。台北：心理。
⑦	Matthew B. Miles & A. Michael Huberman. (1994). Qualitative Data Analysis: An Expanded Sourcebook. Thousand Oaks, CA: Sage.
⑧	David C. Howell. (1997). Statistical Methods for Psychology. Belmont, CA: Wadsworth.

四、導讀暨討論會甘特圖

工作項目	週次															
	第1週	第2週	第3週	第4週	第5週	第6週	第7週	第8週	第9週	第10週	第11週	第12週	第13週	第14週	第15週	第16週
《心理與教育統計學》第1~																
《心理與教育統計學》第6~																
《心理與教育測驗》第2章 《教育測驗與評量》第3、5 英文期刊一篇																
《心理與教育測驗》第4章 英文期刊一篇																

六、計畫參與人員：

編號	學號	系所	姓名
1	M946050020	教研所	黃怡菁
2	M966050001	教研所	陳香廷
3	M966050007	教研所	王薪惠
4	M966050006	教研所	陳怡安
5	M966050011	教研所	曾韻琴
6	N966050101	教研所	林曉瑩
7	N966050102	教研所	張珮蓉
8	M966050005	教研所	許筑鈞
9	M976050016	教研所	鄭媖穗
10	N976050109	教研所	周秀玉
11	N976050113	教研所	林慧雅

七、活動地點&時間

自 98/2/16~98/6/31，導讀暨討論會每週舉辦一次，時間為每週五下午 2~5 時，討論內容為該次導讀內容。地點在社科院二樓教育研究所會議室，共計舉辦 16 次。

本學期擬安排三次座談會，座談時間為每次三小時，地點在社科院二樓 2001 教室。

972 教育測驗與評量研究讀書會紀錄（一）

◎主 題：統計基本概念

◎時 間： 2009/02/20

◎地 點：社 4042-1

◎導讀人員：黃怡菁

◎紀錄人員：陳香廷

◎內容紀錄：

變項

變項：指人或事物的特徵，這些特徵在質與量方面往往變動不定，所以有不同的特質。而變項的種類很多，分類的方式不一，常依需要而定。

依據用途來分：

- 1.自變項：實驗者操縱的變項。
- 2.依變項：被預測的變項，隨自變項的改變而變化。
即，自變項是前項，依變項是後項。

在研究設計與執行過程中，區分為：

- 1.主動變項：研究者操縱的變項。
- 2.屬性變項：研究者無法操縱，只能加以測量，有關人的所有特徵皆是，例如：
智力、性向。

在研究設計與分析資料方面分成：

- 1.等距變項/連續變項(Interval Variable)：某一範圍內呈現順序數值者，如：智力、score.
- 2.類別變項/名義變項(Categorical Variable)：測量中的類別量尺，依有無具備所界定的特徵作為分類的依據。屬於一種名義測量，在這種測量中具有同一性質的東西就歸為一類。如：性別、姓名。
- 3.次序變項 (Ordinal Variable)：如第一名、第二名。
- 4.比率變項 (Ratio Variable)：

其他

- 1.機體變項：接受研究者本身就存在的特徵，是實驗者無法改變的特徵，如：
性別、智力。也稱為屬性變項。
- 2.中介變項：無法操縱或測量的，但可從推論中得知其存在的變項。

- 3.控制變項：由研究者決定會影響結果的自變項以外的變項。
- 4.居中變項：與上述變項可能重疊的變項之一為居中變項。

以上的變項並非互相排斥，任何研究並非限定單一的自變項、機體變項、控制變項、中介變項或依變項。

研究本身是一種觀念的溝通歷程，也就是研究者本身嘗試將自己的研究結果報告給他人知道。由於研究所涉及的概念或變項對不同的人具有不同的意義，因此，研究中所使用的概念或主要變項必須加以界定清楚。界定變項的意義有兩種：

- 1.概念性定義：以一個概念來界定另一個概念，而不是根據可觀察或可操作的特徵來界定研究變項的意義。例如：將智力界定為〈抽象思考的能力〉或〈學習的能力〉就是一種概念性定義
- 2.操作性定義：根據可觀察、可測量或可操作的特徵來界定研究變項的定義。就是將研究變項的抽象形式轉變為可以觀察、可以測量和可以操作的形式。例如：將智力界定為〈比西量表所測得的分數〉就是一種操作性定義。

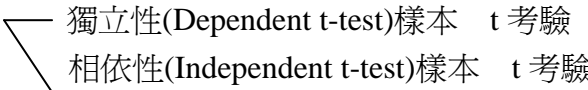
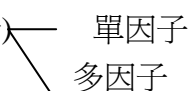
統計基本概念

1. 測量量尺 Scale/Variable 分為四種：

- (1) 名義量尺(名義變項、類別變項)(Nominal/Categorical Scale) ex: 性別 (男, 女)..etc..
- (2) 次序量尺(次序變項)(Ordinal Scale) ex: 第一名、第二名..etc..
- (3) 等距量尺(等距變項)(Interval Scale) ex: 成績 50 分、60 分……100 分, ..etc.
- (4) 比率量尺(比率變項)(Ratio Scale) ex: 重量、年齡(age)、工作經驗長短、身高、溫度…etc..

2. 常使用的教育統計方法：

- (1) Descriptive statistics: Frequency, 平均值 \bar{X} , 標準差(SD), 變異數(Variance, SD^2/S^2), 標準分數, figures。
- (2) 試題分析(Item Analysis):
 - a. 難度分析 (Difficulty Index)
 - b. 鑑別度分析 (Discrimination Index)
 - c. 誘導度分析 (Distracting Index)
- (3) 信度分析 (Reliability Analysis):
 - a. 內部一致性分析 (Internal Consistent Cronbach Alpha)
 - b. 折半信度(Split-half Reliability)

- (4) 效度分析(Validity Analysis):
- 內容效度 (Content Validity): 雙向細目表
 - 建構效度 (Construct validity): Compare score to know groups.
 - 效標關聯效度(concurrent Validity): 與權威問卷做相關 如果 r 值很高 (>0.7), 就代表這份問卷很好.
 - Internal Validity: Based on your study data result.
 - External Validity: Your study
- (5) 相關 Pearson Correlation Coefficient / Pearson Product-moment Correlation Coefficient (皮爾遜積差 or Spearman 相關係數)
- (6) t 考驗 
- (7) 卡方考驗 : X Square(觀察值 vs.期望值)
- (8) Analysis of Variance(ANOVA 變異數分析) 
- 經常使用之 “事後考驗 Follow-up Test / Post Hoc Comparison”
- Bonferroni : 最適中
 - S-N-K: OK to use
 - Scheffs : More serious
 - Turkey: Too lose
- (9) 因素分析(Factor Analysis): 編試題(問卷)時, 相同性質題目編為一組 ,
- (10) 共變數分析 (Analysis of Covariance, ANCOVA):實驗研究/準實驗研究法時 必須使用 ANCOVA 調整分數.
- (11) 預測 (Predict) : Regression (迴歸分析)
- 逐步多元迴歸(Stepwise Regression)
 - 前進式迴歸分析(Forward Regression)
 - 後退式迴歸分析 Backward Regression
 - 移除式迴歸分析 Remove
 - 簡單迴歸分析 Enter
- (12) 卡方考驗 (Cross Tables) : 名義變項 or 類別變項適合用之 Study :
- 1.想要了解男性與女朋友互動的方式, 其互動方式共有三種: 自立的、聯結的、相互的

n=101

E : 期望值(Expectation)

O : 觀察值(Observation)

	觀察次數(O)	期望次數(E)	差異(O-E)
聯結的	50	33.67	
自立的	26	33.67	
相互的	25	33.67	

972 教育測驗與評量研究讀書會紀錄（二）

◎主 題：統計分析及 APA 格式介紹

◎時 間： 2009/02/27 ◎地 點：社 4042-1

◎導讀人員：陳香廷 ◎紀錄人員：王薪惠

◎內容紀錄：

APA 格式

一、文章結構

依據 APA 的格式，文章的結構包括封面、摘要、本文、參考文獻、附錄、以及作者註記六部份。

- (一) 封面部份依次包括報告題目、作者姓名及單位三部份，首先呈現的是報告主題，題目要能確切反映研究的變項或主要問題，避免不必要的贅詞；其次是作者的姓名，在作者姓名之後不加任何名銜及學位名稱；最後是作者的服務單位。
- (二) 摘要部份依文章性質，分別規定不同的摘要內容，實證性文章摘要的內容包括：研究問題、研究對象、研究方法、研究結果(含顯著水準)、及結論與建議；評論性文章或理論性文章摘要的內容包括：分析的主題、目的或架構、資料的來源、及最後的結論。英文摘要的字數，實證性文章以 100 至 120 字為限，評論性文章或理論性文章以 75 至 100 字為限，均不分段，但一率為兩倍行距。摘要的撰寫宜力求忠實反映本文內容，用詞精簡明確，且不添加作者本身的意見，敘述而非條列式。
- (三) 本文部份包括緒論、研究方法、研究結果、結論與建議。緒論部分包括：研究問題與背景、研究變項的定義、研究目的與假設。研究方法包括：研究對象、研究工具、實施程序。研究結果在忠實呈現資料分析的結果，結論與建議應先指出研究結果是否支持研究假設，其次依據研究結果的一致性與差異性，以及相關文獻，引出結論，再依據結論作建議，其中也可略述研究的限制。
- (四) 參考文獻部份所列的每一筆文獻，必須均在本文中引用過。在本文中呈現容易使讀者分心離題，但確又有助於讀者瞭解本研究的資料。
- (五) 附錄部份，例如自行設計的電腦軟體、未出版的測驗、複雜的數學證明、實驗用材料或器具。

(六) 作者註記部份，包括介紹每一作者的服務單位，註明協助完成本研究的單位或人員，以及與作者聯絡的方式。

二、資料引用

(一) 作者為一個人時，格式為：(中文文獻部份為筆者的意見)

1. 英文文獻：姓氏 (出版或發表年代) 或 (姓名，出版或發表年代)。
2. 中文文獻：姓名 (出版或發表年代) 或 (姓名，出版或發表年代)。

(二) 作者為兩人以上時，必須依據以下原則撰寫：

1. 原則一：作者為兩人時，兩人的姓氏 (名) 全列。
2. 原則二：作者為三至五人時，第一次所有作者均列出，第二次以後儘寫出第一位作者並加 et al.(等人)，同一段內則年代再省略。
3. 原則三：作者為六人以上時，僅列第一位作者並加 et al.(等人)，但在參考文獻中要列出所有者姓名。
4. 原則四：二位以上作者時，文中引用時作者之間用 and (與) 連接，在括弧內以及參考文獻中用 &(、) 連接。

(三) 作者為組織、團體、或單位時，依下列原則撰寫

1. 易生混淆之單位、每次均用全名。
2. 簡單且廣為人知的單位，第二次以後可用縮寫 (如 APEC、行政院教改會)，但在參考文獻中一律要寫出全名。

(四) 未標明作者 (如法令、報紙社論) 或作者為無名氏時，依據下列原則撰寫：

1. 未標明作者的文章，把引用文章的篇名或章名當作作者並加雙引號，如：(Educational Leadership, 1994) 或 (領導效能, 民 84)
2. 未標明作者的書、期刊、手冊或報告，把書名、期刊名、手冊名稱或報告名稱當作作者並劃線，如 (Total Quality Maanagement, 1995) 或 (全面品質管理, 民 84) 或 (師資培育法, 民 83)。
3. 作者署名為無名氏時，以「無名氏」當作作者。

(五) 作者姓氏相同時，相同姓氏之作者於中文引用時均引用全名，以避免混淆，如：R.D.Luce (1995) and G.E.Luce (1988)，中文部份則可加註文章名稱或書名以之區別。

(六) 括弧內同時包括多筆文獻時，依姓氏字母 (筆劃)、年代、印行中等優先順序排列，不同作者之間用：號分開，相同作者不同年代之文獻用，分開。如：(Razik & Swanson, 1993a, 1993b, in press-a, in press-b, Pautler, 1992) 或研究顯示 (吳清山、林天祐, 民 83, 民 84a, 民 84b; 劉春榮, 民 84, 印行中-a, 印行中-b)。

(七) 引用資料無年代記載或古典文獻時，(1) 知道作者姓氏，不知年代，以 n.d. 代替年代，(2) 知道作者姓氏，不知原始年代，但知道翻譯版年代時，引用譯版年代並於其前加 trans.，(3) 知道作者姓氏，不知原始年代，但知現用版

- 本年代時，引用現用版本年代並於其後註明版本別，(4)古典文獻不必列入參考文獻中，文中僅說明引用章節。如： Aristotle(n.d.)argued，或 (Aristotle.trars.1945) 或 (Aristotle.1842/1945)或論語子路篇。
- (八)引用特定文獻時，如資料來自特定章、節、圖、表、公式，要一一標明特定出處，如引用整段原文獻料，要加註頁碼。如：(Shujaa.1992.Chao.8) 或 (Lomotey.1990. p.125)或(Lomotey.1990.)(p.125) 或 (陳明終，民 83，第八章) 或 (陳明終，民 83，頁 8)。
- (九)引用個人紀錄時，不必列入參考文獻中。引用時要註明：作者、個人紀錄類別、以及詳細日期。如：(T.A.Razik. Diary. May 1. 1993) 或 (林天祐，日記，83 年 5 月 1 日)。
- (十)其他方面，如：(see Table 2 of Razik & Swanson. 1993. for complete data) 或 (詳細資料請參閱：林天祐，民 84，表 1)。

三、參考文獻

(一) 期刊、雜誌、新聞文獻：

- 1.中文期刊格式 A：作者（年代）。文章名稱。期刊名稱，期別，頁別。
- 2.中文期刊格式 B：作者（印製中）。文章名稱。期刊名稱，期別，頁別。
- 3.英文期刊格式 A：Author，A.A,Author.B.B & Author，C.C1995
Title of article.Title of Periodical .xx (xx) . xxx-xxxx.
- 4.英文期刊格式 B：Author，A.A,Author.B.B & Author，C.C (in press) .Title of article.Title of Periodical .xx (xx) . xxx-xxxx.
- 5.中文雜誌格式：作者（年月日）。文章名稱。雜誌名稱，期別，頁別
- 6.英文雜誌格式：Author，A.A,Author.B.B (1996.January 9) Article title.Magazine Title. xxx,xx-xx.
- 7.中文報紙格式 A：作者（年月日）。文章名稱，版別。
- 8.中文報紙格式 B：文章名稱（年月日）。報紙名稱，版別。
- 9.英文報紙格式 A：Author，A.A, (1996.January 9) .Article title.Newspape_title，p.xx
- 10.英文報紙格式 B：Article title (1996 .January 9) .Newspape_title，p.xx

(二) 書籍、手冊、書的一章:

- 1.中文書籍格式 A:作者(年代)。書名·出版地點:出版商。
- 2.中文書籍格式 B:作者(年代)·書名(版別)·出版地點:出版商·
- 3.中文書籍格式 C:單位(年代)，書名(編號)·出版地點:作者。
- 4.中文書籍格式 D:書名(年代)。出版地點:出版商·
- 5.英文書籍格式 A:Author，A·A(1993)·Book title.Location:publisher·
- 6.英文書籍格式 8:Author，A.A (1993)·Book title. (2nd ed)
Location:Publisher，

- 7.英文書籍格式 C:Institute · (1993) , Book title(No.123)Location:Publisher
Author
- 8.英文書籍格式 D:BooK title · (1993)Location:Publisher ·
- 9.中文書文集格式:作者(主編)(年代) · 書名 。 出版地點:出版商 。
- 10.英文書文集格式 A:Author , A,A.(Ed) . 。 (1995) Book title.Locaion:Publisher
- 11.英文書文集格式 8:Author , A.A.·& AuthorB.B.(Eds ·) · (1995)Book
title.Location :Publisher
- 12.中文百科全書或辭書格式:作者(主編)(年代) 。 書名(第 2 冊) · 出版地點 :出
版商 。
- 13.英文百科全書或辭書格式:Author , A.A,(Ed) . Title(3rd.ed.Vol.1) . Location
Publisher ·
- 14.中文翻譯書格式 A:原作者中文譯名(譯本出版年代) · 書名(版別)(譯者譯)
出版地點:出版商 , (原著出版年:1984 年)
- 15.中文翻譯書格式 B:書名(譯本出版年代):(譯者 譯) 、 , 出版地點:出版商(
著出版年:1984 年)
- 16.英文翻譯書格式 A::Author , A.A . (1996) Book title , · B.Author , Trans ·) .
Location:PUBlisher · (Orginal work publish 1983 ·)
- 17.中文書文集文章格式 A:作者(年代) , 文章名稱 · 載於文集作者 (主編) , 書
名(頁別) , 出版地點:出版商 ·
- 18.中文書文集文章格式 B:作者(年代) , 文章名稱 。 載於文集作者 (主編) , 書
名 . (章別) · 出版地點:出版商;
- 19.英文書文文章集格式 A:Author , A.A.; (1993) .Article title In B . B .
Author(Ed ·) , BooK title (pp , ×××-×××) . Location:Publisher ·
- 20.英文書文文章集格式 B:Author . A.A. (1993) ,Article title. In B .
BAuthor(Ed.) , BooK title(chap.3) :Location : Publisher ·

(三) 專門及研究報告

- 1.中文政府報告格式 A:單位(年代) 。 報告名稱(報告編號: ×× · 出版地:作版地:
作者或出版商 。
- 2.中文政府報告格式 B:作者(年代) 。 報告名稱(○○ 單位報告編號: ××) 。 出版地 :
作者或出版商 。
- 3.英文政府報告格式 A:Institute(1996).Report title (Rep.No) , Location :
Publisher ·
- 4.英文政府報告格式 B:Author . A.A . (1996).Report title(Rep,No) Locatiobn :
Publisher ·
- 5.ERIC 報 告 格 式 : Author .A.A (1995) Report title(ReportNo ·
××××-××××).Eugene OR Univerity of Oregon , ERIC Clearing house on ,
Educational Management. (EA ××××)

(四) 會議專刊或專題座談會論文

- 1.已出版之會議專刊文章格式:依性質分別與書文集或期刊格式相同。
- 2.中文專題研討會文章格式:作者:年月,論文名稱·研討會主持人 主持人·研討會主題。研討會名稱·舉行地點,
- 3.英文專題研討會文章格式:Author, A.A.1995.April, Report title.In B.B,Author·Chair, Symposium topic.Symposium title,Place·
- 4.中文會議發表論文格式:作者(年月)·論文名稱 會議名稱,會議地點。
- 5.英文會議發表論文格式:Author, AA.(1995, April) paper title.Paper Presented in the Meeting Title, Place·

(五) 學位論文:

- 1.DAI 微縮片格式:Author . A.A.(1995).Dissertation title,Dissertation Abstracts International,xx(xx)·xxxx, (University Microfilms No xx-xxx
- 2.DAI 原文格式:Author, A.A.(1995).Dissertation title·(Doctoral Dis-sertation, University Name.1995) . Dissertation Abstracts International xx.xxx·
- 3.中文未出版學位論文:作者(年代)·論文名稱·未出版碩或博士學位論文·大學名稱·大學地點。
- 4.英文未出版學位論文:Author, A.A.(1995) . Dissertation title.Unpublished doctoral dissertation.University Name . place

(六) 視聽媒體資料:

- 1.中文影片格式:製作人姓名(製作人)·導演姓名(導演)·(年代)·影片名稱【影片】(影片來源·及詳細地址),
- 2.英文影片格式:Author . A.A.(Producer)、. Author . B.B . . Director)(1995) .Film title [Film] · (Avail from Company Name . Address)
- 3.中文電視節目格式:節目製作人姓名(製作人), (年月日), 節目名稱·電視台地點:電視台名稱。
- 4.英文電視節目格式:Author, A.A.(Executive Producer), (1996.May 1) program title.Place:Television Company

(七) 電子媒體資料:

- 1.中文線上查詢格式 A:作者(年代)·文章名稱·期刊名稱【線上查詢】·期別·線上查詢的詳細程序。
- 2.中文線上查詢格式 B:作者(年代)文章名稱【線上查詢】·線上查詢的詳細程序。
- 3.英文線上查詢格式 A:Author, A.A.(1996), . Article title.Periodical Name[on-line]xxx· Available:Specify path .
- 4.英文線上查詢格式 B :Author, A.A.(1995).Title[On-line].Available: specify path ·
- 5.中文 CD-ROM 文章摘要格式:作者(年代)·文章名稱【光碟】期刊名稱·期別,頁別·光碟資料庫別以及編號

6.英文 CD-ROM 文章摘要格式:Author,A,A.(1995) .Article title[CD-ROM] .
Journal Title · xx xx-xxxx . Abstract from:Source and retrieval number ·

(八) 法令:

- 1.中文法令格式 A:法令名稱(公布或發布年代)
- 2.中文法令格式 B:法令名稱(修正公布或發布年代)
- 3.英文法院判例格式:Name vs.Name , Volume Source Page(Court data) ,
- 4.英文法令格式:Name of Act,Volume Source §××× (1995) ·

972 教育測驗與評量研究讀書會紀錄（三）

◎主 題：常模、信度與相關期刊論文

◎時 間： 2009/03/06 ◎地 點：社 4042-1

◎導讀人員：王薪惠 ◎紀錄人員：陳香廷

◎內容紀錄：

信 度

一、**Definition**：是指測驗分數的可靠程度。信度是指測量的一致性（consistency）
測驗信度是指在使用測驗來測量個體心智能力時，所得測驗分數其可靠的程度。

二、**信度的意義**：

- （一）、信度是指評鑑工具所獲得「結果」的可靠性，而非指工具本身。（任何一項工具，均有不同的信度，會因 Sample 之性質及測驗時情況之不同而異）。
- （二）、每一信度的估計值，僅指某一特定類型的一致性，而非一般的一致性。
- （三）、**信度是效度的必需條件，但不是充分條件**（Reliability is a necessary but not sufficient condition for validity）。**信度不高，則效度亦不高；但高的信度並不保證必然有高的效度**，因為所編製的測驗可能無法測量到所預期的目標。例如：「普通常識測驗」如果用詞太難或句型結構太複雜，則儘管有高信度，但已變成「閱讀能力測驗」，自然沒有理想的效度。
- （四）、「效度」的考驗方法包括邏輯的分析與統計的應用，而「信度」的考驗則完全採用統計方法。

三、**信度理論模式：真分數理論的信度理論模式。**

$$O = T + E$$

O：觀察分數 or 實得分數（Obtained score）

T：真分數（True score）

E：測量誤差分數（Measurement error score）

測量誤差分數與測驗分數的信度成反比。

最理想的信度 = 零誤差的測量
信度指數 (Reliability Index, P_{or})

四、信度的種類與考驗方法：決定信度最理想的方法是在相同的情境下去獲得兩組測驗分數，再比較兩組分數的差異情形。

信度種類：

(一)、重測信度法 (Test-retest Reliability)：將同一個測驗以同樣方法，
卻

在不同時間重複實施同一樣本，即同一個測驗，對同一受試者施測二次。

- 又稱穩定係數。
- 完全正相關 $r = 1.00$ ，完全負相關 $= -1.00$ 。
- 一般標準化的性向測驗和成就測驗，在同一年內實施兩次，其 test-retest reliability 大多在 0.80 以上。
- 重測時間：不宜太短 (1~2 天內)，不宜太長 (1 年以上)。

The more the results will be influenced by change in the Pupil characteristic being measured and the smaller the Reliability coefficient will be。

The best time interval between tests will depend largely on the use to be made of the results。

(必須視測驗之性質和目的而定)

- 教師自編測驗不太可能求得 test-retest reliability。
- 當選擇標準化測驗時，測驗分數的穩定性是一個重要的標準。
- 一年內的穩定性較高，必須檢視實施測驗日期及該測驗的穩定性，才能作為輔導學生之依據 (例如：性向測驗)。

(一) 複本信度 (Equivalent-forms Reliability)：以複本法來估計信度，要
以

等值 (題數、型式、內容、難度及鑑別度均一致) 但不同題目的測驗，來測量同一群受試者，此種相關係數即為「等值性量數」，可指出兩個測驗能測出相同的行為層面之程度。(Give two forms of the test to the same group in close succession)。

- 複本法 (Equivalent-forms method) 無法指出學生在所測量特質之穩定性。
- 複本信度係數是表示試題之取樣是否具有充份的代表性。
- 教師選擇標準化測驗時，如果發現有複本 (A, B)，卻沒有提出複本信度或是複本信度很低，則這兩份測驗是無法相互替代使用

的。

(二) 內部一致性相關信度

折半信度 (split-half Reliability) : Give test once score two equivalent halves of test (e.g. : odd items and even items) ; correct reliability coefficient to fit whole test by spearman-Brown formula) , 將測驗均勻地分割成兩個部分測驗 (e.g. : 奇數題/偶數題) , 再分別計算每一位受試者兩個部分測驗分數, 再以兩個部分測驗分數計算積差相關係數, 做為信度係數的估計。

— 因為 items 減半, 而低估了真正的測驗信度。

— 測驗信度在沒有複本而具只能實施一次測驗的情形下, 可以採用折半法來估計。

— 以折半法來估計 reliability 確實較 Test-retest and Equivalent-forms 來得簡便。因為後者必需要實施兩次測驗, 不但較不經濟, 而且學生的動機, 疲勞、厭煩等因素均會增加實際估計信度的困難。因此一般教師自編測驗大多採用折半法來估計信度。

— 此種信度係數無法提供時間穩定性的資料。

Spearman-Brown 斯布校正公式：

$$\text{信度估計值 } r_{tt} = (2 r_{hh}) \div (1 + r_{hh})$$

$$\text{例如：} = 2 (0.60) \div [1 + (0.60)] \\ = 0.75$$

庫李信度 (Kuder and Richardson formula 20 ; KR 20)

— 根據所有題目之一致性來估計信度。

— Give test once。Score total test and apply kuder-Richarson formula。

盧氏公式 (Rulon) :

$$r_{tt} = 1 - \frac{\text{兩半測驗之總分的}}{\text{兩半測驗之差的變異數}} (S_d^2) / (S_t^2)$$

蓋氏公式 (Guttman) :

$$r_{tt} = 2 - \left[1 - \frac{S_a^2 + S_b^2}{S_t^2} \right] \\ = 2 \left[1 - \frac{(S_a^2 + S_b^2)}{S_t^2} \right]$$

五、變異數與標準差

- **變異數** (variance) (SD^2) : 一組分數與平均數距離的遠近。變異數是每

個分數與平均數差異的平方的平均數。

計算變異數 Steps：

- (1)、離差分數 (deviation Score)：每個分數與平均數的差 (X - M)。
- (2)、離差分數平方 (Squared deviation Score) → (X - M)²。
- (3)、離差分數平方和 (sum of squared deviations) → SS = Σ (X - M)²
(離均差平方和)
- (4)、變異數：將離差分數平方和除以總數→

$$\text{Variance}(SD^2) = \frac{\Sigma (X - M)^2}{N}$$

- 標準差 (Standard Deviation, SD)：標準差的大小，就能瞭解分數分配的分散情形。標準差等於變異數正的平方根。

$$SD = \sqrt{\frac{\Sigma (X - M)^2}{N}}$$

$$SD = \sqrt{\frac{SS}{N}}$$

◎ 如何計算變異數和標準差？

個別分數	—	全班平均數	=	離差分數	離差分數平方
80		75.2		4.8	23.04
75		75.2		- 0.2	0.04
95		75.2		19.8	392.04
45		75.2		- 30.2	912.04
60		75.2		- 15.2	231.04
72		75.2		- 3.2	10.24
85		75.2		9.8	96.04
90		75.2		14.8	219.04
96		75.2		20.8	432.64
54		75.2		- 21.2	449.44
N=10 Σ = 752		Mean = 75.2		Σ = 0	Σ = 2765.6

$$\text{Variance}(SD^2) = \frac{2765.6}{10} = 276.56$$

$$SD = \sqrt{276.56} = 16.63$$

六、測量標準誤 SE (Standard Error of measurement)：信度愈高，測量標準誤；反之信度愈低，測量標準誤愈大。

$$SE (\text{Standard Error of measurement}) = SD \sqrt{1 - P_{xx}} = SD \sqrt{1 - r_{xx}}$$

例如： $16.63\sqrt{1-0.7} = 9.11$

◎解釋測驗分數應該以「一段分數」(a band of scores)而非以一個特定的分數來代表學生的表現。測量標準誤越大，則「一段分數」的範圍也就愈大，這樣對我們的信心就愈小。

◎當測驗題數逐漸加長，其誤差之比例漸小。例如：當 items=50 題測驗之 SE = 4，而 items = 100 之 SE = 5。雖然試題增加了一倍，但標準誤則僅增加四分之一 → **測驗題數越多，其結果越可靠。**

◎以測量標準誤來表示測驗結果的信度有兩個優點：

(1)、SE 之單位與測驗分數相同，故在解釋個別 Score 時，可直接指出所允

許的誤差範圍。

例如：95% 的信賴區間 (Confidence interval, CI) → 當有 95% 的信心，可以估計學生的成績落在此段分數。

Z	Area between mean and Z	Area beyond Z	Confidence Interval, CI
Z = 1.96	0.4750	0.0250	95 %
Z = 2.58	0.4951	0.0143	99 %

(2)、不同群體間的測量標準誤比信度係數穩定，比較不受群體變異性的影響。

◎缺點：

- (1) 無法比較不同測驗間信度之高低 (除非轉換成相同單位，如均轉換成 T 分數)。
- (2) 如果要評鑑測驗品質及比較不同測驗信度之高低一定要採用信度係數 (correlation coefficient)。只有在解釋個別分數時，才採用測量標準誤。

七、影響信度的因素：

- (一)、測驗的長度：測驗越長，其信度越高。若受時間限制或是受試者年齡太小或是特殊的 Sample，不得不採用較短的測驗。在這種情況下，最好多實施幾次測驗，便可獲得較可靠的結果。
- (二)、分數的分佈情形：分數分佈範圍愈大，其信度估計值就越高。

Group A (信度高)	Group B (信度較低)
95	95
90	94

86		93
82		93
76		92
65		91
60		89
56		88
53		86
47		85

如何加大分數的分散範圍？

- (1)、編製較困難的常模參照測驗。
 - (2)、著重在複雜學習結果的測量（例如：Problem solving evaluation.....）。
 - (3)、增加題目數。
- (三)、測驗的難度：

容易的 test	困難的 test	理想的 test
100	100	100
75	75	75
50	50	50
25	25	25
0	0	0
Mean = 88	Mean = 13	Mean = 50
全距 = 25	全距 = 25	全距 = 100
r = .50	r = .50	r = .90

測驗分數的分佈與信度的關係（Gronlund，1976）

- ◎理想的常模參照測驗，其試題的難度要能使受試者的得分，散佈在整個量表上，才能提高信度。
- ◎100 題的 True/False 是非題中，其「期望的機遇分數」是 50，總分是 100，因此理想的平均分數為 75。
- ◎100 題的五選一之選擇題（multiple choice），其「期望的機率分數」是 20，Total = 100，二者之中點為 60，故理想的平均分數為 60。
- ◎全體受試者得分之散佈範圍趨於最大，這樣可提高測驗結果的可靠性。
- ◎一般編製嚴謹的 Standard test，均會提出不同年級的難度資料，以供選擇測驗之參考。

例如：選擇較低年級水準的 test 來測量啓智班學生。
 選擇較高年級水準的 test 來測量資優班學生。

- (四)、客觀性：指評分結果的一致性。
- (1)、選擇、是非、填充、配合等，均屬於客觀類型，富有客觀性不受評分者判斷與意見影響。
 - (2)、論文式試題 or 觀察評分法：主觀類型，主觀評量法。須儘量使評量程序客觀化。
 - 1、根據所預期的學習結果來評量論文的答案。
 - 2、以「分數記分法」來評量「限制反應式」論文題的答案，並使用模擬答案作為評分的指引。
 - 3、以「等級評定法」來評量「擴展反應式」論文題的答案，並以明確的標準作為指引。
 - 4、閱卷時應逐題評閱，要等到所有學生同一題之答案均閱完畢，才評閱下一題答案。
 - 5、應避免事先知道學生的姓名。（彌封 or 將姓名寫在背面）
 - 6、可能的話，每一題宜由兩人以上來評分。（例如：老師交換批閱，重復批閱）。

972 教育測驗與評量研究讀書會紀錄（四）

◎主 題：效度與相關期刊論文

◎時 間： 2009/03/13 ◎地 點：社 4042-1

◎導讀人員：陳怡安 ◎紀錄人員：王薪惠

◎內容紀錄：

效 度

一、**Definition**：效度（Validity）是指一個測驗能測量到它所欲測量的特質或功能的程度，效度高表示該測驗能達到它所要測量的目標。測驗的效度，又稱之為測驗的正確性。研究測驗效度的目的，就在了解測驗分數的意義與內涵，所以效度可說是探究測驗分數意義過程（Messick, 1989）。

二、 效度的性質（Gronlund & Linn, 1990）

（一）、效度是解釋某一團體之測驗或評量工具之結果的適合性（appropriateness），而不是指工具本身。事實上「測驗效度」應該是「解釋測驗結果的效度」。

（二）、效度是程度上的問題（a matter of degree），而非「全有或全無」（all-or-none）。我們應該用程度去描述效度，如高效度、中效度及低效度。

（三）、效度對於某些特殊的用途或解釋都是獨特的，沒有一個測驗對於所有的目的都是有效的。當我們在評鑑或描述某一個測驗效度時，需考慮到該測驗結果的特殊用途（例如：選擇、安置或學習評鑑等）（Validity is always specific to some particular use）。

（四）、效度不是由實際測量而得，而是從已有的證據去推論而得。效度是一種單一的概念（Validity is a unitary concept）。

三、影響效度的因素：

- （一）、不清楚的指導語。
- （二）、詞彙或句子結構太難。
- （三）、測驗的題目難易度不適當。
- （四）、測驗題目粗製濫造。
- （五）、試題曖昧不明。

- (六)、試題無法測量真正的成果。
- (七)、答題時間太短。
- (八)、測驗題目太少。
- (九)、試題安排不妥當。例如：先出現較難的試題。
- (十)、答案項安排不當。

三、效度的種類：

(一)、內容效度 (content validity)：是指測驗內容的代表性或試題取樣的適切性。(Content validation is the process of determining the extent to which a set of test tasks provides a relevant and representative . Sample of the domain of tasks under consideration. 總之：「內容效度是一個測驗能否測量到具有代表性的教材內容和所預期的行為改變」。例如：使用托福測驗量考生的英語能力。每年入學考試結束，補習班就會分析各科試題，每一冊出現的比例，每種題型（記憶、理解、分析、綜合、評鑑）的佔分比例，就是驗證該考試科目的內容效度。

1. 提高內容效度的方法：內容效度特別適用於成就測驗的效度考驗。在編製成就測驗，如能遵循下列步驟，將可提高內容效度：
 - (1) 分別列出教材內容的各項重點和所要測量的各類學習結果。
 - (2) 各項教材重點各類的行為改變（學習結果）要以相對的重要性來加權。
 - (3) 編製一個如下表那樣的細目表。表中各細格的數字代表各項教材重點及行為改變的相對權數。
 - (4) 根據細目表來編製測驗，所編製的測驗越是符合細目表的比重，則內容效度愈高。

國小自然科成就測驗雙向細目表

教材內容大綱 (教學內容)	(教學目標) 行為改變 (Instructional objectives)			Total
	Knows concepts 概念的介紹	Comprehends concepts 概念的理解	Applied Concepts 概念的應用	
植 物	8	4	4	16
動 物	10	5	5	20
天 氣	12	8	8	28
地 球	12	4	2	18
天 空	8	4	6	18
Total	50	25	25	100

* 內容效度又稱為**合理效度**（Rational validity）或**邏輯效度**（Logical validity）。因為要評鑑內容效度，需要相當時間來進行測驗內容的邏輯分析和比較，故稱之 logical validity。

2. **考驗**內容效度的方法：編製雙向細目表（ie，教學內容&教學目標），利用細目表來檢驗內容效度。表中的百分比指出在編製測驗時各教學內容及各類教學目標（行為改變）的相對重要程度。如果我們將試題逐步分析和歸類，編製一個測驗試題分類表，則只要比較「雙向細目表」和「試題分類表」各細格（cell）題數之比率是否類似，即可了解該測驗內容效度的高低。
3. 內容效度的通用範圍：內容效度特別適用於成就測驗，但是在性向、興趣、態度及個人—社會適應等測量量表的編製也需考慮到內容效度的高低。編製量時，要符合所要測量的目標，先分析測驗的材料和所欲測量的結果，然後判斷二者符合的程度。
4. **表面效度**（face validity）：指一個測驗從外表看來，直覺地認為這個測驗是在測量什麼，完全與內容效度不同。嚴格說來，face validity 不能算是一種效度，但就測驗實施而言，face validity 也是很重要的。如果受試者在感覺上認為主試者所實施的測驗與測驗目的無關，則可能會排斥該項測驗，而影響測驗的功能。表面效度雖有其必要性，但不能視為測驗的真正效度。

（二）**效標關連效度**（Criterion-related validity）：測驗分數與測驗所欲測量之特質的外在指標（external indicator）之間的相關程度。所謂外在指標，稱之為**效標**(Criterion)。它代表測驗所要呈現的行為、表現或統計量數。效標是考驗效度的一個參照標準。效標分數和測驗分數如大約在同一時間獲得，即為「同時效度」（concurrent validity）；反之，如果效標分數和測驗分數二者獲得之時間有很大差異（通常測驗分數在前，效標分數在後），即為「預測效度」（Predictive Validity）。

* 測驗實務上，常經由實徵性的統計方法計算出測驗分數與效標之間相關係數，所以效標關連效度，又稱之為「實徵或統計效度」（empirical or statistical validity）。兩者相關係數愈高，代表測驗的 criterion-related validity 較高，也代表測驗分數愈能解釋效標。

1. 效標的種類：

1. 學業成績（例如：期中考、期末考、聯考成績常是智力測驗、性向測驗、成就測驗的效標。）
2. 具代表性的測驗得分—公認好測驗 or 代表性的測驗（例如：比西智力量表）作為自編智力測驗之效標。
- (3) 現時的工作或行為表現（例如：可以用學生操作機械的表現，作

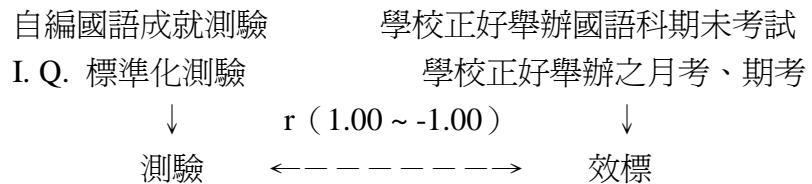
為機械性向測驗的效標。此種效標大都用來驗證「同時效度」。

- (4) 未來的訓練成績或行為表現—例如：職業訓練成績，教師教學表現可做為各種性向測驗的效標。→印證「預測效度」。
- (5) 他人評等—例如：教師對學生的評等，精神科醫師對病患人格異常症狀的評等，常是人格測驗的良好效標。

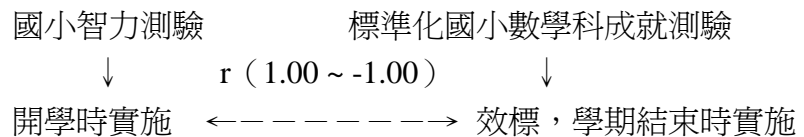
2. 效標關聯效度的種類：

- (1) 同時效度 (concurrent validity)：測驗分數與測驗施測同時取得之效標之間的相關程度。此種相關，大都使用積差相關 (product moment correlation)。代表同時效度的相關係數愈大，代表測驗分數愈能解釋效標所呈現的行為。

例如：



- (2) 預測效度「Predictive Validity」：測驗分數與測驗施測後一段時間所取得效標之間的相關程度。通常性向測驗、成就測驗、智力測驗常會使用 Predictive validity 來驗證其效度，但此種效標取得不易，因為受試者常流失。例如：某老師要考驗學術性向測驗分數（如：國小智力測驗）能否有效預小六的數學成績。



如果國小智力測驗（學術性向）較高的學生，其數學成就亦高，而學術性向較低的學生，其數學成就亦低，那就表示學術性向測驗對數學具有預測力。也就是學術性向測驗結果具有效標關連效度。

- (三) 建構效度 (Construct Validity)：指一個測驗能夠測量到它所欲測量之建構或特質的程度；易言之，一個測驗能符合其賴以編製之理論的程度，就稱之。因此凡是依據心理學在建構，來解釋和分析測驗分析所代表的意義，即為建構效度的考驗。

「**建構**」(Construct)，就是用來解釋人類行為的理論架構或心理特質。例如：「推理能力」、「智力」、「科學態度」、「批評思考」、「閱讀理解」、「研究能力」、「性向」、「動機」、「焦慮」、「社會性」、「內向性」、「外向性」、「自尊」...等心理特質均屬之。這些心理特質均是抽象的，假設性的特質，無法觸摸到。但每一個特質，都有其心理學上的理論基礎，可據以描述或解釋人類的行為。

1. 考驗建構效度的步驟：

- (1) 建立理論架構，以解釋受試者在測驗上的表現。
- (2) 根據理論架構，推演出各種有關測驗成績的假設。
- (3) 以邏輯和實徵研究的方法來證驗假設。

例如：說明建構效度的考驗方法，ex：我們要考驗所編的智力測驗，則從智力建構理論中，可以導出下列預測：

- (1) 測驗分數隨著 age 增加。
- (2) 測驗分數能預測學業成就。
- (3) 測驗分數和教師評定的智力等級有正相關。
- (4) 測驗分數和其他智力測驗有正相關。
- (5) 測驗分數能鑑別不同能力的 Groups（例如：資優生 & 智能不足）。
- (6) Score 和道德判斷能力有正相關。
- (7) Score 和 SES 有正相關。.....etc。

可以採用實徵研究的方法搜集資料。逐一加以驗證，如果每項預測得到支持，則綜合這些證據。即可確定新編的智力測驗所測量的行為與努力的建構相符合，也就是具有「建構效度」。

2. 考驗建構效度的方法：

- (1) 分析受試者解答試題時所需的心路歷程。
- (2) 比較不同群體在測驗分數上的差異。
- (3) 比較實驗處理前後分數的差異。（差異檢定－單因子變異分析）。
- (4) 求測驗分數與其他測驗分數的相關。（相關檢定－Correlation）。
- (5) 因素分析法。（factor Analysis）

五、效度與信度的關係：效度和信度是評估測驗品質兩個重要的要素，二者缺一不可。

Reliability：是評估 test scores 的**誤差**程度。

Validity：是評估 test scores 的**正確**程度。

高信度，未必有高效度，例如：某智力測驗實施多次，得到 Test-retest 達 0.90 以上，但若將這個測驗拿來測量受試者的焦慮，其效度當然是異常低。

高效度，必然高信度。

因為信度是效度的必要條件，而非充分條件。

因為**信度低**，效度一定低；**信度高**，效度不一定高。

效度高，信度一定高；**效度低**，信度不一定低。

972 教育測驗與評量研究讀書會紀錄（五）

◎主 題：試題分析與相關期刊論文

◎時 間： 2009/03/20 ◎地 點：社 4042-1

◎導讀人員：曾韻琴 ◎紀錄人員：陳香廷

◎內容紀錄：

試題分析

一、良好的測驗應具備下述十項品質：

- 1、適切性 (Relevance)：所選用的問題類型，要能適切地測量所期望的學習結果。
- 2、平衡性 (Balance)：每一教材領域所包括試題比率要符合細表中所預定的比率。
- 3、有效性 (Efficiency)：試題的編製、評分以及學生作答時間要經濟。
- 4、客觀性 (Objectivity)：試題要明晰，答案要明確，只要是該領域的專家，應可得到滿分或接近滿分。
- 5、特殊性 (Specificity)：即要正確解答試題，一定具備該領域的特殊知識。
- 6、適當的難度(Difficulty)：試題須難易適中。
- 7、良好的鑑別度(Discrimination)：試題要具有區別高低學習成就者的能力，成就較高者，其答對每一個試題的比率要高於成就低者。
- 8、信度(Reliability)：測驗的結果要有一致性或可靠性，不可因時間及內容之不同，而有太大的改變。
- 9、效度(Validity)：所編製或選用的工具要能真正的測得所欲測量之特質或學習結果。
- 10、非「速度性」(Speededness)：即非速度測驗的試題數量要配合作答的時間，使學生得分不會受到作答速度的影響。

二、題目分析(Item analysis)的功能：

1. 題目分析的資料有助於測驗結果的討論。
2. 題目分析的資料可作為實施補救教學的依據。
3. 題目分析的資料可作為改進班級教學的依據。
4. 題目分析的程序可增進編製測驗的技能。

5. 題目分析可以縮短測驗，以節省測驗的時間，也可以提高測驗的信度和效度。
- *** 試題分析可以分為「質的分析」(Qualitative analysis)與「量的分析」(Quantitative analysis)。「質的分析」是分析試題的內容和形式，包括「內容效度」(content validity)分析。「量的分析」是採統計方法來分析題目的品質，包括「難度」(Difficulty)及「鑑別度」(Discrimination)分析。一個測驗信度與效度之高低完全取決於試題的品質，故經 item analysis，可提高測驗的信度與效度。

三、難度分析 (Difficulty Analysis)：

(一)、以通過百分比表示難度：題目愈容易，則通過的百分比愈高。

【公式一】

$$P(\text{試題難度}) = R(\text{答對該題之人數}) / N(\text{全體受試學生})$$

例如：100名學生中，答對 item#1 的有40名學生，則 item#1 的難度為：

$$P = 40 / 100 = 40 (40\%)$$

另一種求難度方法以高分組答對百分比加低分組答對百分比再除以2。

H 高分組：取最高27%

L 低分組：取最低27%

【公式二】

$$P = (P_H + P_L) / 2$$

P_H ：高分組答對該題的百分比

P_L ：低分組答對該題的百分比

例如：在100個學生當中，被選為高、低分組的各有27人，其中高分組有20人答對第一題($P = 20/27 = 0.74$)，低分組有10人答對第一題($P = 10/27 = 0.37$)，則第一題的難度為：

$$P = (0.74 + 0.37) / 2 = 0.56 (56\%)$$

說明：

(1) P 值越大，則題目越容易

例如：P = 0.6 的 item easier than P = 0.40 的 item。

(2) 通常我們均依據題目的難度，由易而難，順序排序。這樣可以增加受試者的信心，也可以避免受試者浪費時間來解答超出能力範圍的 items，而遺漏了很多可以答對的 items。

(3) 編製測驗，Difficulty analysis 的主要理由是爲了要選取難度適當的 items。大多數的 standard test，均希望準確地區分學生在某種特質上的差異，如果所編製的 test 沒有人能答對或

是所有 items 全部答對，則這些 items 就無法提供受試者個別差異的資料，也不會影響測驗分數的離差，因此對測驗的信度和效度沒有任何貢獻。

(4) 難度指數 (P) 越接近 1.00 or 0.00，就無法區分受試者能力的差異；相反地，P 值越接近 0.50，則區別度越高。

(5) 爲了使測驗具有最大區別力，必須選擇 P 在 0.50 左右的試題以組成測驗較佳。

(二)、試題難度與測驗分數分佈的關係

如果受試者具有代表性，則測驗分數的分佈情形應該接近常態分配 (Normal Distribution)，然而如果所獲得的曲線 (Curve) 並不是常態的，得分偏高或偏低，就是「偏態分佈」(Skewness)。只有在常態分佈的情形下，才能使受試者在測驗上得分之個別差異達到最大。但目前各級學校段考 (尤其是小學)，主要目的是了解學童的精熟度，故區分學童程度之高低並不是主要目的，故沒有必要讓學生的得分成 Normal Distribution。

(三)、試題難度和常模參照測驗

1、由於常模參照測驗 (Norm-referenced test) 是設計來評定學生成就之高低順序，故需儘量使學生得分之分散範圍加大。爲了加大學生得分之範圍，需將所有學生都答對或都答錯的題目去掉，而選擇那些能使學生測驗分數之差異達到最大的 items 來組成 test。

2、當平均分數接近可能得分之中點 (Mean = Median)，而且分散的範圍從「接近 0」至「接近滿分」時，學生的個別差異可以達到最大。

3、選擇型測驗，因有猜對的可能，故理想的平均得分應高於 50 分 (滿分 100 分)。一般而言「期望的機會分數」(猜對的比率) 與「最高的可能分數」(測驗總分) 之中點，是理想的平均分數。

4、除了 test 最初幾題，可以選很容易的題目，以提高作答的動機外，其餘 Items 均不能太容易或太難。

Table 5-1 各類型選擇型測驗的理想平均難度

	機會分數	平均難度
二選一試題(False/Ture)	5 0	7 5
三選一選擇題	3 3	6 7
四選一選擇題	2 5	6 3
五選一選擇題	2 0	6 0

六選一選擇題	1 7	5 9
七選一選擇題	1 4	5 7

(四)、試題難度與測驗目的的關係

1、對於某些特殊目的的測驗來說，試題的難度就不宜集中在 0.5 0 左右，而應視 test 的 Purpose 而定。一般而言，用來甄選人員的 test，其試題的難度宜接近錄取率。

例如：

- (1)、如果要選擇最優秀的 20% 應徵者，則 P 就宜集中在 0.20 左右。
- (2)、如果要甄選前 10% 學生給予獎學金，則試題之平均 P 值要接近 0.10。
- (3)、如果要選取 15% 的國中生做補救教學，則試題的難度就要接近在 0.85 左右。
- (4)、如果要了解學生之精熟程度, therefore, P 應該接近 0.8~0.9。
- (5)、如果要做預備性評量之用，P 值縱然接近 0，也不可刪除這些題目。

四、鑑別度分析(Discrimination Analysis)

試題鑑別度是試題分析用來描述試題鑑別能力的指數。它通常會以受試者作答試題結果與整個測驗分數結果一致程度的指數。

通常 Discrimination index 的計算方法區分為三大類：

- (一)、極端組群的方法
- (二)、內部一致性分析的方法
- (三)、刪題的方法

(一) 極端組群的方法（採用最多的題目分析法）

在常態分配下，最適當的比率是高低分組各佔 27%(kelley,1939)。如果分配的形狀較 Normal Curve 平坦，則高、低分組所佔的比率宜高於 27%，大約是 33%(Cureton, 1957)。在一般教師的測驗只要介於 25%~33%即可。不過，如果是標準化測驗的話，習慣仍用 27%。

公式： $D = P_H - P_L$

D：表試題鑑別度測量

P_H ：表高分組答對試題人數的百分比

P_L ：表低分組答對試題人數的百分比

Table 5-2 高分組及低分組答對試題人數百分比及試題鑑別度指數的計算

Student		試題得分						Total Score
Id #		X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	得分
高分組	04	1	1	1	1	1	1	6
	07	1	1	1	1	1	1	6
	08	1	1	1	1	1	1	6
	15	1	1	1	1	1	1	6
	18	1	1	1	1	1	1	6
P _H		1.00	1.00	1.00	1.00	1.00	1.00	
低分組	02	1	0	1	0	0	0	2
	13	0	0	0	1	1	1	3
	14	1	0	1	0	0	1	3
	17	1	1	1	0	0	0	3
	19	1	1	1	0	0	0	3
P _L		0.80	0.40	0.80	0.20	0.20	0.40	
D		0.20	0.60	0.20	0.80	0.80	0.60	
		$D = P_H - P_L$						

Table 5-3 簡單的題目分析表

item	高分組 (H) (20 人)	中分組 (M) (20 人)	低分組 (L) (20 人)	難度 (H+M+L)	鑑別度 (H-L)
1	15	9	7	31	8
2	20	20	16	56*	4
3	19	18	9	46	10
4	10	11	16	37	-6*
5	11	13	11	35	0*
6	16	14	9	39	7
7	5	0	0	5*	5
:	:	:	:	:	:
:	:	:	:	:	:
75	:	:	:	:	:

*表示需要提出來討論的題目 (取自 Anastasi, 1976, P.209)

說明：

- (1) item# 2 太容易—60 人中有 56 人答對。
 item# 7 太難—60 人中，只有 5 人答對。
 item# 4 之鑑別度是 -6，item# 5 之鑑別度是 0，必須加以修正或刪除。

- (2) 教師自編 test 的 item analysis 主要目的是在找出測驗的缺點和教學的缺點。和全班學生討論通常能診斷出問題的所在。如果用字遣詞不當，應加以修改或刪除。有時候討論結果發現試題本身並沒有缺點，而是學生對所測量的概念並不太了解。如此情形下，教師需進行再教學，以澄清學生之觀念。
- (3) D 值介於 1.00~ -1.00 之間，D 值的平均數越高，則測驗的信度就越高。

(二) 內部一致性分析的方法

此即常見的「諧度分析」的方法。其目的在分析個別試題得分與整個測驗分數作用的一致性。

常見的計算方法：

- (1) 計算試題分數與其測驗 Total Score 的「關聯性指數」，常見的有「二系列相關」(Biserial correlation)或「點二系列相關」(Point-biserial correlation)係數，以求得的相關係數(γ_{bis} or γ_{pb})，以表示內部一致性的高低，亦即試題鑑別度指數。

$$\text{公式：} \quad \gamma_{bis} = \frac{\overline{X_1} - \overline{X_0}}{S} \cdot \frac{P(1-P)}{Y}$$

$$\gamma_{pb} = \frac{\overline{X_1} - \overline{X_0}}{S} \cdot \sqrt{P(1-P)}$$

$\overline{X_1}$ ：答對試題組測驗分數之平均數

$\overline{X_0}$ ：答錯試題組測驗分數之平均數

S：測驗分數之標準差

P：全體受試者之答對率

Y：標準常態分配下，累積機率為 P 時，Normal distribution 之高度。

- (2) 計算試題分數與其 Total Score 之積差相關係數 (Product moment correlation)。

(三) 刪題的方法

此種方法是在 computer 廣泛應用後，普遍被採用的方法。

方法：

- (1)、先計算整個測驗的信度係數(α_t)
- (2)、有系統地逐一排除全體測驗中的一題，再重新計算刪題後的信度係數(α_{t-i})。
- (3)、再觀察 α_t 與 α_{t-i} 的差距有多大。

(4)、如果差距很大時，表示被刪除的題度是一個影響信度極大的 item，不宜輕易 delete；反之，即表示該 item 並不影響整個測驗的信度。

* 刪除與 Total Score 相關較低的 items 是使測驗純淨化(purifying)或同質化(homogenizing)的好方法，因為所保留下來的 items，其平均交互相關必然最高，也就是各題目均在測驗同一特質。

五、常模參照測驗題目分析的解釋

- (一) 以測驗總分為依據所求得的鑑別度未必能代表試題的效度。
- (二) 鑑別度指數低的試題未必表試題具有缺點。例如：在測驗理解層次的學習結果，可能鑑別度較低，但仍不宜刪除。因為若 delete 則無法再測驗到理解層次的學習結果。
- (三) 課堂測驗的 item analysis information，只是暫時性的，而非固定不變的特質。試題之價值取決於邏輯的分析，而非統計的分析。

六、題目分析實例

A、固定反應題 (Fixed-response items)

優點：1、適用廣泛。2、計分客觀。3、有效率。

(一)、**是非題**(true-false test)：一個良好的 True/False，其難度在 0.75 左右，其鑑別度約在 0.20 左右。

是非題的否定暗示效果(ie., 錯誤答案對學生的負面影響)以及猜題因素，通常被視為是非題最大的限制。

1、是非題編製要點：

* 咸認是非題是最容易準備的測驗之一，事實上編製不易，要特別用心，以內容而非形式來決定答案。用詞遣字應該是編製者注重的重心，避免提供答題線索。不必要的隱藏字義。藉由練習與細心，可以達到平衡。下面的建議對是非題、選擇題皆有助益：

(1)、避免使用特殊的限定字詞作為線索：

例如：強烈措詞——所有的、總是、決不、不、皆非或沒有一個。

中度措詞——很多、有些、有時候、時常、常常、通常和一般說來。

(2)、避免對與錯的陳述，出現題數不相稱的現象：

a、若干研究發現，答案是「錯」的題目之效度，稍高於答案是「對」的題目，因為許多受試者會有「默從反應類型」，而選「對」。

b、一個 test 的錯誤敘述，應該比正確的多 (cronbach, 1942)。

- c、艾柏 (Ebel, 1965b) 建議 60% 的題目，應該選「錯」。
- (3)、避免照抄教科書上的句子。
 - (4)、避免陷阱題。
 - (5)、一個敘述題測驗一個重點。不要用兩個或以上的刺激來引出一個答案。
例如：鄭成功是「民族英雄」與「詩人」。
 - (6)、避免過度使用否定字詞，以免語意混淆。
 - (7)、避免曖昧的字彙與敘述。
 - (8)、避免艱難字彙和不必要的複雜句子結構。
 - (9)、盡可能用最簡單的方法來標示答案。例如：「0」「x」或「是」「否」。
 - (10)、僅在這類型不會造成主題模糊的情況下使用之。重大的測驗極少只有是非題。

(二)、**複選題** (multiple-choice)：複選題是選擇題之一種變形，每一個選項必須獨立判斷是否同意，也因此，在進行題目分析時，需將各選題分視為獨立的是非題來分析。

	R _H	R _L	D	P
1	10	6	.40	.80
2	6	9	-.30	.75
3	9	7	.20	.80
4	5	10	-.50	.75
5	7	3	.40	.50

標準答案: (1)(3)(5)

R_H：高分組選答那個選項之人數。

R_L：低分組選答那個選項之人數。

$$P = R / N$$

$$P = (R_H + R_L) / N$$

$$D = R_H - R_L$$

說明：1、選項 (2) & (4) 是錯誤的，故其 D 值是負的，第 (1) (3) (5) 的選項是正確的，故其 D 值是正的。

2、具有此種特性的題目，是良好的複選題。

選擇題編製要點：

- (1) 題幹應該包含中心問題和所有屬性，包括每個選重複出現的字彙。
- (2) 題目越短越好，盡可能清晰。
- (3) 要謹慎使用否定陳述的題幹。

例如：「下列人名中，只有一個不是民族英雄，他是誰？」

- (4) 在題幹中，把問題完整的表達出來。
- (5) 尋求最佳答案。

例如：「導致美國加入第二次世界大戰，最重要的原因是、、、？」。
- (6) 題目的閱讀與語文難度，應該要低。
- (7) 盡可能的，用邏輯順序來安排選項。
- (8) 盡可能的統一各選項的內容、形式、長度、明確性與文法結構。
- (9) 一題要有三個以上的選項，3~5 為佳，對小三或以下的受試者選項可以少一些。
- (10) 避免使用課本上的措詞或者是有刻印象的用詞。
- (11) 題目應該避免透露其他題目的答案。
- (12) 避免使用特殊字，比如「通常」、「決不」等。因為有「通常」這個字眼的選項，較常是正確答案。
- (13) 避免照課本安排題目的順序。
- (14) 一個測驗不要包含太多題目，否則會變成速度測驗。
- (15) 「以上皆非」可以用為正確答案之最後選項(Williamson & Hopkins, 1967)，用了它，會使得題目難度提高(Choppin, 1974)。
- (16) 「以上皆是」「以上至少有一個正確」的選項，有時候很好用。因為猜對的機會大大的降低。
- (17) 、一個選項，分做一段，除非所有選項都很短，可以空成一行。
- (18) 、用數字代題號(例如：1、2、3)，字母作選項(例如：a、b、c)

例如：如果 h, k, m, n 都是正數， k 比 m 大， n 比 h 大，下列何者為正確？

$$1、n + h = k + m$$
$$2、k + h = n + m$$
$$3、k + n = m + h$$

a) 2 b) 1, 2 c) 3 d) 以上皆是
- (19) 、選項的標點要正確。

(三)、**配合題**(matching-exercises)：

通常有兩欄，第一欄是題幹，第二欄是選項。配合題的選項通常比題幹多，這樣才能降低受試者的猜對的機會。第一欄題幹有時候是不完整的句子，需要第二欄的字或詞來完成整個句子。配合題在測量專有名詞，事實知識、地理，圖表或表格方面，非常有用。配合題通常只能測量知識層次的教育目標。

配合題編製要點：

- (1) 每個 matched exercise，僅包含同質的題材。
- (2) 仔細檢查每一題，配合項是否透露答案線索。
- (3) 確定學生了解配合題的原理。
- (4) 把題目放在左邊，用數字編號；把選項放在右邊，用字母順序排序。
- (5) 用系統化的順序來安排題目與選項。
- (6) 把配合題的題目跟選項放在同一頁。
- (7) 配合題的題目，限制在 10~15 個以內。

(四)、**簡答題** (Short-answer items)：

是種客觀測驗題，每一題都是用直接疑問句、關鍵字詞，特定的說明、特殊的說明、特殊的問題或是不完整的陳述句或問題等方式呈現。受試者必須自己提出 answer 而不是由教師所提供的選項中選擇。這類測驗題與論文題最大不同在於答題的深度與長度。簡答題的答案通常很短，可能是一個字、數個字或一個句子、數個句子。

優點：熟悉、自然，可以完全去除猜題。對於數學、物理科測驗，特別有價值，因為這些問題都是需要運算。這些以地圖、圖形、表格來表示，學生需要空間來填寫某部分的名稱，或是數字、字母，更能增進「長期記憶」(Long-term Memory)(Gay, 1980)。

972 教育測驗與評量研究讀書會紀錄（六）

◎主 題：行為檢核表編製與使用與相關期刊論文

◎時 間： 2009/03/27 ◎地 點：社 4042-1

◎導讀人員：林曉瑩 ◎紀錄人員：王薪惠

◎內容紀錄：

行為檢核表期刊論文

一、 標題 / 作者：

「自閉症兒童行為檢核表」之編製及相關研究
張正芬、王華沛（國立台灣師範大學）

二、 摘要：

此研究主要目的在發展一自閉症兒童的篩選與行為檢核工具。共分為「學齡前」、「國小低年級」以及「國小中年級以上」三種版本。題目內容包含溝通、社會性和行為興趣三個領域。題目分為四等第：從不如此、偶爾如此、經常如此、總是如此，並針對題目對自閉症狀的敏銳度予以加權。得分愈高代表自閉程度愈嚴重。

此檢核表的樣本包含台灣東西南北四區，取樣對象從學前至國中生，並包含自閉症兒童與一般兒童。信度考驗方面，進行內部一致性考驗，另外「國小中年級以上」版本還進行重測信度與評分者間信度的考驗。效度考驗則進行了「區辨性診斷」、「效標效度」與「建構效度」。

三、 理論背景：

自民國 79 年修訂的殘障福利法，我國才正式將自閉症者列為身心障礙者之一，但是自閉症兒童的就學問題並未受到國內太多的關注。民國 80、81 年間舉辦的第二次全國特殊兒童普查，首度將自閉症兒童列為普查對象之一。其採用的方法為透過家長、教師的初查，再經由精神科醫師的複查。但結果比預估的 0.05 %少了一千餘名左右。可能的原因除了一般家長、教師對自閉症認知不足外，也有可能是因為其被歸於其他障礙所致（張正芬，民 88）。

民國 88 年教育部公佈「身心障礙集資賦優異學生鑑定原則、鑑定基準」，其主要參考依據為美國精神醫學會所主編的精神疾病診斷及統計手冊第四版（DSM-IV），鑑定標準如下：（一）具有顯著的口語、非口語的溝通障礙（二）

具有顯著的社會互動障礙(三)表現固定而有限之行爲模式及興趣。並有以下四特點：(一)明確標示三個核心障礙，並列出不同發展階段的行爲標準(二)強調「質」的缺陷(三)不強調智力問題(四)發生年齡。

雖有上述依據，但自閉症兒童的個別差異相當顯著，一般教師對自閉症兒童難有正確的了解及辨識。而且我國目前所使用的鑑定工具僅有謝清芬、宋維村、徐澄清(民 72)所修訂的克氏行爲量表，其主要由家長填寫，題目內容包括知覺、活動量、物品操作、怪異行爲、社會性…等。但因題數太少，故效度不高。

至於國外的篩選與鑑定自閉症的工具具有以下幾項：

- (一) Autism Diagnostic Interview-Revised(ADI-R, Lord, et al, 1993; Lord, Rutter & Le Couteur, 1994)：以結構式訪談方式晤談自閉症兒童的主要照顧者，適用對象爲 18 個月至成年，主要分成溝通、社會性、及行爲興趣三部份。其缺點爲鑑定心智年齡 18 個月以下兒童時，易產生過度包含現象。
- (二) Autism Behavior Checklist(ABC, Krug, Arick, & Almond,1980)：內容共有 5 領域 57 題項，包括感官、人際關係、肢體及物體使用、語言、社會互動及生活等方面，使用者爲老師及家長。
- (三) Autism Diagnostic Observation Schedule (ADOS, Lord, et al, 1989)：主要評量六歲以上自閉症兒童之教室情境中的行爲，著重兒童社會互動與溝通。題項包括溝通、社會性互動、心情及非特定的異常行爲。
- (四) The Childhood Autism Rating Scale (CARS, Schopler, Reichler, & Renner,1988)：包含 15 個項目，4 個評分等級。切截點分數因年齡而異，兒童切截點分數較高。與 ADI-R 相較，其可診斷出較多的自閉症兒童，但評量者須經過臨床訓練才可減少評量時的誤差。
- (五) Gilliam Autism Rating Scale (GARS, Gilliam,1995)：依據 DSM-IV 的自閉症診斷標準編製而成，可由家長、教師、臨床工作者所使用的行爲評量表，主要內容包括刻板行爲、溝通、社會互動、發展性異常。其評量者堅信度、重測信度及表面信度均高。

本研究即主要以 DSM-IV 對自閉症的診斷標準爲藍本，並參考上述工具之內容，採結構化的觀察量表形式，評量向度則以溝通、社會性、行爲三項爲主，做爲編製篩選量表的架構。

四、 資料分析：

本研究之預試試題爲四點量表，分爲「從不如此」、「偶爾如此」、「經常如此」、「總是如此」。正向題指一般學生較會出現而自閉症學生較缺乏的行爲，而反向題即相反。正向題採反向計分。分數愈高代表愈不利，愈有可能是自閉症兒童。

五、 研究結果：

- (一) 預試使用 70 題的版本，剔除一般兒童與自閉症兒童未達 0.01 顯著差異的題目 4 題，並將檢核表分為甲乙兩式，甲式適用於學前及國小低年級學生，乙式適用國小中、高年級及國中階段的學生。
- (二) 常模資料取得後，將甲式進行第二次的預試分析，發現小一、二與中、小班間的差異顯著，因此將甲式再分為兩個量表，一為學前兒童使用，另一為低年級學生使用。三個量表均能有效區辨自閉症兒童與一般兒童，區辨力均達 90% 以上。
- (三) 三種版本均進行領域內題目與領域的內部一致性考驗，所有相關係數均達顯著水準。中年級檢核表還進行了間隔兩週和一年的重測信度，兩年間的分數差距極小，未達顯著水準，代表穩定性相當高，另外家長與教師的評分者間信度也達 0.81。
- (四) 三量表均能有效區辨自閉症兒童與一般兒童，區辨率均達 95%，具有構念效度。
- (五) 中高年級以上版本與 Gilliam Autism Rating Scale 的相關為 0.72，達 0.01 顯著水準。
- (六) 安置在普通班級者及口語能力較佳者，再自閉症兒童檢核表的三領域及總分得分均較低，顯示本量表具有理論上的建構效度。

六、 優缺點：

(一) 優點

- 1、取樣樣本之特性及年齡說明詳細。
- 2、信、效度解說清楚，易明瞭。
- 3、列出國外多筆鑑定工具，並對其適用範圍、限制及信效度...等加以描述，方便讀者參閱。

(二) 缺點

- 1、樣本之區域說明只知來自台灣東南西北四方，但並未詳加說明其比例各為多少，使讀者不知是否會有區域性差異誤差。
- 2、描述 Gilliam Autism Rating Scale 時，最後有提及其評量者堅信度、重測信度及表面效度均高。令我感到懷疑的是，「表現效度」並沒有嚴密的數理推算公式，那該怎麼定義其具有「高表面效度」？
- 3、效標效度是以 Gilliam Autism Rating Scale 為效標，但只針對中年級以上版本做檢測。若可分別拿三個版本與之對應將會更好。

972 教育測驗與評量研究讀書會紀錄（七）

◎主 題：訪談與焦點團體訪談相關期刊論文

◎時 間： 2009/04/03 ◎地 點：社 4042-1

◎導讀人員：張佩蓉 ◎紀錄人員：陳香廷

◎內容紀錄：

訪談與焦點團體訪談

一、何謂焦點團體？

焦點團體意指針對某特定問題或特定族群組成團體進行資料收集，其優點為省時間，且以小團體訪談，易激發彼此間互動的情形，可談較廣泛的議題，讓研究除了文獻的瞭解外，更可深入明確瞭解研究對象的實際情況，但焦點團體的訪談法缺點則是小團體所談論內容並不能完全代表所有的情況，且由於多人討論，內容難以控制，較難做嚴謹的比較，因此訪談的主持人扮演非常重要的角色。

二、訪談法

訪談法一般可分為三類：第一類為結構式訪談、第二類為半結構式訪談、第三類為低結構式訪談，其中需考慮的就是工具，在質性研究中主持人就是工具，而主持人接觸個案的方法有：一對一或一對多等形式，其中一對一即為個別訪談，而一對多即是所謂的焦點團體，訪談的方式敘述如下：

- (1) 結構式訪談：對於較無經驗的主持人，需使用高度結構的問題，如是否、或給 degree，讓個案回答後再問「為什麼」？
- (2) 半結構式訪談：問什麼問題需先想好，但是回答方式是屬於開放性的。個案可能會講到一個故事或繞著某個話題，但研究者本身就應將個案敘述內容加以分析
- (3) 無結構式訪談：非常有經驗的人，在和個案談論的同時就在分析，會引導個案談論有關所要研究的內容，此種情形通常是研究者臨時遇到個案，必須馬上抓住機會，但研究者必須很有經驗，才能得到所要的資料

上列三種訪談法中目前以半結構式訪談最常被使用。

三、推論式與誘導式

一般質性研究分爲兩種形式，一種是屬於推論式：由上往下來推論；而另一種形式則是屬於誘導式：由原始資料建立概念化的東西，因此需要具有抽象思考的能力，因爲個案在敘述一件事情時，通常會天馬行空，一下講東、一下講西，也可能講到概念外圍的東西，但我們要對所研究的內容有所瞭解，引導個案到我們所要談論的概念、層次、內容或項目中，焦點團體的進行就必須具備有上列的功能，才會具有層次性而所收集的資料才會豐富，進一步才能談論分析。

四、個案尋找

首先對於我們所要研究的母群體分佈必須有清楚的瞭解，由母群體的分佈中找尋我們要的個案，一個焦點團體的進行以 4-6 人最爲恰當，其中以不超過 10 人爲原則，其中對於基本資料要有初步的瞭解，如年齡、性別等，可搭配問卷同時進行。

取樣的方法希望能代表母群體，因此在選取樣本時能有的訊息就盡量收集，不要一開始就排除一些個案而造成誤差。

在舉辦焦點團體時，interview guide 需具備有下列條件：

- 1、 題目的層次性：題目的層次越多越好，其中重要的可做記號
- 2、 題目內容的豐富性：包含個案所關切的問題、研究者不瞭解或需要澄清的問題、個案的需要、別人或醫療人員能給於什麼幫助、對於朋友、住家附近、國家、同事、政府...對個案本身的影響、或對問卷之外的問題有哪些尚未瞭解...等問題皆可列出。

972 教育測驗與評量研究讀書會紀錄（八）

◎主 題：態度測驗編製與使用及其相關期刊論文

◎時 間： 2009/03/06 ◎地 點：社 4042-1

◎導讀人員：許筑鈞 ◎紀錄人員：陳香廷

◎內容紀錄：

態度形成與改變的理論

壹、態度的定義

態度的研究向來是社會心理學中重要的一環，但對態度的定義卻是各家不一，有的強調態度的行為取向，如 Doob（1967）將態度定義為：「個體內在的、具有驅力強度的反應，此反應乃針對特定刺激組型，會影響後來的外顯反應」；或如 LaPiere（1967）更直接定義態度為：「對某種特定社會情境或社會刺激的制約化的行為組型或反應傾向」。有的同時強調態度的認知與行為層次，如 Allport（1967）認為：「態度是經由經驗所組織成的心理與神經的準備狀態，對於個體在所有事物及其相關情境的反應有直接的或有力的影響」。但是更多的學者專家認為，態度同時包含有認知、情感、行為三個成份（張春興，民 77；郭生玉，民 74；Antonak & Livneh, 1988；Horne, 1985；Triandis, Adamopoulos & Brinberg, 1984），如 Lombana（1992）便主張：「態度乃是針對某一事物正向或負向的情感反應，此反應伴隨著特定的信念，並驅使個體對該事物有特定的行為反應」。Horne（1985）認為態度的情感因素指的是個人對於態度對象一種喜歡或是不喜歡的感覺，認知因素指的是個人對於態度對象的了解或是信念，例如對於殘障者的刻板化印象即是，行為因素指的是個人對於態度對象的行為或是行為傾向。本研究係針對國中生對於殘障者態度的認知、情感和行為的三個層次加以探討。

貳、態度形成與改變的理論

態度形成與改變的理論主要有下列四種：

一、訊息處理理論（information-processing theory）

此理論強調溝通說服對態度改變的重要性，而將整個態度的形成分析為：來源（source）、訊息（message）、手段（channel）、接收者（receiver）和目的（destination）等五個部份，態度的改變也分成注意（attention）、理解（comprehension）、順從（yielding）、維持（retention）和行動（action）五個

部份；這個理論的實證研究認為：在相反立場時，直接經驗所形成的態度比間接經驗所形成的態度更能抗拒說服；但在立場相同時，前者比後者更易接受說服（吳正桓，民 78）。訊息處理理論在應用方面，仍然必須考慮個體人格、能力以及情境因素等各方面的影響（Watts，1984）。

二、行為理論（behavioral theory）

行為理論秉持著行為學派的學習理論，認為態度乃是對社會刺激的反應（Horne，1985；Kiesler，Collins & Miller，1969）。態度的改變要包含幾個因素：參與溝通、了解溝通內容、接受所獲得的訊息、保留新的看法、根據新的看法來行動（Horne，1985）。Kiesler、Collins 和 Miller（1969）指出行為理論基於三個因素，其學說較受到限制：（1）其對態度的定義較其他理論狹隘；（2）其研究實驗較其他理論更具計畫性，更脫離真實生活情境；（3）其理論驗證的實驗相當不成功。

三、功能理論（functional theory）

此理論著重於態度所能提供的功能，態度的改變取決於個人的特性和個人所獲得的有關態度對象的訊息，及二者之間的關係（Triandis，1971）。態度有四種功能：

- （1）工具性的、適應或功利的功能，指個體對於能滿足其需求的事物會產生正向的態度。
- （2）自我防衛的功能，指個體在面對真實自我或真實外在世界時，態度會有自我保護的功能。
- （3）認知的功能，態度幫助個體簡化及了解這個複雜的世界，提供有意義的認知組織。
- （4）價值表達的功能，指個體所表達的態度與其價值觀和自我概念相符合時，他能從態度的表達中得到心理滿足（Horne，1985；Watts，1984）。

此理論最受抨擊的是缺乏實證研究的支持。而且這個理論認為要改變態度之前，首先要知道所要改變的態度是屬於哪一種，這在實際上是困難的。而在改變對殘障者的態度方面，有些人對殘障者的偏見是因自我防衛的需求而起，因為自我防衛的態度和事實並無一定的關連，故更難加以改變（Watts，1984）。

四、一致性理論（consistency theory）

此理論認為人們會致力於維持他們的信念、態度和行為間的一致，維持一致的驅力成爲一種動機性因素，使得不一致出現時，人們會做某些改變以重新獲得一致（Watts，1984）。在一致性理論中，最著名的是認知失調理論（cognitive dissonance theory）。認知失調理論的主要論點是：當個體同時有兩種不一致認知時，即產生認知失調。產生失調的來源有四：

- (1) 邏輯性失調，指兩種認知在邏輯上不一致
- (2) 文化習俗，指個體明白自己的行為與文化習俗不合
- (3) 特定意見 (specific opinion) 與全面意見 (general opinion) 的不一致
- (4) 過去經驗，指現在發生的事和過去經驗不一致

當不一致發生時，個體可能會：

- (1) 增加或改變認知以降低不一致或達成一致
- (2) 避免會增加不一致的資訊或情境 (Horne, 1985)。

雖然態度形成與改變的理論發展已久，但少有對殘障者態度改變的方案設計完全是基於某種理論，多數是利用各理論所提出的一些原則作為上課原則。本研究的實驗課程設計所運用的原則是：

- (1) 利用分組方式和活動方式讓學生注意、參與課程
- (2) 以淺顯、生活化的資訊讓學生了解殘障者
- (3) 增加對殘障者新的認知和正確的認知，以引發其認知失調。

本研究使用訪談方式了解對殘障者態度形成與改變的影響因素，亦是驗證理論方式之一。

參、態度的測量

態度測量的目的是要將個體行為的觀察，轉換為代表行為背後的態度之指數 (Antonak & Livneh, 1988)。Antonak & Livneh (1988) 將態度測量分成直接測量和間接測量，間接測量是指受測者並不知道自己正在接受測量的方式，間接測量的方法有行為觀察、投射技術、偽裝技術和生理測量技術等，直接測量指的是受測者知道自己正在接受態度測量的實驗，直接測量的方法有：意見調查 (opinion surveys)、訪談 (interviews)、等級評定 (rankings)、Q 方法 (Q-methodology)、社會計量 (sociometrics)、形容詞檢核表 (adjective checklists)、配對比較量表 (paired-comparison scales)、語意區分量表 (semantic differential scales)、概率評定量表 (probabilistic rating scales)、決定性的評定量表 (deterministic rating scales)。Horne (1985) 則將測量態度的方法分成：態度量表、評定等級量表、Q 技術、配對比較量表、語意區分量表、形容詞檢核量表、社會計量法、訪談、行為觀察、投射技術、回歸主流問卷、生理測量技術等。茲就在對殘障者態度測量中經常被使用到的幾種測量方法進行討論。

一、戈特曼式量表 (Guttman scale)

戈特曼式量表因為編製困難，比較少用而不普遍 (王文科, 民 82; Antonak & Livneh, 1988)，但是運用戈特曼式量表理念編製的社會距離量表 (social distance scale) 和社會容忍量表 (social tolerance scale)，卻常被用以測量對殘障

者的態度。社會距離量表是 Bogardus (1925, 引自 Dawes, 1984) 採用戈特曼式量表的觀念所編製, 其基本理念是個體對某一特定團體所持的偏見愈大, 他與這個團體所保持的社會距離愈大。後來由 Tringo (1970, 引自 Horne, 1985) 修訂為對殘障者的社會距離量表。Westervelt 和 McKinney (1980) 亦自行設計社會距離問卷 (social distance questionnaire), 測量一般兒童對於肢體障礙兒童的態度。社會容忍量表是 Yamamoto 和 Dizney (1967, 引自 Antonak & Livneh, 1988) 所編製的, 用以測量一般人對於心理疾病患者的看法。

二、語意區分量表 (semantic differential scale)

語意區分量表是將所要測量的主題置於量表上端, 其下列出一系列的量尺, 量尺兩端是對立的成對形容詞, 量尺可能被分成三點至九點, 受測者依其對主題感覺強度作答。語意區分量表的量尺可以分成三個向度: 評量 (evaluation), 如: 好--壞; 效能 (potency), 如: 強--弱; 行動 (activity), 如: 快--慢。語意區分量表的優點是易於使用、易於作答, 缺點是受測者容易因應社會需求而偽答、或作答時有集中傾向和慷慨誤差 (王文科, 民 82; Antonak & Livneh, 1988; Dawes, 1984; Horne, 1985) Hersh、Carlson 和 Lossino (1977) 曾以十九題的語意區分量表來測量一群成人對智障者和一般人的期望。另外一個研究則以語意區分量表的方式, 測量一般兒童對於肢體障礙兒童的態度 (Rapier, Adelson, Carey & Croke, 1972)。

三、李克特式量表 (Likert scale)

Dawes (1984) 將語意區分量表和李克特式量表都稱為評定量表 (rating scales), 他認為評定量表是社會和教育心理學領域中最常使用的, 尤其是有關態度的測量經常使用評定量表, 而許多研究態度改變的研究, 其所定義的態度改變往往指的是評定量表行為的改變。

李克特式量表的內容是與測量主題有關的陳述句, 受測者表達他對每一個句子陳述內容的贊同程度, 贊同程度的量尺由兩點至七點不等, 由研究者自行決定, 通常是以預計受測母群的特性來考慮, 例如, 對於年幼兒童而言, 多於兩點或三點的量表, 可能會造成混淆, 但對成人來說, 使用七點量表是可行的, 而研究者亦可收集到比較精密的資料 (Horne, 1985)。由於李克特式量表的編製較為方便, 所耗時間較少, 從事意見或態度調查研究者, 較喜歡此類量表, 所以這種量表是當前意見或態度調查中, 最常使用的一種樣式 (王文科, 民 82)。

由 Yuker 和同事所編定的「對殘障者的態度量表」(Attitude Toward Disabled Person Scale, 簡稱 ATDP), 廣為許多研究所採用 (Byrd & Rhoden, 1981; Lazar, Gensley & Russel, 1971; Leyser & Abrams, 1983; Wilson & Alcorn, 1969), 便是採用李克特式的量表編製方式, 國內的研究者在針對各類殘障者進行研究, 所編製或修訂的態度量表亦多採用李克特式量表的編製方式 (吳武典、梁能, 民 67; 陳皎眉、杜富漢, 民 73, 吳麗君, 民 76; 李真賢、邱英秋、蔡美良、梁

鈞洗、林思雄，民 77；林素美，民 77；黃瑞珍、許秀英，民 81；邱佩瑩，民 83），可見在態度量表的編製上，此量表編製方式廣為研究者所信賴。

四、社會計量法 (sociometry)

社會計量法是評量團體中，個人社會地位與團體結構的一種技術（吳武典，民 67）。在態度的測量上使用社會計量法，可以測量出態度的行為層次，因為這個測量法本身就是受測者就他們對態度對象所採取的行為或行為傾向來回答（Antonak & Livneh, 1988）。設計社會計量法的題目時，要注意題目必須讓受測者認為他所作的決定對他的現實生活會有所影響，否則只能測出態度的情感層次，而不能測出行為傾向（Macmillan & Morrison, 1984），如：要問「你希望誰的座位在你旁邊？」，而不是「你最喜歡誰？」。在班級中使用社會計量法可以看出誰是班上的明星（star）、孤獨者（isolate）、受忽視者（neglectee）、被拒絕者（rejectee），根據測量結果所畫出的社會圖（sociogram），可以看出班級的內在社會結構。

社會計量法有易於實施、計分的優點，因此在回歸主流的情況下，常被用來測量殘障學生在班級中的人際關係（吳麗君，民 76；邱佩瑩，民 83）。但是在比較同班級殘障學生與非殘障學生的社交地位，或是僅在解釋殘障學生的社交地位，將其結果歸因為殘障因素時要非常小心（Macmillan & Morrison, 1984）。

伍、訪談 (interview)

訪談可分為結構式訪談（structured interview）、無結構式訪談（unstructured interview）、半結構式訪談（semi-structured interview）三種。結構式訪談是由訪談者根據訪談計畫向受訪者詢問相同的問題，其回答只有是 -- 否、同意 -- 不同意等，對於訪談的程序嚴格要求標準化與正式化；無結構式訪談很少限制受訪者的回答，有時候鼓勵受訪者自由表達自己的觀點，但是對於收集到的資料難以比較；但在教育研究中，鑑於結構性問題無法深入了解問題的癥結所在，因而有採用半結構式訪談的趨勢，訪談者以開放性問題收集更完整的資料。訪談法的優點為：容易取得完整資料、易深入了解問題的核心、可評鑑答案的真實性、可以掌握問題的次序等。但是也有一些限制，如：訪談者的主觀因素、缺乏標準化、費時等（王文科，民 82）。

972 教育測驗與評量研究讀書會紀錄（九）

◎主 題：智力測驗編製與使用及其相關期刊論文

◎時 間： 2009/04/17 ◎地 點：社 4042-1

◎導讀人員：鄭媿穗 ◎紀錄人員：王薪惠

◎內容紀錄：

智力測驗

一、智力的涵義與結構理論

（一）、智力的涵義

intelligence 的界定：

- (1) 智力是抽象思考的能力（the capacity of abstract thinking）。例如：
比奈 (Binet, 1905)認為正確的判斷，透澈的理解，適當的推理
為智力的三個要素。
- (2) 智力是學習能力或是學業性向（scholastic aptitude）。
- (3) 智力是適應新環境的能力（adaptability）。
- (4) 智力是問題解決（problem-solving）的能力。

（二）、智力的結構理論

1、因素分析理論

（1）二因素說（two-factor theory）

Spearman（1927）發表二因素說：

4. 普通或一般因素（General factor）：G 因素，一切心智活動之主體。
5. 特殊因素（Special factor）：S 因素。指個人的特殊能力。例如：
協調能力、準確性、空間關係等。

Cattell（1963）卡泰爾：

- a、流質智力（Fluid intelligence）：個人的推理思考能力與大腦有關。不受學習經驗、文化背景等影響。
- b、晶質（結晶）智力（Crystallized intelligence）：個人的學習經驗逐漸累積而形成的能力，是流質智力與個人經驗交互作用下的產物，受生長環境與文化背景的影響。

(2) 多因素說 (Multiple-factor theory)

Thorndike (桑代克) 將智力分為三大類：抽象、具體和社會智力。

抽象智力：指瞭解及應用觀念與符號的能力。

具體智力：指學習與應用工具及機械的能力。

社會智力：指管理他人，且善於適應或處理人際關係的能力。

(3) 群因素說 (Group-factor theory)

塞斯頓 (Thurstone, 1938) 認為智力是由一群主要能力因素構成。

(a)、知覺速度。

(b)、空間關係。

(c)、文字流暢。

(d)、語文推理。

(e)、記憶能力。

(f)、歸納能力。

(g)、數字能力。

Guilford 吉爾福德 (1965,1960,1967) 認為智力是由多種因素群所組成的。Guilford 提出智力結構 (the structure of intellect) 模式： $5 \times 6 \times 6 = 180$ 智力因素

5 種內容 (Contents)：引起思維的材料，決定思維的內容

圖形的/視覺的 (Figural/Visual)

符號的 (Symbolic)

語意的 (Semantic)

行動的 (Behavioral)

聽覺的 (Auditory)

6 種思維運作 (Operation)：進行思維的心理活動，決定思維的運作

記憶 (Memory)：記憶保存、記憶收錄

認知 (Cognition)：含後設認知

分析 (Analysis)

評價/評鑑 (Evaluation)

聚斂思考 (Convergent thinking)

擴散思考 (Divergent thinking)

6 種思維結果/產物 (Product)：整理思維的結果，獲得思維的產物

單位 (Units)

類別 (Classes)
關係 (Relations)
系統 (Systems)
轉換 (Transformations)
涵義 (Implications)

Guilford 的智力結構理論，將智力的概念予以擴大，不像傳統的智力測驗偏重於 G 因素的測量，對智力的界定和測量有相當的貢獻。

2、非因素分析論

(1) 發展理論

Piaget 認為智力是指個體對其生存環境的生物適應 (Biological adaptation) 的能力。任何個體在適應環境時，必藉調適 (Accommodation) 和同化 (Assimilation) 兩種功能的交互作用來促成。所謂「同化」是將新 experience 納入既存的心理結構或 schemata (基本圖式)，而「調適」則據個體經驗來修改 schemata。

Piaget 將個體智力發展分為四個時期：

- (a)、**感覺動作期** (Sensorimotor stage) (0~2 歲)。
- (b)、**前運思期** (Preoperational stage) (2~7 歲)。
- (c)、**具體運思期** (Concrete operational stage) (7~11 歲)。
- (d)、**形式運思期** (Formal operational stage) (11~15 歲)。

* 每個時期各有其發展特徵，以階梯性的關係，連貫而漸進發展。

(2) 認知理論：

Das (達斯, 1957) & Kirby 主張對智力的瞭解應從認知的過程 (Cognitive processes) 開始。個體在處理訊息和思考問題時的認知過程可分為兩種類型：

- a. 同時性處理 (Simultaneous Processing)：個人在問題解決情境中，可以同時考慮多種刺激。並作出答案的能力。
- b. 序列性處理 (Sequential processing)：個人在處理問題時，將各種刺激作順序性的安排，再求出答案的能力。

二、智力測驗的意義與解釋

- (一)、智力測驗 (Intelligence test) 是用來測量個人潛在能力的一種測驗，近年來智力測驗逐漸由「心理能力測驗」(Test of mental ability)、**「普通能力測驗」**(Test of general ability)、**「學術性向測驗」**(Test of academic aptitude) 等所替代，使智力測驗的意義更加明確，分為**個別與團體**智力測驗兩種。

(二)、智力測驗分數的解釋：

(1)、智力年齡（或心理年齡 mental age）：早期的比奈西量表（Binet-Scale，1905）是一種個別的心理測驗，其每個 age 皆為一個分測驗，其測驗結果採用 mentalage 來表示，故屬於年齡常模。

(2)、智力商數（Ratio intelligence quotient），由德國斯敦（Stern，1912）所創，但卻由美國心理學家推孟（Terman，1916）運用在他所編製的史比智力量表(Stanford Binet Intelligence Scale)。

$$IQ = \left[\frac{\text{心理年齡 (MA)}}{\text{實足年齡 (CA)}} \right] * 100$$
$$X = 100, \quad SD = 15, 16$$

(3)、百分等級（Percentile rank）：指低於某特定分數以下，所佔百分比。

百分位數（Percentile）：指在團體中，某一等級所得的分數。

(三)、智力商數的分佈

(1)、Stanford Binet Intelligence Scale 史比量表

IQ	%	分數解釋
140 以上	1.3	非常優秀
120~139	11.3	優異
110~119	18.1	高於正常
90~109	46.5	正常或平均
80~89	14.5	低於正常
70~79	5.6	邊緣或臨界
69 以下	2.7	智力不足

Terman (1937)

(2)、魏氏兒童智力量表（Wechsler Child Intelligence Scale）

IQ	%	分數解釋
140 以上	2.3	非常優異
120~139	7.4	優異
110~119	16.5	高於正常（聰明）
90~109	49.4	正常或平均
80~89	16.2	低於正常（愚笨）
70~79	6.1	邊緣或臨界
69 以下	2.2	智力不足

三、個別智力測驗：以比西量表（Binet Simon Scale）和魏氏量表（Wechsler Scale）最為普遍。

(一)、比西量表：

- (1) 1905 年法國 Binet & Simon 編製世界最早的個別智力測驗，共 30 題，由易到難順序排列。
- (2) 1908 年美國 Goddard，將比西量表加以修訂，此為比西量表傳入美國的開端。後來 University of Minnesota 予以修訂，而 Stanford University 的 Terman 及其同仁多次修訂完成，1985 年 Thorndike 又再次修訂，成為最廣泛使用的量表。
- (3) 1985 年 Stanford Scale 個別分測驗的內部一致性信度介於 0.66~0.96 之間。效度方面，Terman 對 1960 年量表做過分析，發現在各年齡組合測驗題目與 Total 間之相關在 0.86~0.96 之間。
- (4) Waddell 瓦德魯 (1980) 分析該量表與魏氏智力測驗之間的相關在 0.52~0.91 之間。Thorndike 亦分析 1985 年量表與 1973 年量表間相關係數為 0.81，與魏氏兒童智力量表的相關係數為 0.83。由此可知 Stanford Scale 無論在 reliability and validity 均具有相當高的係數，值得使用者信賴。
- (5) 民國 65 年在我國最後一次修改，全量表共有 140 題，自 2 歲到 18 歲共有 20 個 age group。每一年齡組有正測驗 6 題，交替測驗 1 題。(郭生玉，民 83 年)

例如： 八歲兒童心理年齡的計算

年齡組	通過題數	每通過一題的心理月數	Total Score
7	6	2	12
8	6	2	12
9	4	2	8
10	3	2	6
11	1	2	2
12	0	2	0

$$IQ = (9.3 / 8) * 100 = 116 \quad \text{mental age} = 8 \text{ 歲 } 16 \text{ 個月} \\ = 9 \text{ 歲 } 4 \text{ 個月}$$

(二)、魏式量表：

- (1)、美國 Wechsler 魏克斯勒在紐約的 Bellevue hospital 工作時為了解病患之智力，於 1939 年編製一套測量成人的智力量表 (Wechsler-Bellevue Intelligence Scale)，該表是美國人自編的第一套個別智力測量。1949 年魏氏編製一套適用於兒童的智力量

表，魏為魏氏兒童力量表（Wechsler Intelligence Scale for Children）。1967年又編製一套魏氏學前智力量表（Wechsler Preschool and Primary Scale of Intelligence）。

(2)、魏氏量表在我國的修訂：

a、民國 57 年簡茂發、盧欽銘曾初步修訂魏氏成人智力量表，惟迄今未正式出版。

b、魏氏兒童智力量表修訂版（WISC-R）則由北師大特教中心主持修訂工作，民國 68 年完成，由中國行為科學社出版。內容列舉如下：

壹、語文量表

常識測驗、類同測驗、算術測驗

詞彙測驗、理解測驗、記憶廣度測驗

貳、作業量表

圖畫完成測驗、圖系編排測驗、方塊設計測驗、

物形配置測驗、符號替代測驗、迷津測驗。

(3)、Split half reliability $r = 0.93$ （語文量表）

$r = 0.86$ （作業量表）

總量表 reliability $r = 0.94$

相隔 1 ~ 1.5 月之 retest reliability $r = 0.49 \sim 0.95$ 之間

Validity 與第四次修訂的比西量表各分測驗的

同時效度介於 0.22~0.85 之間

語文量表為 0.84~0.89 之間

作業量表為 0.55~0.66 之間

總量表 0.87

由此可知魏氏兒童智力量表與比西量表之間相關甚高

四、團體智力測驗

(一) 奧迪斯智力測驗（Otis Quick-Scoring Mental Tests, 1937）

(二) 普通分類測驗

(1) 依據美國「陸軍普通分類測驗」（Army General Classification Test, AGCT）由黃堅厚、路君約共同修訂的。AGCT 是第一次世界大戰時，陸軍為測量士兵的普通能力，請心理學者 Robert 編製的團體測驗。

(2) 該測驗分為 α 和 β 兩種

α ：是屬於文字的測驗，適用識字的士兵

β ：是非文字的測驗，適用不識字的士兵

本測驗主要目的在篩選或分類士兵的智力屬於上智、中智、下智等

三個等級，以便賦予不同的任務或工作。

- (3) 黃路氏將 α (文字) 測驗予以修訂，該 test 分爲語文理解、算術推理和方塊計算三類。每類各 50 題，全部共 150 題。對象爲國一至高中三年級或成人。施測時間爲 40 分鐘。

Split half reliability $r = 0.86$ ($n = 108$)

Retest $r = 0.75$ ($n = 113$)

效度則以本 test 與學科成績之相關，男生爲 $r = 0.5$ ($n = 50$)

女生爲 $r = 0.45$ ($n = 58$) ；與中學智慧測驗之相關爲 $r = 0.54$ 與

Otis 智力測驗之相關爲 $r = 0.61$ 。

(三) 中學智慧測驗：

蕭孝嶸氏依 Spearman 理論，於民國 31 年所編製。民國 55 年由黃堅厚與路君約共同修訂，內容分爲：類推測驗、刪字測驗、算術測驗。

(四) 國民中學智力測驗：

- (1) 由程法泌、路約君和盧欽銘共同編製，內容包括：

語文推理、語文歸納、算術計算、算術推理

- (2) 民 62 年編製完成，民 70 年修訂。適用於國中各年級學生，施測約需 1 小時。

- (3) Retest reliability $r = 0.83 \sim 0.94$ (2 weeks interval)

Equivalent reliability $r = 0.77 \sim 0.84$

- (4) Validity :

與普通分數的相關介於 $r = 0.71 \sim 0.78$ 之間

與比西量表的相關爲 $r = 0.85$

(五) 羅桑二氏語文智力測驗

- (1) 羅吉和桑代克 (Lorge & Thorndike) 於 1963 年編製了一套「多層次智力測驗」(The Multi-level Edition of the Lorge Thorndike Intelligence Test)

- (2) 讓測驗語文部分於民國 72 年由黃國彥、鍾思嘉等人修訂完成，內容分爲字彙測驗、語句完成測驗、算術推理、歸類測驗、語文推理測驗。

- (3) 已建立國小三年級至大學新生常模。測驗時間爲 40 分鐘。

- (4) retest reliability $r = 0.73 \sim 0.91$

Equivalent reliability $r = 0.64 \sim 0.88$

效度：與羅桑非語文智力測驗的相關介於 $r = 0.32 \sim 0.68$

分測驗與總測驗成績相關介於 $r = 0.51 \sim 0.86$

(六) 學業性向測驗 (Scholastic Aptitude Test, SAT)

SAT (Scholastic Aptitude Test)

ACT (American College Testing Program)

} 全美大學甄選學生之工具

GRE (Graduate Record Examination)

MAT (Miller Analogies Test)

} 目前頗為廣泛運用的甄選研究生之工具

(七) 加州心理成熟測驗 (California Short Form Test of Mental Maturity)

此測驗包括七個分測驗，構成非語文、算術和語文三種。民國66年，由台大心理系路約君修訂。測驗時間約39分鐘，適用對象為國三生至成人。

Split half reliability $r = 0.52$ (相對測驗)

Split half reliability $r = 0.32$ (歸納測驗)

$r = 0.66$ (類推測驗)

$r = 0.80$ (算術測驗)

$r = 0.44$ (數字測驗)

$r = 0.54$ (語文理解)

$r = 0.68$ (延宕回憶)

Validity，採 Predictive validity 考驗，發現總標準智力分數與高二 GPA (學業平均分數) 有顯著相關，原始總分對大學聯考分數有預測效度。

972 教育測驗與評量研究讀書會紀錄（十）

◎主 題：性向測驗編製與使用及其相關期刊論文

◎時 間： 2009/04/24 ◎地 點：社 4042-1

◎導讀人員：周秀玉 ◎紀錄人員：陳香廷

◎內容紀錄：

性向測驗

一、**性向測驗 (Aptitude test)**：屬於能力測驗的一種，在測驗個體某種特殊能力的「潛在學習傾向」。性向測驗的結果可以推測個人將來學習的方向及可能，且有助於決定訓練的需要性。

二、**性向測驗**可按其內容與功能分為以下三類：

(一)、普通性向測驗（或稱普通能力測驗）：亦即通常所說的智力測驗。此種 test 的主要功能是藉以了解個人的一般能力。

(二) 綜合性向測驗，是鑑別個人多方面能力的一類測驗。此類測驗結果，多用多個分數，表示個人在多方面的能力。

(1)、區分性向測驗 (Differential Aptitude Test, DAT) ---八向度，最廣泛使用的測驗之一。

(2)、普通性向測驗組合 (General Aptitude Test Battery, GATB) ---分為 12 個分測驗，適用於職業安置與輔導。

(3)、多元性向測驗 (Multiple Aptitude Test, MAT) ---九個分測驗。

(4)、佛氏性向測驗 (Flangan Aptitude Classification Test, FACT) ---16 個分測驗。後來又修訂成為 19 個分測驗，時間需 5 ~ 10 小時，因而限制其實用性。

(5)、羅氏職業性向測驗 (Aptitude Tests for Occupations) ---6 個性向分測驗，供選讀校系之參考，亦可供工商機構遴選人員或安排工作之依據。

(三) 特殊性向測驗：即測量個人在某方面等有潛在能力的一類測驗。現有的特殊性向多用於測量個人在音樂、藝術、機械、創造等方面的特殊潛在的才能。

(1)、美術性向測驗 (Artistic aptitude)

(2)、音樂性向測驗 (Music aptitude)

(3)、文書性向測驗 (Clerical aptitude)

(4)、機械性向測驗 (Mechanical aptitude)

(四) 創造思考測驗 (Creative Thinking test): 測驗個人思想的流暢、歸納、推理和某些特殊的知覺傾向。

(1)、基魯福特(Guilford, 1960)發散性思考測驗。

(2)、拖浪斯(Torrence,1966)創造思考測驗，由於創造思考測驗的評分比較困難而且費時，故採用並不普遍。

三、性向與智力之分別：

「智力」是指個人的抽象思想、理解、推理與記憶等方面的心理能力，此能力與學習有正相關，所以智力測驗的名稱已逐漸由「學業性向測驗」所取代。

「性向測驗」僅測驗某些活動或技能的潛在能力。

四、性向測驗在編制過程中，要特別注意該測驗是否具備「內在效度」與「外在效度」。所謂「內在效度」是指組成性向測驗的各分測驗應該不具顯著的相關，但各分測驗在其所

測的性向因素，須具有可觀的「因素負荷量」(Factor Loading)，以證明該分測驗之效度，亦即該分測驗有其測量上的獨特性。而「外在效度」是指各分測驗與其外在效標，如學業成績、訓練效果、工作表現等保持高度相關，才能建立測驗的效用或實用價值。

五、性向測驗之目的：

(1) 學校輔導學生升學或就業之參考。

(2) 工商界採用人員時篩選、分類和安置的參考。

(3) 發掘個人潛能。

(4) 軍隊用來區分能力，以分配任務或工作。

972 教育測驗與評量研究讀書會紀錄（十一）

◎主 題：興趣測驗編製與使用及其相關期刊論文

◎時 間： 2009/04/24 ◎地 點：社 4042-1

◎導讀人員：林慧雅 ◎紀錄人員：陳香廷

◎內容紀錄：

興趣量表

測驗名稱	適用對象	測驗簡介(功能)	測驗內容	所需時間	題數
生涯興趣量表(C.I.I)	1.大專版：大專以上在學學生 2.增訂版：大專院校高年級及一般社會人士	『生涯興趣量表』是協助生涯探索、生涯規劃的心理測驗，依據何倫職業興趣理論為架構。	1.大專版：生涯活動、學科理念、職活動等向度評估。 2.增訂版：三大類開放式問題、另含興趣量表。	約 30 分鐘	大專版 180 題
生涯信念檢核表	大學、專科、高中、高職學生	此量表是協助個體生涯探討與規劃的好工具。以認知取向來思考省察生涯決定因素。	以完美主義、外在取向、逃避順從、自我價值、工作抱負、刻板印象等向度評估。	約 15-20 分鐘	共 72 題
生涯發展阻隔因素量表	大專學生	幫助受試者覺察到阻礙自己生涯正向發展(如升學、就業、成家等)的各項因素，面對進而克服自己的困難阻礙，獲得較好的生涯發展。	本量表包含八個阻礙生涯發展的因素，包括意志薄弱、猶豫行動、資訊探索、特質表現、方向選擇、科系選擇、學習狀況、學習困擾。受試者就四點量表中，圈選和自己最類似的選項。	需 15 分鐘	76 題
我喜歡做		本測用以探討職業興趣的方向	本測驗內容有分屬九個興趣及十二個	20-30 分鐘	182 題

的事一 職業興趣 量表			專業團體所作預測研究之結果，藉以探討受試者其學習滿意度的預測。		
修訂庫德 普通興趣 量表	適用至高三	本測驗可鑑別受試者的十種興趣，每種興趣可與常模對照，而作自比性比較。了解個人十種興趣中，那些興趣較高，那些興趣較低。	本測驗共有 504 種不同活動，每三種活動成爲一組，共 168 組。可記十種興趣範圍。	35-50 分鐘	168 題
工作價值 觀量表	大專生及大專生在職人員用	大專生及大專生在職人員了解自己的工作價值觀。	題目分成價領域(有自我成長、自我實現、尊嚴取向等三個分量表)，和工作價值領域(有社會互動、組織安全、安定、休閒健康等四個分量表)。	約 25 分鐘	共 49 題
教師信念 量表	在職教師或有志從事教職工作者	了解教師情緒和行爲困擾程度，做爲影響師生關係及教學成效之指標。及早予以輔導，使其在工作崗位上發揮所長。	本量表共 80 題，包括要求美、高自期許、遣責非難、挫折反應、不當情緒、過度焦慮、逃逸、依賴、固執、力求完美等十種分量表；另分類出『理想主義』、『逃避現實』兩項特質。	需 30 分鐘	共 80 題
高職學生 興趣測驗 (參考用)	高職學生適用	本測驗之目的在使學生實施測驗之後，對自己的興趣有所了解，亦可供輔導教師從測驗結果很快的了解學生的興	本測驗包括科學、文藝、法政、工程、商業、農漁、醫護、體育休閒及服務興趣等九項。	約 30 分鐘	270 題

		趣所在。			
國中學生 興趣診斷 測驗 (參考用)	國中生適用	本測驗之目的在使學生實施測驗之後，對自己的興趣有所了解，亦可供輔導教師從測驗結果很快的了解學生的興趣所在。	本測驗包括科學、文藝、法政、工程、商業、農漁、醫護、體育休閒及服務興趣等九項。	約 30 分鐘	270 題
青年性向 測驗 (參考用)	大專學生適用	作為對大專生從事教育或職業輔導與諮詢的重要工具。	共有七個分測驗，合訂兩冊。第一冊：語文推理、數的能力、抽象推理；第二冊：機械推理、空間關係、語文應用 I (錯字)、語文應用 II (文法)。	以不超過 1 週或 2 週為原則	有七個分測驗
青年職業 興趣測驗 (參考用)	大學及專科生	興趣測驗的結果可增進大專青年學生的自我了解，指出他們將來可能滿意的職業範圍和研究領域。	共有 22 種代表不同興趣範圍的記分卡。	40 分鐘	100 題

972 教育測驗與評量研究讀書會紀錄（十二）

◎主 題：成就測驗編製與使用及其相關期刊論文

◎時 間： 2009/05/08 ◎地 點：社 4042-1

◎導讀人員：黃怡菁 ◎紀錄人員：王薪惠

◎內容紀錄：

成就測驗

一、成就測驗特色

成就測驗是測驗學生後天習得的知識與技能，他可以考核學生學習成就的高低情形，亦可激發學生的學習動機，可以讓老師知道學生的學習效果，更可以改進教師教學方法與教材設計的依據，簡言之它可以辨別智愚、分辨才能、客觀評量學生成績、改進教學方法、增進學習動機、診斷與預測。

二、成就測驗分類

成就測驗為眾多測驗中的一種，而測驗在分類上依據施測的功能、編制的程序、科目、對象、材料、測驗的時限、測驗的反應型態、評分方式與分數的解釋之不同，有許多的分法。其中在成就測驗方面較常見的分類方式有葉重新將之分為一般綜合性的成就測驗、分科成就測驗與教師自編測驗三部分，郭生玉對於成就測驗的介紹則分為自編性的成就測驗、標準化的成就測驗、綜合成就測驗、單科成就測驗與診斷測驗等，周文欽等依其測驗的科目、編製程序與施測的目的將之分類為：綜合成就測驗與單學科成就測驗（依科目分）、標準化成就測驗與教師自編成就測驗（依編制程序分）、普通成就測驗與診斷成就測驗（依據施測的目的分）。在特殊教育上對於成就測驗的使用常依賴診斷成就測驗，用以發現特殊兒童學習上的障礙因素作為補救教學的依據，另外在教學上亦常使用教師自編性的成就測驗，以作為特殊學生起點行為的評估、擬定個別化教育計畫及評量個別化教育計畫的執行情形。

三、成就測驗功能

（一）在特殊教育之教學上

1. 讓特教老師能釐清教學目標
2. 提供特教老師回饋，且有助於特殊學生的自我瞭解
3. 提高特殊學生的學習動機
4. 讓特殊學生產生過度的學習
5. 有助於特殊學生學習的保留與遷移

（二）在特殊教育之行政上

1. 協助特教課程的評鑑與研究

2. 協助特教教師做適當的分類與安置
 3. 對特殊學校（班）作有效的品質管制
 4. 協助甄選特殊學生
- （三）在特殊教育之輔導上
1. 診斷特殊學生的性向與能力
 2. 協助輔導人員與特殊學生瞭解輔導之效
 3. 預防特殊兒童不良行為的發生

四、成就測驗編制步驟

1. 決定測驗的目的。
2. 確定所要測驗的學習結果。
3. 以具體可觀察的行為來界定所預測的學習結果。
4. 列出教學大綱。
5. 準備細目表。
6. 根據細目表編制測驗。

而根據以上的步驟，大抵尚可將測驗編制時區分兩個階段：

第一階段為 編制前的階段。

第二階段為 進行編制的階段。

（一）而編制測驗前需做編制的相關準備：

1. 確定編制的目的。
2. 確定所要測量的學習結果。
3. 以具體可見的行為界定預期的學習結果。
4. 準備雙向細目表。

（二）進行編制的相關準備：

1. 參考相關的資料
2. 決定試題的類型
3. 依據細目表進行測驗的編制
4. 依據學生的障礙類型做測驗的調整

成就測驗期刊論文

一、標題/作者(Topic/Author)

The Effect of Test Familiarization on Achievement Test Scores（成就測驗分數中對測驗題目熟悉的效果）／ N. Scott Bishop & David A. Frisbie

二、摘要(Abstract)

此研究是想要探討做部份重疊的題目使學生熟悉特定測驗內容，是否會導致學生在之後的成就測驗實施中處於優勢地位，而這種優勢可能會降低之後成就測驗分數的效度。因此，該研究同時使用部分重疊題目與無重疊題目兩種情況，控制測驗內容和年級階段，僅關注於對測驗題目的熟悉是否影響成就測驗分數。此

研究以 4 所學校作為樣本，並以 PR 值的比較進行資料分析。其主要發現為學生做重疊题目的影響並未受到支持，而且也沒有清楚的模式顯示年級或測驗會對题目的熟悉有影響。

三、理論背景(Theoretical Background)

關於練習效果的研究，在教育和心理學方面已經有一段很長的時間，但如果特別是指關於成就測驗上部分重疊题目的使用的話，此部分的文獻非常稀少，僅有 Bond(1989)和 Nevo(1976)的研究較為有關。

- (一) Bond(1989)：指出練習效果會被幾個因素影響，包括受試者特質、測驗的特性、兩次測驗的間隔時間、重複施測的次數。更重要的發現是，隨著測驗的類型由完全相同的測驗、相似或可比較的測驗、試題內容屬於相同範圍、到完全不同的測驗，其測驗的練習效果會逐漸減少。
- (二) Nevo(1976)：利用 202 個以色列大學生的樣本，以 10 個月為重測間距，比較他們在性向測驗分數的改變上，3 種練習情況的影響。研究中使用 13 個不同的性向測驗，其平均數為 10，標準差為 3。結果發現練習效果逐漸減少與測驗類型有關，依序是重做完全相同的測驗、類似的測驗、到相同範圍內的不同試題的測驗。
- (三) 研究動機：研究者認為做部分重疊题目的測驗會影響之後測驗的成績，使其失去效度，因為題目重複會使學生對特定測驗內容變得敏感，特別是第二次回答這些題目時會較有利。這種類型的題目引發的問題如下：
 1. 題目曝光增加學生下次的填答速度，使他們完成更多題目，或更快完成並有時間去檢查。
 2. 學生記得特定問題答案，會不再考慮就根據之前答案作答。
 3. 兩次測驗間，學生對特定測驗內容的記憶產生一些新想法。
 4. 這些學生甚至比常模團體更具優勢。

由於此種類型的測驗長久以來很常被使用，但引發的問題很多，甚至影響之後成就測驗的效度，因此，研究者認為有必要對此問題做研究。

四、研究問題(Research Questions)

- (一) 由於先前測驗題目曝光導致對题目的熟悉是否影響之後的成就測驗分數？
- (二) 如果這樣的影響存在，則它是否視學生的年級階段或測驗內容範圍而定？

五、資料分析(Data Analysis)

本研究對象選取 4 所學校並分派至兩個實驗情境，配對方式是基於樣本大小、測驗完成的年級、測驗時間、與前一年測驗的學校平均。實施部分重疊题目的組別，依年級來看樣本人數是從 38 到 86 人；而實施無重疊题目的組別其人數是從 77 到 95 人。此 ITBS 測驗主要包含字彙、閱讀理解、拼字、大小寫、標點符號、用法和表達、數學概念和估算、數學問題解決和資料解釋等 8 種，實施於

3~8 年級，而另外還有社會研究測驗與科學測驗僅被實施於 7~8 年級。此測驗從 3~8 年級測得內部一致性信度，其 KR20 係數大約為.82~.93。

此研究欲比較實施部分重疊題目的組別與無重疊題目的組別之學生表現，在此使用的資料分析方法為百分等級(PR)，即了解 ITBS 測驗上百分等級的改變。由於是要比較兩組的差異，所以使用此方法來描述學生的成就測驗成績，不會扭曲實際表現，能夠加以比較。

六、結果/結論(Results/Conclusion)

(一) 研究結果

此研究以實施部分重疊題目與無重疊題目兩組中，比較百分等級 50 的改變程度，來看先前測驗題目曝光導致對題目的熟悉是否影響之後的成就測驗分數，其結果如下：

- 1.結果發現百分等級的改變少於一半，表示學生做重疊題目的影響未受到支持。
- 2.沒有清楚的模式顯示年級或測驗會對題目的熟悉有影響。

(二) 結論

根據此研究問題與結果，可作成以下結論，沒有證據可以支持學生對題目熟悉的影響。並且，這些資料也無法支持這樣的影響存在，會視學生的年級階段、測驗內容範圍，或成就水平而定。

七、優點/缺點(Advantages/Disadvantages)

(一) 優點

- 1.清楚交代為何使用某種資料分析方法：此研究並非僅描述使用什麼方法進行分析，研究者詳細說明使用百分等級(PR)來作分析的好處、與此研究問題的適合性、以及為何不使用其他方法分析的原因，如年級當量(GEs)或常態曲線當量(NCEs)。
- 2.詳細介紹研究中所使用的測驗：文獻中即詳細介紹 ITBS 測驗的發展過程、基本原理、適用對象、及測驗本身的優缺點，之後在介紹研究工具時，更清楚說明包含的分測驗種類與分別的適用對象，還有測驗的信度檢驗結果。

(二) 缺點

- 1.樣本分佈不平均：4 所學校中，有一所學校的 3 年級未接受測驗，導致實施部份重疊題目的組別人數明顯少於實施無重疊題目的人數。
- 2.未清楚說明研究對象與抽樣方法：此研究僅提到研究樣本是 4 所學校的學生，僅描述樣本大小與分派情形，但並未說明樣本來自哪裡，以及用什麼方法進行抽樣，所以無法得知該樣本的特質與適用程度。
- 3.未做任何統計考驗：此研究僅進行兩組學生百分等級(PR)的比較，並無看到有做任何的統計考驗，比較後的差異大小很難確定是真的差異還是誤差範圍。
- 4.研究工具未提到效度檢驗情形：研究工具僅提到內部一致性信度檢驗的結果，沒看到任何與效度有關的描述，無法確定工具的好壞。

八、本研究對我的啓示

- (一) 讓我進一步了解學生對測驗題目熟悉其對成就測驗分數的影響：從文獻可發現，練習效果的確會對測驗分數有所影響，而且不同練習類型的測驗對之後測驗成績的影響程度會有差異，包括完全相同的測驗、類似或可比較的測驗、同一範圍內不同試題的測驗，到完全不同的測驗。但此研究卻無顯著差異，應再探討導致此結果的可能影響因素。
- (二) 應探討此研究結果沒有顯著差異的可能原因：先前的研究有支持研究假設，此研究卻沒有，可能是此研究與先前的其他相關研究有所差異，例如：
 1. 先前是指一般練習效果，這裡特別是指對測驗題目熟悉的概念。
 2. 先前是針對性向或能力測驗而不是成就測驗。
 3. 受試者是年紀較大的學生，如高中生，而不是 3-8 年級的小學生。
 4. 之前研究的再測間隔沒有像此研究 1 年那麼久。各方面都有差異，無法保證先前的研究獲得支持，此研究就也能獲得支持。
- (三) 可再探討不同測驗類型與不同題型對測驗練習效果的影響程度：相關研究中，有人使用性向測驗做研究，而此研究是針對成就測驗，我覺得可以再探討更多種類測驗的練習效果之差異。另外，成就測驗中不同題型的題目其測驗的練習效果也可能不一樣，可以進一步探討與比較。

972 教育測驗與評量研究讀書會紀錄（十三）

◎主 題：人格測驗編製與使用及其相關期刊論文

◎時 間： 2009/05/15 ◎地 點：社 4042-1

◎導讀人員：王薪惠 ◎紀錄人員：陳香廷

◎內容紀錄：

人格測驗

一、心理測驗之種類：

心理測驗，可依其所以測量的心理特質，分為認知與情意兩大類。

(一) 認知測驗—性向與智力同屬於認知測驗。

(二) 情意測驗--人格測驗、興趣測驗、態度測驗、動機測驗、人際測驗及價值量表等屬於情感運作的問題。情意測驗在測量受測者的行為方式或習慣，所以在測驗上答案並無「對」或「錯」的區分，而僅是「是」、「不是」、「有」、「沒有」的不同，以表示受測者的典型行為。

二、人格的涵義及特性：

(一) 人格(personality)一辭是由拉丁字 Persona 衍生而來，意指戲劇人員所使用的一種面具。由 Persona 很容易聯想到一個人的外在表現。

(二) 人格(personality)是一組持久而獨特的個人特性之集合，它可以因應情境之不同而變化。奧爾波特(Apport)對人格所下的定義即指各人對環境適應過程中，身心

系統動力的組織，它不僅包含個人身體和心理各自方面的特性，還包括個人在人生舞台上所扮演的角色，以及他人對其思想與行為的評價。

(三)人格的特性

(1)、人格的一致性

(2)、人格的獨特性

(3)、人格的統整性：人格是各種特質的整合體。

(4)、人格的發展性：人格不是生來就固定不變的，而是在個體的生長過程中逐漸發展而形成的。

(5)、人格的持續性：少年期的人格特質是兒童期人格特質的延伸青年期的人格特質也是少年期的人格特質的擴大。

三、人格衡鑑：

(一)「衡鑑」---意指某事物一個評價。人格衡鑑是心理學應用在實際生活上的

重要領域之一。最好的人格衡鑑技術必須符合三個條件：

- (1) 標準化 (standardization)：實施一套衡鑑工具時情境與程序之一致或統一。
- (2) 信度 (reliability)：對一套心理測驗工具反應的一致性。
- (3) 效度 (validity)：指一套測量工具能測到多少它所想測的。就實用觀點而言，最重要的一種效度就是「預測效度」—測驗分數在預測未來行為上精確程度。

「內容效度」---和測驗的個別項目或問題有關。內容效度的方法是評估每一個題目，檢驗它是否和此測驗的目標有關。

「建構效度」---則與一個測驗在測量某建構 (Construct) 時的能力有關；所謂「建構」---是指一種假設性或理論上的行為成份，例如：一個特質或動機。

(二)、人格衡鑑主要取向：

- 兩大主流 {
- (1)、自陳量表或客觀量表 (Self-report Inventory)。
 - (2)、投射技術 (Projective Technique)。
 - (3)、臨床晤談 (Clinic Interview)。
 - (4)、行為衡鑑程序 (Behavioral assessment)。

四、人格測驗的分類：

(一)、自陳量表法 (Self-report Inventory) (Self-report questionnaire)：

這種自陳問卷式的人格測驗是提供許多結構性問題或敘述語句，要受試者依照測驗作答說明來反應，以表明該敘述語句是否符合受測者個人的情況。故自陳量表又稱客觀式的人格評量法。

- (1)、明尼蘇達多相人格量表 (Minnesota Multiphase Personality Inventory, MMPI)：由於它所測量的偏重在人格多方面之特性，所以稱為「多相人格量表」，此量表自 1941 年出版以後，採用者甚多。無論在人格研究上或臨床使用上的價值均受肯定。1983 年修訂後於我國亦由路君約氏於民國 56 年修訂，柯永河亦依據其臨床經驗並參照 MMPI 編製「柯氏性格量表」分為 11 個分量表共 228 題。迄今 MMPI 已譯成 120 多種文字，全世界約有 50 個國家正在使用。堪稱極為被廣泛運用的量表。
- (2)、加利福尼亞心理量表 (California Psychological Inventory, CPI) (Cough, 1957)：本量表以測量正常人的性格特徵為主要目的，與 MMPI 為測量異常人的行為症狀不同，它可適用於十三歲以上的兒童、青少年和成人。讓量表在我國民國 67 年由楊國樞和李本華兩氏將原來十八個分量表加以修訂，稱為「青少年心理測驗」，由 480 個改為 294 個項目，建立國中及高中常模，建立國中及高中常模，以供中學輔導學生之用。爾後又建立成人常模，使本量表可適用至成人。
- (3)、明尼蘇達諮商量表 (Minnesota Counseling Inventory, MCI)：由我國路

君約氏加以修訂。該量表在了解學生的適應情形，作為教師輔導學生之參考。該量表共有 355 個項目，包含九個分量表：

1. 疑問分數（計算未答項目數）
2. 效度分數（鑑別受試者是否誠實）
3. 家庭關係分數
4. 社會關係分數
5. 情緒穩定性分數
6. 服從性分數
7. 現實適應分數
8. 情緒分數
9. 領導能力分數

測驗方式則以個別和 Group 均可，測驗時間約 30 分鐘。

Retest $r = 0.44 \sim 0.87$ (Males)

Split half reliability $r = 0.50 \sim 0.90$ (Females)

Validity：與貝爾適應量表的相關係數在 $0.53 \sim 0.98$

由此可知，本量表具有頗高的信度和效度。

(4)、基督性格量表（Guilford-Zimmerman Temperament Survey）：

該量表包括 10 個人格特質，每一特質 30 個項目，共有 300 個，每個項目有三種可能的 answer，分別為「是」、「？」、「否」。10 個人格特質即：

- a、一般活動（General activity）
- b、抑制性（restraint）
- c、支配（ascendance）
- d、社會性（sociability）
- e、情緒穩定性（emotional Stability）
- f、客觀性（objectivity）
- g、友善性（friendliness）
- h、深思性（thoughtfulness）
- i、個人關係（Personal relation）
- j、男性氣概（masculinity）

* 基督性格量表曾於民國 55 年由賴保禎陸續修訂，建立國中、高中及大專學生的 norm，另外增加虛偽量表（Lie Scale），共分為 13 個分量表，各個分量表的項目為 10 個，總共為 130 個項目。若 Lie Scale Score 得高分數，則該份測驗視為無效。

* 本測驗的各分量表之 Split half reliability $r = 0.75 \sim 0.93$
retest reliability $r = 0.53 \sim 0.80$

* 民國 81 年賴氏將該量表的項目內容重新編訂，並再建立新常模，稱為「賴氏人格測驗」。以應國中以上學校實施輔導工作之需

，目前廣泛為學校及輔導機使用。

* 自陳量表的評估：

優點：

- (1)、自陳量表的人格測驗比性向測驗或智力測驗更為客觀。
- (2)、記分較客觀、解釋有 Norm 為依據，較為簡易。

缺點：

- (1)、受試者會因個人的情緒反應，思考方向常隨時間與空間而改變，所以測驗結果亦受影響。
- (2)、受到社會期待 (Social desirable) 影響，受試者常會依照社會期待的方式回答，而影響其真實性。

(二)、投影法測驗 (Projective test)

(1)、投影法測驗工具都由一組或一系列不明確的結構材料或作業所組成，它在含義上頗為曖昧，使受測者對於測驗的目的不易了解。例如：施測者將一些墨漬或模糊不清的圖片予以受測者，並以中性的指導語詢問受測者，要求受測者根據他（她）在圖片上所見到的印象，編製一個故事。投射法測驗的分析著重在尋找受試者在反應上較為突出而一致的答案型態，然後對其人格特性加以解釋，以了解受測者的人格特性。

(2)、投射法測驗用來作為發掘受測者潛意識的工具，因此這類測驗的解釋多以心理分析的人格理論為根據。

(3)、投射法測驗的分類：

a、聯想技術 (Association techniques)：

要求受測者以測驗刺激所引發的第一個字詞或墨漬作個反應。例如：單字、單詞測驗或墨漬測驗。

b、構造技術 (Construction techniques)：

要求受測者對其情況或圖片編造一個故事。例如：主題統覺測驗、圖畫挫折測驗。

* 主題統覺測驗 (Thematic Apperception Test, TAT)：

乃是慕雷等人 (Murray & Morgan) 於 1943 年在美國哈佛大學心理診療所出版的，該 test 主要特點是施測者提供一系列圖片 (30 張)，均為黑白的，另外有一張是空白的，共計 31 張。再由受測者依自己的 feeling 編一個故事來說明圖畫中的內容。受測者的回答內容通常將其過去的經驗，目前的需求，同時亦會顯露他潛意識中的幻想，使施測者瞭解其內在心理問題。

c、完成技術 (Completion techniques)：

要求受測者將測驗刺激的來完成語句或其他作業加以完成。例如：語句完成測驗即屬此類。例如：「我的母親...」、「我的家庭...」

、「我想....」。

d、表達技術 (Expressive techniques) :

要求受測者實地去進行一項活動。例如：繪畫活動----一種藝術品往往會將作者的內心世界、心理歷程和問題投射出來。

茲將人形畫的假設略述如下 (路約君, 81)

- a、用筆很輕，表示可能是人格困擾。
- b、將圖畫在整張紙的右邊，表示取向於「將來」
將圖畫在整張紙的左邊，表示取向於「過去」
將圖畫在上面右邊，表示想壓制不愉快的過去
將圖畫在下面左邊，表示抑鬱而緬懷過去，以求逃避
- c、大眼睛或大耳朵，表示疑心病，不相信別人。
- d、繪男生而胸部特別大，表示有戀母情緒問題。
- e、細長而顯著的領帶，表示性侵犯或過份補償無能的恐懼感。
- f、特別強調衣服的「鈕扣」或凹形的嘴唇，表示依賴性強或比較幼稚的人格特性。

(三)、行為評鑑法 (Behavioral Assessment)

1960 年代末隨著行為改革學派的盛行而運用到臨床上面來。

壹、分類：

(1)、行為觀察：

自然觀察 (Naturalistic observation) : 在自然環境中實施。

設計觀察 (Contrived observation) : 在實驗室中進行。

(2)、行為問卷 (Behavior questionnaire) : 大多數屬於非正式工具，未經標準化程序，沒有建立 Norm，未檢定其信度、效度。

(3)、行為檢核 (Behavior Checklist) 與行為評定量表 (Behavior rating Scales) :

行為評定量表比較精細，觀察者不僅指出受測者有否此行為尚進一步對該行為作高低評定。評定等級一般分 1 ~ 7 級。

貳、行為評鑑的評估：

由於行為檢核表或評定量表，可在觀察時進行，亦可對個案行為作累積結果的觀察。由於使用方便，耗時不多，所以被廣泛運用。同時在內容上涵蓋面廣，對個案的行為能有全面性了解，又可將所搜集的資料加以量化，有助於對行為的解釋。故一直被行為主義心理學者視為值得採用的方法，亦為學校輔導及研究個案行為者所廣泛使用。

972 教育測驗與評量研究讀書會紀錄（十四）

◎主 題：IRT 介紹及其相關期刊論文

◎時 間： 2009/05/22 ◎地 點：社 4042-1

◎導讀人員：陳香廷 ◎紀錄人員：王薪惠

◎內容紀錄：

IRT 介紹

測驗理論主要可以分為古典測驗理論與試題反應理論兩大類，古典測驗理論數學模式簡單，計分容易而廣受採用。但是，在測量的品質上，古典測驗理論仍有非線性與樣本依賴、試題依賴的缺點。相較之下，Rasch 模式在測量上有優越的特性，能夠達到線性與客觀性的要求。

一、測驗理論簡介

測驗理論主要可以分為兩大類：古典測驗理論(classical test theory, CTT)與試題反應理論(item response theory, IRT)。

試題反應理論又稱為潛在特質理論(latent trait theory, LTT)，目前許多重要的入學考試與證照測驗，例如國內的國中基本學力測驗、美國的托福測驗、多益測驗、臺灣理財規劃顧問認證測驗等，其測驗編製與考生能力估計都是採用試題反應理論來進行。

二、古典測驗理論

法國心理學家比奈、賽門在 1905 年發展的第一個心理測驗，是心理計量學的濫觴。其理論模式的發展已為時甚久，且發展得相當規模，所採用的計算公式簡單明瞭、淺顯易懂，適用於大多數的教育與心理測驗資料，以及社會科學資料的分析，為目前測驗學界使用與流通最廣的理論依據。

古典測驗理論有下列幾項假設：

- (一) 藉由真分數模式可知，隨機誤差即是觀察分數與真實分數之間的差距。
- (二) 如果樣本數夠大，我們可以假設隨機誤差的平均數為零。
- (三) 真實分數與誤差之間應是獨立的，任何兩個複本測驗的誤差之間亦是獨立。

古典測驗理論雖然數學公式較為簡便，並且容易利用於教師自編測驗之施測與解釋，但是該理論確實存在不少的缺點，引起學者廣泛的探討。

三、心理測量的客觀性

心理測量不同於物理的測量，心理測量的測量標的物為「潛在特質」，物理的測量為「物理特性」。相較於可直接測量的物理特性，潛在特質無法直接測量，

也無可避免的隱含較大的測量誤差。即便如此，心理計量學者仍然努力改善測量的特質，希望能夠達到物理測量的客觀和等距的特性。

心理測量是藉由測量工具，將受試者的能力校準到量尺。這樣的測量才符合線性的要求，其測驗分數才達到「等距量尺」的特性。

就測量特性而言，原始分數並不能達到有意義測量的基本要求：單向度、線性、客觀性。心理測量必須滿足單向度、線性、客觀性：

- (一) 單向度是指測量必須只描述測量物的單一屬性。
- (二) 線性指的是測量單位必須均等，也就是必須達到至少等距量尺以上的要求。
- (三) 客觀性是測驗必須以達到樣本獨立及試題獨立，也就是說試題的校準不會因為受試者不同而有不同的結果；受試者能力的估計也不會因為試題的不同就產生不同的結果。

972 教育測驗與評量研究讀書會紀錄（十五）

◎主 題：質性研究相關期刊論文

◎時 間： 2009/05/29 ◎地 點：社 4042-1

◎導讀人員：陳怡安 ◎紀錄人員：陳香廷

◎內容紀錄：

質性研究

壹、質性研究歷史演變

一、傳統時期(1900-1950)

這個時期的主要思想是實證主義，強調追求研究的客觀和真實性。是對殖民地所屬的社會進行「客觀描述」的田野觀察，並企圖在看似毫無秩序的現象中，找出其規律和通則。

二、現代主義期（1950-1970）

受後實證主義影響，所謂「真實」的社會現象是研究者對世界認識的部分，是不確切的。只有不斷地從各個不同的角度對社會現象進行考察，將研究結果進行證偽，才有可能逐步接近客觀真理。

三、領域模糊期(1970-1986)

此時期受到人類學理論變化的影響，社會科學與人文科學間的界限問題受到關注，許多學者也開始注意到跨領域的理論與方法，學科與學科間的界限已日益模糊。社會科學中原本強調「統一」的時代已經過去了，也就是說，不再以科學實驗的方法來尋找「正解」，而是要分析其隱含在內的意義。

四、表述危機期(1986-1990)

學術傳統被認為是種族、性別、階級、權力等結構所作用而成的，因此逐漸受到質疑，研究者開始反思自我意識與研究立場之間的關係，也更加重視語言的詮釋問題。

五、後現代主義期(1990-)

這個時期認為沒有一個理論是放諸四海皆準的，其凸顯了後現代的不穩定、非連續及多元性。研究者更加注重不同的人群，特別是弱勢族群的心聲，也更重視行動研究。

貳、質性研究觀點取向

一、現象學

研究的目的是不在於建構獨立於主體之外的客觀世界，而是要以對話和辯證的方式，來對社會事實進行協調與取得共識。知識的來源是基於了解人們如何建構和解釋日常生活經驗。

二、符號互動論

在一個團體中，不同地位和立場的人會以不同的方式來看待同一事件。因此研究者若想要瞭解人們的行為，就必需進入研究對象身處的生活世界，並以他的想法和語言為思考中心，如此才能體會到被研究者的內心世界。

三、文化

人們對於意義的詮釋並非是分開獨立的，因此人們所生活的意義網絡是具有整體性的，而文化正是這個意義網絡。其研究結果不是採實徵主義，以某個優勢文化來檢證或修訂另一個文化，而是要來凸顯不同文化的意義與價值，進而體會與瞭解。

參、質性研究者的角色與主觀性

基於質性研究的資料蒐集方法和分析異於量化研究，所以質性研究者在角色上會關注對於人、事、物敏銳度和熱愛研究主題的要求。強烈的感性動機，導致過於主觀而成為限制；此亦為質性研究的特性和優勢，它會激起閱讀者的動機是最寶貴的關鍵。研究設計上要小心，盡量全備；來使降低偏誤產生。

在研究者就是研究工具的前提之下，研究者對於研究問題的個人意識和相關理論的本質和切入所在必須深入的、持續的思考。例如：想了解九年一貫之藝術教育的施行情況，要先釐清自己想深入了解現象的問題意識；認識九年一貫政策的來由和藝術與藝術教育的相關理論，並且這個問題使用深入訪談還是量表調查來了解，需不需要再將研究範圍縮小？

研究者角色和研究興趣的本質之深入認識議題：為什麼人需要去了解人？為什麼研究值得研究者投入？莊老師說明此作為一世界或者社會公民的使命感，是必須做的事。曉瑩的詮釋是：每個人有與生俱來的不同才華，透過互助促進人類最大幸福就是美事；找到自己最適合發揮的場域去深入，也許是一輩子的追求，人生夫復何求！？

肆、研究設計之策略

研究設計時，因不同的研究問題，我們要有各種策略來對假設的情況進行解決方案的預想。例如：欲進入重症病房，了解重症病人對死亡的看法。如果預

設研究場域（醫院）的進入遭到拒絕。可以調整的方案：去透過親友的介紹找到願意接受訪問的個案！評估自己的資源（人力資源、財力資源、時間資源）並透過指導教授的指導在文獻探討時，聚焦自己的對於相關理論的認識、批判和整理的功夫；可幫助強化獨立思考的能力。而自我檢視的一個有效的方法是透過概念框架來發現自己的思路矛盾和漏洞。例如：欲了解台灣原住民兒童藝術教育在原鄉的發展現況和自我價值的關聯。

伍、研究資料之蒐集

資料蒐集中，訪談並非純聊天是帶有目的性的；可以多種訪談技巧加入。以上主題為例：如欲進行俗民誌訪察來認識這個區域的生活文化，需要在當地居住半年到一年以上詳實紀錄。若採用焦點團體訪談瞭解藝術教育的施行與他們對理想性藝術教育的看法，可針對原鄉的藝術文化老師和藝術文化工作者為對象。使用現象學訪問法求得對當地文化的發展現況和了解，即可的透過原鄉的新移民、流動師資、鄉公所人員、長輩的互動求得他們的詮釋方式。若利用兒童訪談來了解兒童對於藝術教育的想法，利用藝術創作品提問，問題要具引導性。輔以訪問之外使用非語言的影音紀錄，可以幫助回想並蒐集到可能遺漏的關鍵資料。

陸、進行研究注意事項

無論什麼規模研究總需要人力、物力的付出，當研究結果呈現時我們期待它的正當性合法性和真實價值的展現；所以特別也注重研究者的能力、品格和研究者與研究場域的互惠與倫理，期待能在研究進場、出場之際皆能使研究者和受訪者獲得雙贏。

柒、研究結果

研究的結果可信與否，容易落入質疑。除了利用相關方法盡量說服閱讀者之外，我相信透過誠實面對自己的過程，可以產生謙卑和勇氣整個研究過程會因而富有生命力和治癒效果；所以我不打算在這個部份過慮。

捌、資料補充

一、IRB

獲得聯邦政府資助的大學或研究機構，進行以人為研究對象的研究前，都需經過IRB(institutional review board)的同意，IRB 主要工作為衡量研究所可能帶給參與者的傷害，檢查研究計畫中是否打算告知必要資訊，取得同意，以及檢查是否打算在計畫結束後將過程向參與者報告。

二、HHS

美國健康及人類服務部(美國衛生部)(Department of Health and Human Services；HHS)是維護美國公民健康，提供公共服務的聯邦政府行政部門。專門規範以人為實驗對象的指導原則。免受 HHS 指導原則限制之計畫種類：

- (1) 在教育場所施行的研究。Ex：教學法的研究。
- (2) 使用教育測驗且匿名的研究。Ex：認知、診斷、性向及成就測驗的匿名研究。
- (3) 調查或訪問的過程，但若下面三種情況任一種存在時，就需受限制：
 - a、可以指認出參與者。
 - b、參與者的回答若被公布之後，可能使某個參與者受刑事或民事的控告，或可能影響參與者的經濟或職業上的名聲。
 - c、研究內容包括參與者行為上某些「敏感的部份」，Ex：違法行為、吸毒、性行為、酗酒。

玖、三份指定閱讀教材摘要

陳向明 社會科學質的研究 ch5-7	李政賢 質性研究與計畫撰寫 ch3-4	Designing Qualitative Research
-研究設計組成部分 -研究對象抽樣 -研究者個人因素對研究影響	-建立研究設計 (研究者角色、倫理、互惠) -資料蒐集方法	-傳統質性研究不普遍 -批判質性研究： a.樣本數少無法推論 b.主觀 -質性研究辯護 a.可滿足對真相結果的關注 b.可延伸出相關的題目 -質性研究對護理教育有貢獻

拾、提問與討論

陳向明 社會科學質的研究 ch5-7

Q1：質性研究當中特別注重研究者主觀性的問題，想請問在質性研究中主觀與客觀的議題是否有所謂的平衡點或是通行方法存在？(如量化研究可採取雙盲設計)

Ans：質性研究者就以本身專業對於事物的敏感度和自己的發現為基礎出發點，當然每個不同背景的人從不同角度去看待同一件事物難免會有許多不同觀點和意見，這也是質性研究中所面臨的挑戰之一。

Q2：進入異文化場域，研究者如何調整自己？

Ans：因為有異文化的生活經驗，我覺得需要研究者自己先當觀光客；先前的疑問透過和自己背景相關的當地同儕交流，確定自己的偏見被過濾；再請教當地的重要相關人士並輔與觀察，放開自己不帶偏見地去試著和他們一起生活。

Q3：抽樣策略採關鍵個案研究，目的為推論至其他個案，該如何確保推論的邏輯性與推論的程度？是否會有盲點？如何因應樣本數量與其他變項干擾等評判？P.145

Ans：找到相關資料能為研究個案確立其在該領域的影響力，並且採取全面性多元角度的觀點，設計研究計畫。

Q4：如何分別前設與偏見，並且該如何檢視研究者自己的視角與前設？p.175

Ans：前設：為研究者在進行研究前，即在潛意識所存有的見解與假設。

偏見：一種毫無理由的觀點，一種聽不到論證，就作出的判斷，一種不知來處的莫名其妙的情緒。前設與偏見差別：理論根據的有無。與同儕或專家進行討論，並盡可能不帶個人的負面觀點進入研究領域。

Q

李政賢 質性研究與計畫撰寫 ch3-4

Q1：何謂「互惠常模」，既是常模，是否有一套標準？p.109-110

Ans：又稱互惠規範、互惠基準(Norm of Reciprocity)。

Shumaker and Brownell (1984) 認為互惠規範是一種互相受惠的意識，個體通常互給利益，以確保持續進行交換。根據 Chiu et al. (2006) 對互惠規範所下的定義，是指知識交換是互相的，且被成員認知是公平的。Wasko and Faraj (2005) 使用二個問項來衡量互惠因素，Chiu et al. (2006) 參考 Wasko and Faraj(2005) 之研究問項而發展出四個問項來衡量互惠規範，互惠的基準是個人決定是否有責任針對所獲得的東西給予回報。當個人於接受後直接回報就產生平衡交換(balanced exchange)，而也是經濟活動的基礎。

Q2：通篇提及「必要的彈性」，研究者於研究前如何訂定彈性的寬度？

Ans：取決於對研究問題的理解程度

Q3：研究中，如何脫離自己的價值判斷？P.137

Ans：暫時脫離自己的價值判斷，不容易但是詳實紀錄並適時讓自己從觀察者角度重新看見自己，我覺得戲劇的方式是很好的訓練和發現。

Q3-1：延伸問題：研究主題：原住民高中生對於高中生活的適應？採焦點訪談。

Ans：藉由此篇文章，了解須廣泛去閱讀文獻，才能找到最適合的方法。

Designing Qualitative Research

Q.4：質性研究除在護理領域有所貢獻，還有哪一個領域也十分適用質性研究？

Ans：取決於研究主題

972 教育測驗與評量研究讀書會紀錄（十六）

◎主 題：總檢討

◎時 間： 2009/06/05 ◎地 點：社 4042-1

◎紀錄人員：王薪惠

◎內容紀錄：

The Program Evaluation Standards

Utility Standards

The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.

U1 Stakeholder Identification--Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.

U2 Evaluator Credibility--The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.

U3 Information Scope and Selection--Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.

U4 Values Identification--The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

U5 Report Clarity--Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

U6 Report Timeliness and Dissemination--Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.

U7 Evaluation Impact--Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

Feasibility Standards

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

F1 Practical Procedures--The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.

F2 Political Viability--The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

F3 Cost Effectiveness--The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

Propriety Standards

The propriety standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

P1 Service Orientation--Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

P2 Formal Agreements--Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

P3 Rights of Human Subjects--Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.

P4 Human Interactions--Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.

P5 Complete and Fair Assessment--The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.

P6 Disclosure of Findings--The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results.

P7 Conflict of Interest--Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

P8 Fiscal Responsibility--The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

Accuracy Standards

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.

A1 Program Documentation--The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.

A2 Context Analysis--The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.

A3 Described Purposes and Procedures--The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

A4 Defensible Information Sources--The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.

A5 Valid Information--The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.

A6 Reliable Information--The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.

A7 Systematic Information--The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected.

A8 Analysis of Quantitative Information--Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

A9 Analysis of Qualitative Information--Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

A10 Justified Conclusions--The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.

A11 Impartial Reporting--Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

A12 Metaevaluation--The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

各項統計方法的使用目的及使用時機

1.積差相關(product-moment correlation)

使用目的：了解兩個變數之間關係密切的程度。

使用時機：適用於兩個連續變數。

例子：同一組人其智力和學業成績之間的相關。

2.獨立樣本 t 考驗(t-test, independent samples)

使用目的：兩個平均數的差異考驗。

使用時機：用在兩個互為獨立的母群的差異比較。

例子：比較男、女選手學習動機得分的差異。

3.重複量數 t 考驗(相依樣本 t 考驗)(t-test, repeated measures; t-test, correlated samples)

使用目的：兩個平均數的差異考驗。

使用時機：用在一個母群中兩次得分的差異比較(或是配對組、雙胞胎得分的差異比較)。

例子：同一組人前、後測得分的差異比較；同一組人在紅光和綠光反應時間的差異比較(或是將相同 I.Q.的受試者配對然後做某種特質的差異比較；雙胞胎在紅光和綠光反應時間的差異比較。此種相依樣本雖然不是同一個人，但因配對的關係或是雙胞胎的關係，亦可視為同一個人，此即相依樣本。)

4.獨立樣本單因子變異數分析(one-way ANOVA, independent samples)

使用目的：比較三個(含)以上的平均數的差異。

使用時機：用在三個(含)互為獨立的母群的差異比較。

例子：比較啓發式教學法、演講式教學法及欣賞式教學法在教學效果上的差異。

5.重複量數單因子變異數分析(one-way ANOVA, repeated measures)

使用目的：比較同一個群體三個(含)以上的平均數的差異。

使用時機：同一個群體，每個受試者都有三次(含)以上的得分。例子：比較某個實驗組在紅光、綠光及黃光反應時間的差異(每個受試者都必須做紅光、綠光及黃光的反應時間)。

6.獨立樣本二因子變異數分析(two-way ANOVA, independent samples)

使用目的：了解兩個自變項(或屬性變項、類別變項)對於某個依變項(觀察變項)交互作用的影響。

使用時機：當有兩個因子時，且這兩個因子互為獨立，若要了解其對某個觀察變項有何交互作用的影響時，可使用此項統計方法。

例子：想要了解 A、B 兩種藥品在使用不同的劑量(輕、重)時對於治療高血壓是否有交互作用影響。

7.混合設計二因子變異數分析(two-way ANOVA, mixed design)

使用目的：了解兩個自變項(或屬性變項、類別變項)對於某個依變項(觀察變項)交互作用的影響。

使用時機：若有兩個因子，其中一個是獨立樣本(如男、女；或有、無回饋)，另一個為重複量數(如每一個受試者均接受紅光、綠光、黃光反應時間的實驗處理)，想要了解其對某個觀察變項有何交互作用影響。

例子：了解有、無回饋是否對於不同色光(紅光、綠光、黃光)的反應時間有所影響。有回饋組和無回饋組的受試者，每人都必須對三種色光做反應。

8.重複量數二因子變異數分析(two-way ANOVA, repeated measures)

使用目的：了解兩個自變項(或屬性變項、類別變項)對於某個依變項(觀察變項)交互作用的影響。

使用時機：若有兩個因子，皆為重複量數，想要了解其對某個觀察變項有何交互作用影響。

例子：想要了解釘鞋的釘子長短和起跑架的角度對於 100 公尺短跑速度的影響，其中每個受試者均需穿長短不同的釘鞋，並使用不同角度的起跑架(一種是 60 度，另一種是 45 度)各跑一次 100 公尺。

9.變異數同質性考驗(test of homogeneity of variance)

使用目的：想要了解各個群體之間的變異數是否同質。

使用時機：在做獨立樣本的差異性比較時，都需先做此項考驗。若各個群體的變異數不同質，則所得的結果將和實際的結果有所不同。

例子：如在進行獨立樣本 t 考驗、獨立樣本單因子、二因子、多因子的變異數分析之前，都需先做此項考驗，等確定各組都同質時，才可進行差異性的比較。

10.杜凱法事後比較(Tukey method)

使用目的：確認那兩組之間有差異存在。

使用時機：當各組人數相等，而且變異數分析達到顯著水準時(當計算的 F 值 $< .05$ 時)，進一步確認何組之間有差異存在。

例子：啓發式教學法、演講式教學法及欣賞式教學法在人數相同的實驗組所產生的教學效果其差異若達顯著時，需進一步比較何組之間有差異存在。

11. 薛費法事後比較(Scheffe' method)

使用目的：確認何組之間的平均數有差異存在。

使用時機：當各組人數不等，而且變異數分析達到顯著水準時($p < .05$ 時)，進一步確認何組之間有差異存在。或是不同合併組進行差異比較達顯著水準時。

例子：啓發式教學法、編序教學法、演講式教學法、欣賞式教學法在人數不同的實驗組所產生的教學效果其差異達顯著水準時，需進一步比較何組之間有差異存在。

12. 共變數分析(analysis of covariance; ANCOVA)

使用目的：將會影響數個自變項的某一個變數抽離出來，以便比較這數個自變項對依變項的影響。

使用時機：當無法做到實驗控制時，而且發現某個變數會影響到此項實驗，就以統計控制的方法將此變數控制住(從自變項中抽離出來)，然後比較這數個自變項對依變項的影響。

例子：在三個智力不同的班級實施不同的教學法，看其學習效果是否有所不同。因無法將三個班級打散隨機編班(亦即無法做到實驗控制)，因此只好將智力的因素先加以排除，然後才比較三個班的學習效果。

13. χ^2 考驗(χ^2 test)

使用目的：可用在(1)適合度考驗(在單一向度中，就某一變項考驗其觀察次數是否與其期望次數相符合)；(2)百分比同質性考驗(考驗研究者所感興趣的 J 個群體在 I 個反應的百分比是否一樣)；(3)獨立性考驗(考驗兩個變項之間是否互為獨立)；(4)改變的顯著性考驗(用來考驗同一群受試者對同一事件前後兩次反應之間的差異情形)。

使用時機：適合用在間斷變數(如人數)的統計。

例子：(1)適合度考驗(test of goodness of fit)：考驗體研所的研究生對於四個組喜歡的人數是否有所不同。(2)百分比同質性考驗(test of homogeneity of proportions)：想了解運動選手、家長、裁判對於服用

禁藥的反應(包括贊成、沒意見、反對)，其中贊成的百分比是否相同。
(3)獨立性考驗(test of independence)：了解性別和學習動機之間是否互為獨立(亦即想驗證性別和學習動機之間是否有所關聯)。(4)改變的顯著性考驗(test of significance of change)：了解學生在學期初和學期末對體育統計喜歡的人數是否有所不同。

14.多元逐步迴歸分析(multiple stepwise regression analysis)

使用目的：主要是用在探討哪些變項較能有效預測某個效標變項。

使用時機：在不確定那些預測變項較能有效預測某個效標變項時，可將這些預測變項用多元逐步迴歸分析的方式篩選出較具預測力的變項。一般多用在探索性的研究方面。

例子：想了解學生的智力、學習動機、學習習慣、學習策略、考試技巧、學習態度、成就目標、自我觀念、父母的社經水準、父母的期望、教師的教學態度、教師的期望等變項，究竟是那些變項較能有效預測其學習表現。

15.多元同時迴歸分析(multiple simultaneous regression analysis)

使用目的：主要是了解所選出的預測變項對於某個效標變項的聯合預測力。

使用時機：當某些預測變項已被確定對某個效標變項有相關時，可將這些預測變項同時投入迴歸模式中，看其對效標變項的變異量可以解釋多少百分比。若解釋的百分比越多，則表示這些預測變項對此效標變項有較大的預測力。但在做多元迴歸分析時，必須考慮預測變項之間是否有共線的情形(若預測變項之間有中度以上的相關，就有可能產生共線。若有共線的情形，兩個變項就要選擇高的預測變項保留下來，另一個則予以剔除。)

例子：以三分球、二分球、罰球、防守籃板球、進攻籃板球、阻攻、助攻、抄截、犯規、失誤等籃球十項攻防技術來預測籃球比賽的得失分。

16.多元階層迴歸分析(multiple hierarchical regression analysis)

使用目的：主要是了解所選出的預測變項對於某個效標變項的聯合預測力。

使用時機：當某些預測變項已被確定對某個效標變項有相關時，可將這些預測變項同時投入迴歸模式中，看其對效標變項的變異量可以解釋多少百分比。但投入迴歸模式的順序並不是根據預測變項和效標變項相關的高低，而是建立在理論的基礎上。多元階層迴歸分析亦需注意共線性的診斷。

例子：以成就動機、目標接受和目標難度來預測後測分數，由於成就動機是屬於人格特質的一種，是人類比較穩定的特質，因此就第一個投入迴歸模式裡；目標接受是一個人對於所分派的目標接受的程度，是屬於一個人的態度，由於態度較會受外界的影響，不是一種穩定的特質，因此第二個投入；目標難度是由實驗者所分派給受試者的目標，是屬於實驗者操弄的變項，此為受試者自己無法控制的變項，因此最後一個投入。多元階層迴歸分析和多元同時迴歸分析在統計方法上非常類似，只是前者必須按預測變項的特質指定進入的順序。

本文作者為國立體育學院王俊明教授

討論照片舉隅



討論照片舉隅



教育專題講座資料

◎ 講者：林陳涌教授

◎ 講題：科技與人文的橋樑—生物科技教育

◎ 時間：2009.4.29

◎ 地點：社 2001

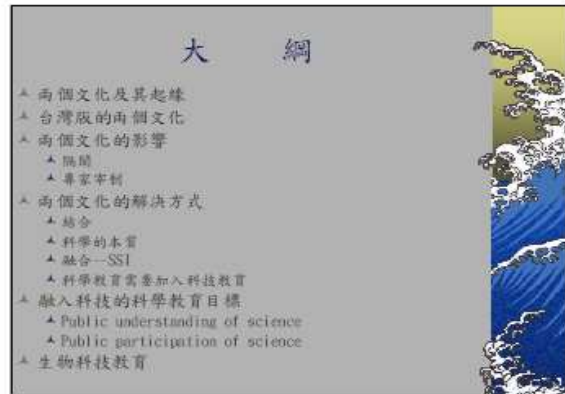
◎ 作業人員：張瑋婷、許筑鈞



科技與人文的橋樑
生物科技教育

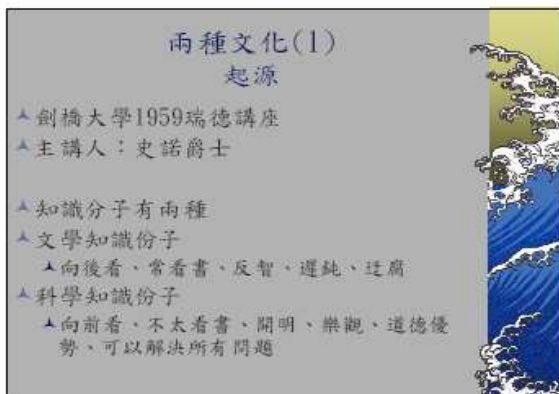


林陳涌
國立台灣師範大學
生命科學系 科學教育研究所



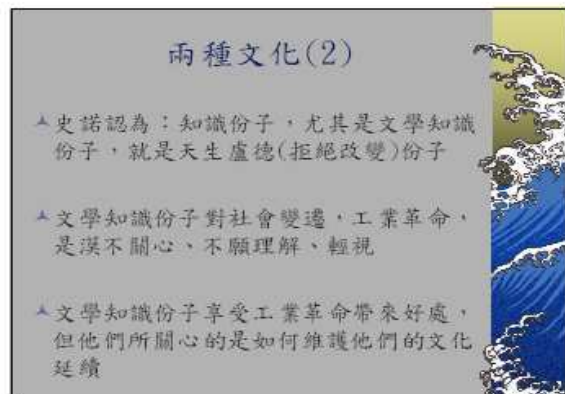
大 綱

- ▲ 兩個文化及其起緣
- ▲ 台灣版的兩個文化
- ▲ 兩個文化的影響
 - ▲ 隔閡
 - ▲ 專家宰制
- ▲ 兩個文化的解決方式
 - ▲ 融合
 - ▲ 科學的本質
 - ▲ 融合—SSI
 - ▲ 科學教育需要加入科技教育
- ▲ 融入科技的科學教育目標
 - ▲ Public understanding of science
 - ▲ Public participation of science
- ▲ 生物科技教育



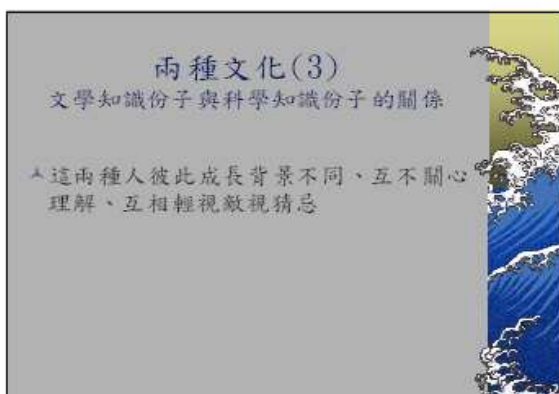
兩種文化(1)
起源

- ▲ 劍橋大學1959瑞德講座
- ▲ 主講人：史諾爵士
- ▲ 知識分子有兩種
- ▲ 文學知識份子
 - ▲ 向後看、常看書、反智、遲鈍、迂腐
- ▲ 科學知識份子
 - ▲ 向前看、不太看書、開明、樂觀、道德優勢、可以解決所有問題



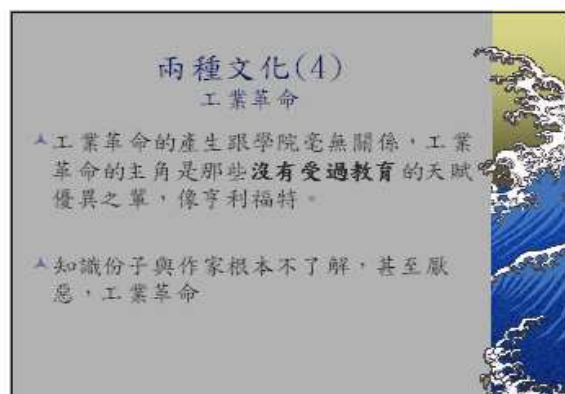
兩種文化(2)

- ▲ 史諾認為：知識份子，尤其是文學知識份子，就是天生盧德(拒絕改變)份子
- ▲ 文學知識份子對社會變遷，工業革命，是漠不關心、不願理解、輕視
- ▲ 文學知識份子享受工業革命帶來好處，但他們所關心的是如何維護他們的文化延續



兩種文化(3)
文學知識份子與科學知識份子的關係

- ▲ 這兩種人彼此成長背景不同、互不關心理解、互相輕視敵視猜忌



兩種文化(4)
工業革命

- ▲ 工業革命的產生跟學院毫無關係，工業革命的主角是那些沒有受過教育的天賦優異之輩，像亨利福特。
- ▲ 知識份子與作家根本不瞭解，甚至厭惡，工業革命

兩種文化(5)

科學革命

- ▲史諾把科學跟工業革命連結，以塑造科學的解決社會問題功能的地位
- ▲工業革命是第一次革命來自黑手草根
 - ▲不知不覺的進行
- ▲科學革命是第二次革命，會更科學化
 - ▲真正的科學應用到工業領域

兩種文化(6)

教育

- ▲史諾認為也處於兩種文化的鬥爭
- ▲更坦白的說，教育是控制在文學知識份子手中

第三種文化的興起

- ▲社會科學
 - ▲介於人文與科技之間
- ▲主流的假設不是很實證主義
- ▲有更多空間的詮釋或歷史文化分析

史諾兩種文化的小結

- ▲史諾提出兩種文化概念的企圖
 - ▲用科學文化霸權取代人文文化霸權
- ▲兩種文化的關係
 - ▲隔閡、敵視、欺壓
- ▲兩種文化的關係依然如此嗎？在英國或整個歐美

歐美對科技的反思

- ▲二戰原子彈後的反核 (1945以後...)
- ▲1970s的戰爭、貧窮、環境破壞
- ▲為甚麼科技發達沒有自動增進人類的幸福與建立更美好的世界
- ▲近年來美國歐洲還是認為社會不太了解與支持科技，大力鼓吹 public understanding of science, science communication 試圖讓社會更了解科學，以增進對科學的支持

在21世紀的台灣，我們要探討

- ▲台灣兩種文化是不是存在？
- ▲若存在，台灣兩種文化的面貌為何？
 - ▲哪兩種？誰是優勢？
- ▲若存在，科學教育應扮演甚麼角色？
- ▲若存在，整體教育應如何調整因應？

台灣版的兩種文化(1)

- ▲ 根源
 - ▲ 特殊的歷史與教育，甚至自然
 - ▲ 把握現在，有錢最重要
 - ▲ 工業經濟發展
 - ▲ 經濟垮了，台灣就垮了 發展再發展
 - ▲ 科技是經濟之母
 - ▲ 所以科技很重要，當然科學教育也很重要
- ▲ 這就是台灣版的兩種文化的基礎
 - ▲ 科技知識(份子)優於人文知識(份子)
 - ▲ 目前一流高中社會組的班級數

台灣版的兩種文化 (2)

- ▲ 工業經濟優先，民生生活第二
 - ▲ 節約能源支持工業
 - ▲ 節能減碳的對象是人民，隨手關燈、不能烤肉
 - ▲ 這些用道德包裝的環保政策，背後真正的目的是故意忽視了工業的污染，讓人民覺得環保不好是人民的責任，而不是政府與大企業的责任。
- ▲ 發展優先，環境第二
 - ▲ 蘇花高
 - ▲ 雪山隧道對礁溪發展的衝突
 - ▲ 不僅得不到好處，反而讓社區更不適合人住

台灣版的兩種文化 (3)

- ▲ 社會常用科技威權
- ▲ 大眾媒體的利用
 - ▲ 洗髮精廣告，保養品廣告

我也常用科技的威權

- ▲ 對師資培育或教科書開放的立場
 - ▲ 達爾文的演化論
- ▲ 競選社團社長的演講
 - ▲ 學科學的人是客觀公正的一路走來始終如一
 - ▲ 科學是始終如一
- ▲ 用**科學的威權**來支持人與觀念
 - ▲ 他很科學
 - ▲ 他很生物學

解決兩種文化的方法 (1)

- ▲ 教育的通識教育
- ▲ 學術界：國科會1996舉辦的「科技與人文對話」論壇
 - ▲ 科學與靈異現象
 - ▲ 文學、藝術與科學對話
 - ▲ 理性育成性、人文與科學教樣的融合
 - ▲ 宗教與科學


解決兩種文化的方法 (2)

- ▲ 全國科技會議
 - ▲ 四年開一次，有兩次提到人文
- ▲ 1991元月第四次全國科學技術會議
 - ▲ 議題五：調和人文社會與科技發展
- ▲ 2009第八次全國科學技術會議
 - ▲ 議題一：結合人文科技，提升生活品質

▲上述的解決方式的基本觀念：科技人文兩種文化根本就是不同

▲也就是科技與人文本質上是兩個不同的世界

▲所以我們只能結合跟調合



兩種文化的影響

▲「兩種文化」誰得利？

- ▲專家!專家!專家!
- ▲專家不應凌駕民意 vs. 民意不應凌駕專家
- ▲專家決定，庶民閉嘴
 - ▲請相信專家、請交給專家
- ▲傅大為：陽明大學社會與科技所
 - ▲科學太重要了，不能只讓科學家決定。

兩種文化的影響

▲以簡單的二分法化約來面對複雜的社會現象

▲科技的内容交給專家，不是一般人可以理解與參與的

▲科學是中立的，科技所帶來的弊病不是源自於科技，而是源自於科技被人的誤用，所以是人文的問題

▲科技與人文真的是兩個不同的世界嗎？

- ▲科技沒有人文社會內涵嗎？

科學與科技的區別

▲科學	▲科技
▲理論的(基本)	▲應用的(衍生)
▲上游的(先)	▲下游的(後)
▲學術性	▲實用性
▲生活無關	▲生活有關
▲價值中立	▲價值不中立？
▲沒有對錯	▲有對錯

科技與科學的關係

▲科技與科學屬於不同領域

- ▲平行
- ▲科技可用來促進科學
- ▲科技是工具
- ▲科技是應用科學*
- ▲科技是衍生
- ▲目前科學教育的主流觀念
- ▲科技與科學密不可分**
- ▲一體兩面
- ▲科學教育逐漸推廣的觀念

「科技與科學」的緊密關係

▲科學經常被視為有無窮想像空間的科技應用

- ▲科學法科技的光
- ▲科學需要經費，而其最大的支持論點是其潛在的應用性

▲科技與科學的距離越來越近

- ▲材料系 Material science and engineering, 畜產系改名 Animal science and technology, 生物系改成生命科學系或生物科技系, 台大物理系長期在崩潰邊緣
- ▲科技靠向科學? 科學靠向科技?

▲能避免嗎? 難! 因為錢

- ▲冷戰後美國兩部科學大計畫(太空、高能物理)面臨經費問題
- ▲台灣非美國軍型計畫的壓力
- ▲台灣光子源(加速器)實驗室 70億

問題

- ▲ 如何解決兩種文化的問題？
 - ▲ 重視科學社會議題 Social scientific issues
- ▲ 我們的教育，尤其科學教育，必須做什麼改變？
 - ▲ 引入科技教育，如生物科技教育

如何解決兩種文化的問題？ 錯誤的觀念與方法

- ▲ 過去將兩種文化視為完全不同不同的領域，因此需要結合、調合
- ▲ 所以產生要雙方的人來對談，
- ▲ 或是要理學院的人去修人文的課，文學院的人去修理學院的課

如何解決兩種文化的問題？ 正確的觀念與方法

- ▲ 科學社會議題 Social scientific issues (SSI) 本身就是科技與人文融合在一起，要解決SSI就是同時要考慮科技與人文層面
- ▲ 教育要引入SSI，作為培養具有融合兩種文化素養公民的教學內容

誰來執行SSI教育？

- ▲ 人文科：地理、歷史
- ▲ 社會科學：公民
 - ▲ 不是公民與道德
- ▲ 科學科目：
 - ▲ 99課綱：基礎生物(2)

科學教育的目標

- ▲ 科學相關專業從業人員 vs. 科學素養的公民
 - ▲ 兩個目標都需要補充
- ▲ 科學相關專業從業人員
 - ▲ 數理資優班：保送理學院
 - ▲ STEM, science, technology, engineering, mathematics
- ▲ 科學素養的公民
 - ▲ Public understanding of science
 - ▲ Public participation of science

科學教育的內容

- ▲ 1950—1970
 - ▲ 生活化教材
- ▲ 1980—1990
 - ▲ 美國科學課程大改革
 - ▲ 以科學概念為主
- ▲ 1990—至今
 - ▲ 目標觀念雖改，內容沒有變

科學教育的內涵

- ▲ 重新界定科學教育內涵
 - ▲ 科學教育要不要增加科技內容
- ▲ 如何呈現科學或科技
 - ▲ 科技與人文社會是一體的
- ▲ 科學教育應重視科技的人文與社會面
 - ▲ 科學本質、科學史、科學之美
 - ▲ 科技政策形成機制、科技倫理、科技法律、科技管理或科技(產品)的設計。
 - ▲ 教育不是融合科技與人文，或只是提升國民科技原理原則的知識，而更應該促進討論參與科技的人文社會面

99高中生物課綱

- ▲ 基礎生物(一)
- ▲ 基礎生物(二)(應用生物學)
- ▲ 選修生物
- ▲ 以往通常採用螺旋式課程

基礎生物(一)、(二)目標

- ▲ 一、引導學生探討各種生命現象及其原理以及生物體的構造和功能，以培養基本的生物學素養，激發探究生物學的興趣。
 - ▲ 二、引導學生認識現代生物學知識的發展，以及欣賞生物學社群研究活動的成就。
 - ▲ 三、引導學生了解人類生活與大自然之間的關係，以培養尊重生命、愛護自然與永續發展的理念。
 - ▲ 四、培養學生的科學過程技能，批判思考與解決日常生活問題的能力。
- ▲ 一、引導學生探討生命科學的原理，以及**存在生活與真實上的感觸**，以促進學生對生命科學的瞭解與興趣。
 - ▲ 二、引導學生探討**生活中科學社會、法環及治理相關議題**，以培育現代國民基本科學素養。
 - ▲ 三、引導學生了解人類生活與環境之間的關係，以培養尊重生命、愛護自然與永續發展的理念。
 - ▲ 四、培養學生的科學過程技能，批判思考與解決日常生活問題的能力。

選修生物目標

- ▲ 一、引導學生探討各種生命現象及其原理以及生物體的構造和功能，以培養基本的生物學素養，**奠定探究生物學的基礎**。
- ▲ 二、引導學生認識現代生物學知識的發展，以及欣賞生物學社群研究活動的成就。
- ▲ 三、引導學生了解人類生活與大自然之間的關係，以培養尊重生命、愛護自然與永續發展的理念。
- ▲ 四、培養學生的科學過程技能與批判思考，**以發展從事生物學研究**與解決日常生活問題的能力。

基礎生物(一)內容

- ▲ 生命的特性
 - ▲ 生命現象、細胞的構造、細胞的生理、細胞與能量
- ▲ 遺傳
 - ▲ 染色體與細胞分裂、性狀與遺傳法則、遺傳物質、基因轉殖技術及其應用
- ▲ 演化與生物多樣性
 - ▲ 生物的演化、生命樹、生物多樣性

基礎生物(二)內容

- ▲ 生物科學與農業
 - ▲ 生物品系的培養與繁殖、生物病蟲害與疫病的防治、基因改造生物
- ▲ 生物科學與食品
 - ▲ 微生物與食品生產、酶在食品上的應用、基因改造食品
- ▲ 生物科學與醫藥
 - ▲ 抗生素與疫苗、器官移植、遺傳疾病的篩檢與治療、人類基因體於醫學上的應用
- ▲ 生物科學與環境
 - ▲ 入侵外來種、環境污染物質、生質能源

教科書編輯遇到的問題

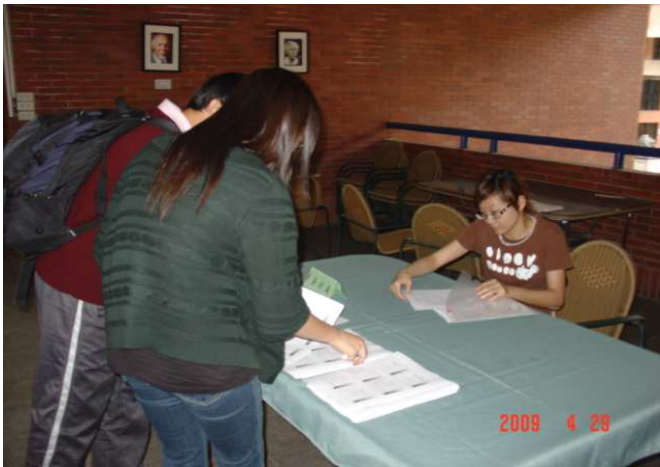
- △傳統理、生科學院的教授
- △醫、農學院教授參與

科學教師遇到的問題

- △台灣的科學教師大部分都是理學院(生命科學院)畢業
- △如何處理科技教材？
- △如何處理科技社會議題？

謝謝聆聽

教育專題講座照片



同學簽到情形



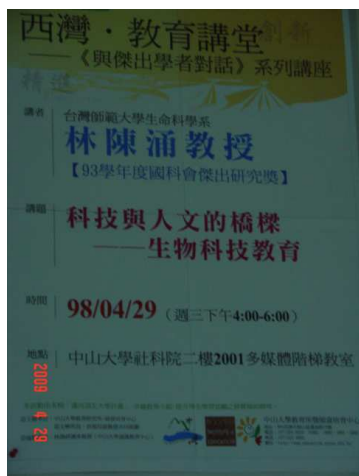
台下同學專注聽講



林陳涌教授與台下熱烈互動



林陳涌教授認真說明



活動精美海報



與教授合影