

以上四節分別說明尺度、效度、信度、與常模四個概念及其相關的方法與程序。這四個概念都直接關聯到測量工具的編製與使用。尺度是編製測量题目的依據，效度與信度是保證測量結果之可靠性與適切性的指標，而常模則是解釋測量結果的架構，四者缺一不可。因此在教育研究中使用評量工具時，必須確實了解這四個概念的意義、性質、以及應用的方法。

## 二、評量的主要方法

從研究的角度來看，評量工具的應用，旨在蒐集資料，藉以了解事物或變項的特徵。在教育研究中，最常使用的評量工具有問卷(questionnaire)、量表(scales)<sup>⑤</sup>、以及測驗(test)。換言之，研究者常採用問卷、量表、與測驗來評量事物或變項的特徵。因此，從事教育研究工作時，必須熟悉問卷、量表、與測驗的編製方法與使用要領。以下分節說明，供讀者參考。

### (一)問卷的編製與應用

「問卷」是法文questionnaire一字的中譯名稱，原意是「一種為了統計或調查用的問題表格」，若直譯的話，可譯成「問題表格」。目前大家習慣用「問卷」一詞，可說是相當達意的譯法。在應用調查法(survey research)從事教育研究時，問卷是蒐集實徵資料的主要工具，因此研究者必須熟悉問卷的編製與實施方法。

乙份問卷是由許多題目(items)組成的，每一個題大都包含「問題」(questions)與「答題」(answer)兩部分，因此在應用問卷從事調查研究時，必須先會編寫題目。以下先就「問題」的型式與「答題」的型式略加說明，然後再分別提示問卷編製及實施的程序與要領。

#### 1.問題的型式<sup>⑥</sup>

問題是一個题目的骨幹，故也稱為「題幹」(Stem)，一般而言，問卷所採用的問題，因其性質或內容不同，大致可作幾種分類，茲舉例如下：

##### (1)直接問題與間接問題：

- ①您對學校功課感到興趣嗎？（直接）
- ②您覺得學校的活動中什麼事令您愉快？（間接）
- ③您覺得學校功課對未來生活有益嗎？（間接）

##### (2)特定問題與普通問題：

- ①您喜歡爬山嗎？（特定）
- ②您喜歡從事休閒活動嗎？（普通）

##### (3)事實問題與意見問題：

- ①您家裡訂了幾份報紙？（事實）
- ②您認為一個家庭應該訂幾份報紙？（意見）
- ③您平時在家陪伴子女做功課嗎？（事實）

- ④您認為家長有必要在家陪伴子女做功課嗎？（意見）
- (4)行為問題、知識問題，與態度問題：
- ①您自本學期以來打過電動玩具嗎？（行為）
  - ②您知道本校所屬社區有幾家電動玩具店嗎？（知識）
  - ③您贊成在校園裡放置電動玩具嗎？（態度）

(5)威脅性問題與非威脅性問題：

- ①您這學期來考試作弊過嗎？（行為、威脅性）
- ②您知道成人的正常血壓是多少？（知識、威脅性）
- ③您對婚前的性行為有何看法？（態度、威脅性）

上述三個例句中，第①個問題涉及道德規範；第②個問題是現代人的基本知識，不知道就不好意思；第③個問題也涉及社會價值與期望。當答題者回答這三個問題，心中都可能有所顧忌。因此這幾個問題都是威脅性的問題，在編寫時要小心處理，設法在文字及形式上避免引起答題者的心理防衛。至於非威脅性問題係指與道德規範、社會價值、及個人隱私無關的問題，不致引起填答者的心理防衛。

(6)獨立問題與關聯問題：

- ①您喜歡戶外運動嗎？
- ②您喜歡那一種戶外運動？（關聯問題）

從上述兩個問題可以看出，第②個問題是依附在第①個問題之下，只有在第①個問題回答「喜歡」時，才有機會回答第②個問題。所以這是一個關聯性問題。

(7)疑問句與敘述句：

- ①您認為中學男女分校應予取消嗎？（疑問句）  
應予取消 不應取消 無意見
- ②中學應取消男女分校：（敘述句）  
贊成 不贊成 無意見

## 2. 答題的型式

一般常用的答題型式可以大分為兩類：一是「開放式答題」(open-ended response)，一是「封閉式答題」(closed response)。所謂開放式答題，就是不限定答題的方法，允許填答者在問題有關的範圍內，用各種方式自由回答；而封閉式答題則不僅限定答題範圍，也限制答題的方法。請看下述幾個例子：

- ①您為什麼要繼續升學？
- ②您為什麼要繼續升學？（可以複選）  
 (1)為了充實自己  
 (2)為了習得一技之長  
 (3)父母的期望  
 (4)其他
- ③您父親的職業是：\_\_\_\_\_

④您父親的職業是：\_\_\_\_\_

(1)軍  (2)公  (3)教  (4)農  (5)工  (6)商  (7)其他

由上述的例子可以看出，例①是完全無結構的、開放式答題，答題者可以從各個角度充分說明升學的理由。例③也算是開放式答題，但範圍比較明確，填答者只要填寫幾個字即可，故稱為「填充式」(fill-in)答題。至於例②與例④則都是封閉式答題，填答者只要依據自己情形，在適當的空格裡勾選即可。

一般說來，開放式答題型式採用「問答」或「填充」；至於閉鎖式的答題型式則以「是非」、「選擇」為主，包括「檢核法」(checklist)、「類別法」(categories)、「評定法」(rating)以及「等級法」(ranking)。以下分別舉例說明封閉式答題的基本型式：

(1)檢核法：通常用在初步了解有關行為與知識的事實。

您平時做什麼戶外運動？（可以複選）

- |                               |                                |                               |
|-------------------------------|--------------------------------|-------------------------------|
| <input type="checkbox"/> ①爬山  | <input type="checkbox"/> ⑤打籃球  | <input type="checkbox"/> ⑨打棒球 |
| <input type="checkbox"/> ②游泳  | <input type="checkbox"/> ⑥打網球  | <input type="checkbox"/> ⑩慢跑  |
| <input type="checkbox"/> ③騎馬  | <input type="checkbox"/> ⑦打羽毛球 | <input type="checkbox"/> ⑪其他  |
| <input type="checkbox"/> ④打排球 | <input type="checkbox"/> ⑧騎單車  |                               |

(2)類別法：通常用在行為或知識已有明確類別時；但若態度或意見可以明確分類時，亦可採用。

您目前就讀的學校是：

- ①公立學校                       ②私立學校

您贊成中學男女合班嗎？

- ①贊成                       ②無意見                       ③不贊成

(3)評定法：通常用來評量態度的程度。

您認為下列因素對一個人事業成功的重要性如何？（請圈選適當的號碼）

重要性	低				高
教育	1	2	3	4	5
智力	1	2	3	4	5
人格	1	2	3	4	5
家世	1	2	3	4	5
運氣	1	2	3	4	5

(4)等級法：通常用來了解態度的相對程度。

請將下列因素依其對一個人事業成功的重要性，依序排列等級（等級愈低，表示重要性愈低；等級愈高，表示重要性愈高）

___ 教育	___ 家世
___ 智力	___ 運氣
___ 人格	

### 3.編製問卷的程序

一份良好的問卷，必須能幫助研究者蒐集到預期蒐集的資料，藉以達到研究目的。因此編製問卷的時候，必須確實掌握研究的目的，以研究目的為起點構思問卷的題目，其程序大體如下：

(1)確定所欲探討的變項：在編製問卷之初，須將研究中有關的變項明確列出，然後根據變項的性質，逐一編寫題目。舉例言之，若想採用問卷調查國民小學的「學習環境」時，須先列出研究變項，包括物質環境與心理環境兩類變項。前者有「校園美化程度」、「學習空間大小」、與「專科教室之設置」；後者則有「班級氣氛」、「師生關係」、與「同儕關係」。確定這六個變項之後，再逐一構思題目。

(2)選定問題的形式：問題的形式很多，但每一種形式均有利弊，故如何決定問題形式，難有定則。基本的要領是：依據變項的性質、並考慮統計分析的方法，選擇一種較能引發填答者真實反應的題型。

(3)選定答題的形式：相同的問題可以採用不同的答題形式，前已述及。如何選擇答題形式，也無固定法則可資遵循。通常，可以從幾方面來考慮：其一，填答者是否容易填答？其二，要花費多少時間填答？其三，填答結果如何統計分析？所以，無論選定那一種答題形式，一定是填答者會填、也願意填，而且有適當的分析方法。

(4)撰寫問卷題目：在選定問題形式與答題形式之後，就要一題一題地撰寫題目。這是編製問卷最難的部分，也是最重要的部分。撰寫題目時，必須以變項為依據，並兼顧文字的適當運用，同時考慮題目的順序。

(5)撰寫指導語：在一份問卷裡，除了問卷的標題、編製者資料，以及題目之外，還要有一段指導語，置於題目之前。指導語一方面說明問卷的目的，一方面說明填答的方法，使填答者因了解這份問卷的意義、價值、與重要性，而能熱心據實填答，也使填答者因了解填答方法而能以適當的形式填答，增加問卷資料的效度。若問卷中有特殊題型及填答方法者，則在每個題目之前說明填答方法。

(6)預試與修正：問卷編擬完成後，一定要經過預試，以確定題目形式、內容，以及文字的適切性。預試的對象必須在研究的母群體中選取，也要略具代表性，人數不定，五至十人亦可。經過預試並加以修正後的問卷才能使用。

(7)建立效度與信度：問卷的信度大都以「重測信度」(test-retest reliability)為主。在問卷經過預試修正後，在研究的母群體中選取適當人數填答，大約二週後，同樣的

人再填答一次，比較兩次填答結果的一致程度而決定信度。若爭取時間，在第一次填答之後，改以訪問方式進行，然後比較填答與訪問結果的一致性而定信度的高低。至於效度，可請專家分析每個題目的適切性，建立「專家效度」。若能蒐集先前相關研究的調查資料作為效標，也可建立效標關聯的效度證據。

#### 4. 撰寫問卷題目的原則

在上述編製問卷程序中，撰寫研究題目可說是最難的部分，也是最重要的部分。撰寫問卷題目時，必須以研究的變項為基礎，並兼顧文字的簡潔流暢，同時也要考慮題目的順序，以利填答者填寫。以下幾個原則可供參考。

(1)要運用填答者能了解的文字、名詞、或概念；敘述要扼要、通順；避免使用學術性專門名詞。

(2)問題所傳達的信息要明確，不可語意含糊、目標曖昧；如涉及抽象概念，必須提示具體指標。

(3)避免「雙管」(double-barreled)的問題：原則上，一個題目只能問一件事、一種態度、或一個概念。如果在一個問題中包含並行的兩件事或兩個概念，就成為雙管的問題，填答者無從填答。例如，「您的教育抱負與職業抱負受誰的影響？」、「您贊成國民小學取消早自修與午休嗎？」都是雙管的問題，必須各拆成兩個問題來問。

(4)避免在問題中暗示或引導填答的方向。譬如，「爬山是一種很好的戶外活動，您喜歡爬山嗎？」這個問題的前半句就具有引導作用，容易影響填答者的反應，而造成偏差。

(5)對於敏感的、有威脅性的問題，要妥用文字。通常，使用較長的句子、或提供合理化的線索，有助於減低敏感性與威脅性。譬如，若開門見山就問「您有過自慰的行為嗎？」填答者可能有所顧忌而不敢據實回答；若改為「根據醫學報導，自慰並非病態行為，而且大部分身體健康的人都有過自慰的經驗，請問您有過自慰的行為嗎？」，不但句子加長，也有合理化的線索，可以減輕或紓解填答者的心理威脅而據實填答。

(6)在題目的順序方面，通常是容易回答的問題在先，複雜的問題在後；一般性的問題在先，特定的問題在後，呈漏斗狀；普通的問題在先，敏感的、涉及社會期望的問題在後；封閉式問題在先，開放式問題在後。同時，問題的排列要有邏輯順序。

(7)題目的長度、型式要有變化，藉以維持填答者的興趣；同時，要避免採用固定的答題型式，以免填答者受填答趨向的影響，而不仔細閱讀題目內容。

(8)問卷必須匿名填答；如需要填答者的背景資料，則置於題目的最後來問。

#### 5. 問卷的實施要領

問卷的實施有兩種方式，一是郵寄，一是當面實施。郵寄問卷可以簡省經費與人力，但無法控制填答情境，故回收率(return rate)低；當面實施雖然花費較大，但能掌握填答狀況，故可以有相當高的回收率。然而，當面實施只適合於團體施行，所以必須填答者有團體組織時才方便實施，譬如，以學校學生為對象時，若經校方同意，即可安排團體實施問卷。

郵寄問卷的最大困難是無法掌握回收率。一般的問卷調查，應有50%以上的回收率，資料的分析才有意義，若能得到70%以上的回收率，才算理想。不過，當樣本愈小時，要有較高的回收率才具有代表性。

問卷的回收率固然與問卷本身的内容及形式有關，如内容過於複雜、填答不易，就會減低填答的動機，但若在實施過程中予以適當留意，也有助於提高回收率。首先，在郵寄問卷時，一定要附上書明地址的回郵信封，讓填答者在填答完畢之後方便寄回。其次，要做追蹤催促的工作。雖然問卷係採匿名填答，但研究者可事先在每一份問卷上編號，作為識別填答者的依據。通常在問卷郵寄二、三週後，須查明未填回者姓名資料，再補寄一份問卷，請求填答。若問卷之實施要求完全保密，故不但匿名，而且問卷上不能編號時，則在郵寄問卷二、三週後，再寄一份問卷給全部填答者，並附一封「提醒信」(reminder letter)，說明如已填妥寄回者可不必理會此信，若因事忙碌而未克填回者，則請費神填答後寄回。在第一次催促後二、三週，如覺回收率仍嫌不足，則可再一次以同樣方式催促。總之，在問卷實施過程中要盡一切努力提高回收率。若最後結果回收情況不甚理想時，在研究報告中要據實說明。

## (二)量表的編製與應用

在教育研究中最常用的量表是「態度量表」。一般常用的態度量表有三種型式：一是塞斯通式量表(Thurstone scale)，二是李克特式量表(Likert scale)，三是語意區分法(Semantic differential technique)。以下分述各種型的特徵：

### 1.塞斯通式量表(Thurstone scale)

塞斯通量表係心理學家塞斯通(L. L. Thurstone)所創，採用「相等出現間隔法」(method of equal-appearing interval)來編製態度量表，其步驟如下：<sup>⑦</sup>

(1)針對特定的態度對象(人、事、物)，編寫大約200條有關的積極性及消極性的「敘述」。

(2)請專家(大約50人)判斷這200條「敘述」。判斷的方法是將這200條「敘述」依其積極性(即贊成程度)平均分為11組(堆)，每一組各給一個代碼，積極性最高者為11，最低者為1。這個代碼也就是每個評判者給予每一條「敘述」的分數。換言之，每一條「敘述」都得到50個分數(最小是1，最大是11)。

(3)分別將每一條「敘述」的得分作成次數分配，取其中數(median)(即次數居於50%位置的分數)，作為「量表值」(scale value)；並計算其「四分差」(semi-interquatile)(即次數居於75%之分數與次數居於25%之分數的差之一半，亦即 $(Q_3 - Q_1) / 2$ 作為「模糊指數」(ambiguity index)。模糊指數愈大，表示評判者意見愈不一致。

(4)將「四分差」太大之「敘述」刪除，保留20條「敘述」作為題目，即可構成一個態度量表。下圖是一個「教學態度量表」的部分題目，每一個題目均註明「量表值」(scale value)。

### 教學態度

以下各題敘述，如果您同意，就請在題號前畫✓，若您不同意，則請在題號前畫×  
量表值

- 10.2 1.教學是社會上最迫切需要的一種專業。  
10.0 2.教師是國家的領導者。  
8.7 3.教師是社會的工程師。  
5.0 4.如果教師嚴格依照教科書教學，學生會學得更多。  
2.6 5.就教學方法言，教學專業大約落後20年。

但在實施塞斯通量表時，量表值不可以顯示出來，以免受試者在表示「贊成」或「不贊成」的意見時，受到量表值的暗示。俟受試者填答之後，計算其「贊成」的題目之量表值，並以這些題目的量表值的中數作為該受試者的態度分數。

#### 2.李克特氏量表(Likert scale)

李克特式量表就是大家熟悉的評定量表(rating scale)，係由社會心理學家李克特(R. Likert)所創。®下圖是李克特育兒態度量表的一部分題目：

下列各題的敘述，請依您贊成或不贊成的程度，在適當位置作✓號。

	非常贊成	贊成	不一定	不贊成	非常不贊成
1.家裡有小孩，使夫婦更為親密	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2.教小孩子學習各種事物是件愉快的事	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3.小孩子本質上有許多惡性，所以要嚴格	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4.照顧小孩使做母親的失去許多社交機會	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5.當遭遇生活困境時，小孩是鼓勵與希望	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

李克特量表的編製方法比塞斯通量表簡單，故在應用上相當普遍。其編製程序如下：

(1)針對態度對象編寫題目。題目中要包含正向（如上圖例題中的第1、2、5題）與反向（如第2與4題）的題目，同時儘量避免太多中性或極端傾向的題目。

(2)決定評定的等級（通常採5點或7點評定量表）。

(3)選取一群受試者（大約100人左右）填答上述題目。

(4)決定計分方式後記分。通常以正向題目選答「非常不贊成」者給1分，依次往上，填答「非常贊成」者則給5分；若是反向的題目，則反之，亦即填答「非常贊成」者則給1分，填答「非常不贊成」者給5分。因此，得分愈高，表示態度愈積極，得分愈低表示態度愈消極。以上圖的題目為例，得分愈高，顯示對「育兒」的態度愈積極、愈有責任感；得分低則顯示消極的育兒態度，不喜歡、不願負責等。

(5)計算每一名受試者的總分（亦即將每一個題目所得的分數加起來），並進一步分別計算每一個题目的得分與總分之相關係數(correlation coefficient)。相關係數太低的題目即予刪除。這個過程也就是題目分析(item analysis)。

(6)經題目分析後保留的題目就構成一份態度量表。實施時，只要依據受試者填答情形換算分數，計算總分即可了解該名受試者態度傾向。必要時也可依據個別受試者的總分計算群體的平均數，作為比較群體間態度傾向的指標。

### 3.語意區分法(semantic differential technique)

語意區分法係奧斯古(C. E. Osgood)等人所發展的測量技術，旨在區分特定概念(concept)的意義，也可以用來測量態度。<sup>⑨</sup>應用語意區分法時，須以兩極化(bipolar)的形容詞(或副詞)為題目，讓受試者依自己的想法與感受，在兩極之間標定位置，藉以表達對特定概念的看法。其基本題型如下圖所示：

		教 師												
①快樂	_____	:	_____	:	×	:	_____	:	_____	:	_____	:	_____	悲哀
②不公平	_____	:	_____	:	_____	:	_____	:	_____	:	×	:	_____	公平
③和平	_____	:	×	:	_____	:	_____	:	_____	:	_____	:	_____	兇猛

在上圖中，「教師」係受評的概念；而每一對兩極化的形容詞即為題目，每一題目區分為七個間隔，代表趨向兩極的不同程度；間隔上的「×」號，就是要受試者標記的符號。計分的方法係從1分至7分，依題目的積極性增加。以上圖的題目為例，第①題靠「悲哀」的一端可得1分，靠「快樂」的一端可得7分，其中依次為2, 3, 4, 5, 6分；第②題則方向相反，靠「不公平」的一端為1分，往右遞增分數，至「公平」一端則為7分。計算每一名受試者的總分，即可知其積極為消極之間的偏向。

奧斯古等人在其原著中提出50對形容詞，經因素分析(factor analysis)後歸為三類：一是「評價」(evaluation)，如「誠實——不誠實」、「勇敢——懦弱」、「美——醜」等；二是「潛力」(Potency)，如「大——小」、「輕——重」、「深——淺」等；三是「活動」(activity)，如「主動——被動」、「快——慢」、「和平——兇猛」等。其中也有些形容詞可以同時歸在不同類別。不過，在應用語意區分法測量態度時，儘量選用評價性形容詞作為題目。至於填答方式與計分方法則與上述一般語意區分法相同。根據受試者的填答結果，可以了解個別受試者對特定概念(如教室、教學、校園、學校風氣等)的態度傾向與強度，也可以比較個別受試者對不同概念的態度差異。同時，也可以作團體間的比較，或比較同一團體對不同概念的態度。總之，語意區分法編題簡易，施行方便，又能作多種分析，可以作為蒐集教育研究資料的方法。

### (三)測驗的編製與應用

在教育研究中，常須應用測驗(test)來了解研究對象的特質與屬性。大體而言，測驗有兩類：一是心理測驗，一是教育測驗。心理測驗包括人格測驗、智力測驗等；教育測驗則以學科學習性向測驗與成就測驗為主。由於心理測驗的編製是一項艱鉅的工作，不僅程序繁瑣，而且費時費力甚多，故常非研究者獨力所能負擔；況且目前學術界及民間均有專門機構編製各類心理測驗出售應用，故通常研究者只要依據研究目的選擇適用



的測驗即可，不一定要自行編製。因此，研究者須具備選擇與使用心理測驗的能力與技巧。至於教育測驗之應用，因其內容較具特殊性，現成的測驗未必盡符研究之需要，而常須研究者自行編製，故研究者應熟悉編製的方法與程序，以下分就心理測驗的選擇與實施以及教育成就測驗的編製程序加以說明。

### 1.心理測驗的選擇與實施

就基本性質來看，一般心理測驗大都是「標準化測驗」(standardized test)。標準化測驗具有三個基本特徵：其一，它是經由標準化的程序編製而成的；其二，它有標準化的實施程序與計分方法；其三，它有解釋測驗結果的標準。就編製程序言，標準化測驗係經由客觀、嚴謹的編寫測驗題目，並經「預試」加以修正，然後定式，故具有相當程度的效度及信度。就實施程序及記分方法言，標準化測驗均明確規定實施程序與記分方法，故任何人主持測驗及評計測驗結果均無二致。就測驗結果之解釋言，標準化測驗均有「常模」(norm)作為解釋的標準。這些特徵就是心理測驗的基本性質，也是「好測驗」應具備的條件。

選擇測驗時，要從兩個問題來考慮：其一，這個測驗的性能好不好？其二，這個測驗合不合本研究之用？蓋測驗的性能係從測驗的信度、效度、實施程序、計分方法、以及常模幾方面來檢查。性能好的測驗，一定要有相當程度的信度與效度，一定有明確簡便的實施程序與計分方法，也一定有適當的常模。通常，這些資料都會在測驗手冊(manual)中詳細說明。因此，在選擇測驗時，一定要詳細閱讀測驗手冊，了解測驗的性能。同時，要看看文獻中有關該測驗的使用效果及評論，作為選擇應用的參考。

其次，從實用的觀點來看，一個測驗是否適用，必須從研究目的、研究對象，以及研究資源來考量。我們要考慮，測驗的性質與功用是否符合研究需要？測驗的適用對象是否符合研究對象？測驗所需的時間、人力、經費是否符合研究的條件？如能在這幾方面充份檢討，才能選到一個性能好的，而且適用的測驗。

至於心理測驗的實施要領，基本上要依照測驗手冊上規定的標準化程序執行。無論是個別施測抑或團體施測，都要安排一個適當的施測場所，避免外界的干擾；然後依照測驗手冊上的指導語說明作答的方法，以及注意事項，並鼓勵受試者認真作答。施測過程中，不可自作主張的補充解釋題目，或暗示作答方法，否則將影響測驗的客觀性與標準化；再者，也要確實遵照規定，控制作答時間。時間一到，務必收卷，不可提前、也不可延後。同時，測驗結果之整理與計分，也要依據測驗手冊上提示的方法來執行。唯有依據標準化的程序實施測驗，測驗結果才可靠，作為研究資料才有意義。

### 2.教育成就測驗的編製程序

教育成就測驗之編製程序，大致可以分成五個主要步驟：(1)界定測量目標、(2)發展題庫、(3)預試、(4)分析題目、以及(5)建立信度與效度。以下逐步說明：

#### (1)界定測量目標

測量什麼？這是編製成就測驗所要考慮的第一個問題，由於教育成就測驗必須以學習的材料、以及預期的學習目標為依據，因此在編製時，先要熟悉學習材料的內容，並

確定學習的目標，進而依據內容與目標製成一個「雙向細目表」(two-way specification)，如下圖，作為選題及分配題數的架構。

目標 題數	內容	第1單元	第2單元	第3單元	第4單元	第5單元
1.了解名詞與字彙 2.了解事實與信息 3.了解普遍原理 4.解釋關係 5.計算與解題 6.預測與推論 7.評鑑與判斷						

在上述雙向細目表中，橫排代表各項學習目標，直行代表學習內容（可直接填上單元名稱）。據此就可周全且均衡的分配測驗題數，填在「目標」與「單元」交叉的適當位置上，作為編選測驗題目(item)的依據。同時，也要考慮測驗的題型與難度，一併在雙向表上適當位置註明。

#### (2)發展題庫

在確定測量目標，決定測量項目之後，就要編寫測驗題目，建立題庫(item pool)。一般而言，研究用的成就測驗大都以「客觀測驗題」(objective-test item)為主，而較少採用申論的題目(essay)，除非測量的目的與寫作能力有關。常用的客觀測驗題包括：是非題，選擇題，配合題，以及簡答題（類似填充題的簡答）。編寫這類題目時，必須講求命題的技術，始能發揮題目的測量作用。以下列舉一般性的命題原則以供參考，至於各類題目的命題方法與要領，務請參閱教育測驗的專門書籍。編寫成就測驗題目的一般原則如下：<sup>⑩</sup>

- ①試題之取材宜均勻分佈，且應包括教材的重要部分。
- ②試題文字力求淺顯簡短，題意須明確，但不可遺漏解題所依據的必要條件。
- ③各個題目須彼此獨立，不可相互牽涉。
- ④試題應有不致引起爭論的確定答案。
- ⑤試題之中不可含有暗示本題或其他題目之正確答案的線索。
- ⑥試題文句須重新組織，避免直接抄襲課文或原來材料。
- ⑦試題宜注重基本原理之了解與活用，不可偏重零碎知識的記憶。

#### (3)預試

初步編寫完成的題目，必須經過預試，鑑別題目品質後，始能採用。預試的對象必須從研究的母群體中選取，也要有代表性。實施時，作答時間可以稍為加長，讓受試者均有完卷機會，但要紀錄完卷時間，以供決定測驗長度之參考。在受試者作答過程中，應容許

適當的發問（尤其是與題意有關的問題），藉以確定受試者是否完全了解題目文字意義。總之，預試時要儘量保持正式測驗時的情境與條件，但也要注意施測過程中受試者的反應，作為修訂測驗題目或施測程序之參考。

#### (4)分析題目

分析題目的目的，是要了解每一個題目的「難度」(difficulty)與「鑑別度」(discriminability)，作為取捨的依據。通常，這個步驟稱為「題目分析」或「項目分析」(item analysis)。所謂「難度」，是指題目的難易程度。如果有一個題目，全部受試者都答對了，顯然這個題目太容易；如果全部受試者都答錯了，顯然題目太難了。因此，由答對人數的比率可以推論題目的難度，難度太高或太低的題目都不是好題目。至於「鑑別度」則是指測驗題目能鑑別出受試者能力的程度。理論上，能力高的受試者比較可能做對題目，能力低的受試者比較可能做錯題目。因此，若有一個題目能力高的受試者答對了，能力低的受試者答錯了，那麼這個題目就有鑑別力；相反的，如果能力高的受試者答錯了，但能力低的受試者卻答對了，那麼這個題目就缺少鑑別力。鑑別力低的題目就不是好題目。

在分析題目難度與鑑別度時，係以預試的結果為依據。先計算每一受試者的測驗總分，然後按分數高低排列。分數在前27%者，稱為「高分組」，分數在後27%者，稱為「低分組」。然後依據下列公式分別計算每一個題目的難度：

$$\text{題目難度}(P) = \frac{\text{高分組答對人數} + \text{低分組答對人數}}{\text{高分組總人數} + \text{低分組總人數}}$$

至於題目的鑑別度，則依下列公式計算：

$$\text{題目鑑別度}(D) = \frac{\text{高分組答對人數}}{\text{高分組總人數}} - \frac{\text{低分組答對人數}}{\text{低分組總人數}}$$

通常，題目難度以.50最為適中，而鑑別度則愈大愈佳。不過，在實際選擇題目時，均定一個範圍，難度以.33至.67之間，鑑別度則要大於.30，凡逾越此範圍的題目，須予以刪除。刪除後留下的題目則按難度排列，自易而難。

預試的功用是在測試題目的適當性，故作為預試用的題目數量要多，以免在刪除修正後，所遺可用題目不足。若經預試修正後的題目確實不夠用，則須增編題目，再經預試，並作題目分析後，始可應用。若勉強選用難度及鑑別度均不適當的題目，將減低測驗結果的有效程度。

#### (5)建立信度、效度，並編寫手冊

標準化測驗必須具備相當程度的信度(reliability)及效度(validity)，故在編製成就測驗時，也要提示有關信度及效度的資料。有關信度及效度的性質，以及建立信度及效度的方法、本文第一部分已有說明，請自行參閱。

在建立信度與效度之後，必須編寫測驗手冊，作為未來正式實施測驗，以及計算分數的依據。通常，測驗手冊的內容，除了說明測驗的性質、目標、內容、編製經過、以及信度與效度資料外，要明定實施程序與計分方法。由於研究者應用成就測驗蒐集研究資料時，未必自己主持測驗之實施與計分，故這兩項一定要在手冊中說明清楚，以免影響測驗實

施與結果計分的可靠性。至於常模，並非每一項研究都需要。因為應用成就測驗所蒐集的資料，常作為團體的比較與分析，而較不需要了解個別分數的團體地位，故研究者可斟酌實際情形決定是否建立測驗常模。一旦測驗編製完成，即可藉之蒐集研究所需資料。至於實施成就測驗的要領，大體與一般心理測驗的實施相同，力求客觀、標準化。

以上各節分別說明評量的基本概念與主要方法，旨在增進有意從事教育研究者有關評量的了解，俾在選擇與應用各類評量工具時，能做最適切的決定與表現，且在自行編製評量工具時，能做最周密的考慮，藉以提昇教育研究的品質，唯以篇幅所限，本文之說明大都提綱契領，故難免不夠詳盡，如果讀者覺得本文有所不足，務請自行參閱參考書目中所列有關測驗與評量的專書。再者，本文所示觀點與資料，部分摘自拙著「教育研究：基本觀念與方法之分析」乙書，特此註明。（吳明清撰）

#### 附註：

- ①參見Stevens, S. S.(1959). *Measurement, Psychophysics, and Utility*. In C.W. Churchman & P. Ratoosh(Eds). *Measurement: Definition and Theories*. New York: Wiley, P.19
- ②這四種測量尺度自Stevens於1959年提出後，沿用迄今。參見上註PP.18-63。
- ③美國心理學會(American Psychological Association). 教育研究學會(American Educational Research Association)與全國教育測量委員會(National Council on Measurement in Education)共同於1985年修正出版「教育與心理測驗標準」(Standards for Educational and Psychological Testing)，對於效度的解釋，強調效度的「單一概念」(Unitary Concept)，故捨棄傳統有關效度的分類方式，改以「效度證據」代之。詳見：American Psychological Association(1985). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Psychological Association, P. 19。
- ④測量誤差係指測量時實得分數（或稱觀察值）與真實分數（或稱真值）的差距。通常以實得分數的變異量中所占比率來說明。測量誤差愈小，測量的信度就愈高；反之則反。參見：郭生玉（民78），*心理與教育測驗*（4版）。台北：精華書局，68頁。
- ⑤英文的scale一字具有「尺度」與「量表」兩種意義。尺度是測量的準則，而量表則指由尺度構成的測量工具。
- ⑥有關問卷題目的分類、型式、與設計，請參閱S. Sudman & N.M. Bradborn(1982). *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass.
- ⑦參見L. L. Thurstone & E. J. Chave(1929). *The Measurement of Attitudes*. Chicago:University of Chicago Press.
- ⑧參見R. Likert(1932), A Technique for Measurement of Attitude. *Archives of Psychology*, 22,140,PP.1-55.

- ⑨參見C. E. Osgood, et al.(1957). The Measurement of Meaning. Urbana, IL:University of Illinois Press.
- ⑩引自簡茂發與郭生玉撰（民67），編製測驗題目的技術。刊於楊國樞等編著，社會及行為科學研究法。台北：東華書局，頁442- 444。