測驗統計年刊，民82.1期33-42頁

標準化測驗的編製發展程序

歐　滄　和

教育與心理測驗的出版品在整個教育出版事業中所佔的份量是微不足道，縱使像美國那麼大的國家，純粹出版測驗的機構也是屈指可數；因此本文將介紹標準化測驗的發展過程，以增加讀者對它的了解與重視。

本文所謂的「測驗發展」是專指由學術機構或商業出版機構編製且發行的客觀計分的標準化測驗的產銷過程，但並不包括大學入學考試或專業證照及檢定考試。

要發展測驗先要組成一個團隊，這團隊通常由具有博士學位的測驗專家主持，裏面的成員可能有統計學、心理計量學、研究方法學、學科專家或特殊領域之專家（視測驗類型而定），除此之外，也需要有文書和編輯方面的高手，才能勝任測驗發展過程中繁多的文書工作。雖然近年來，電腦的廣泛運用，已減輕了很多工作負荷，但測驗發展仍然是一種技術密集、勞力密集的工作。

一、提出測驗發展計畫

發展測驗的原初構想可能來自於大學教授作研究上的需要，也可能來自研究生或一般教師寫論文時收集資料的工具，更可能是來自商業機構或社會福利機構在人員甄選上或臨床診斷上的需求。但不論原初的構想如何，如要獲得出版社的支持，編製者就得把他籠統的概念化成具體的計畫。出版社所要求的計畫，通常要包括下列資料，以便評估計畫的可行性：

1. 說明編製此測驗的目的及理由
2. 對此測驗的描述
   - a. 所測量的特質
   - b. 受測對象的年齡或年級範圍
   - c. 測驗結構成分測驗內容
   - d. 施測時間
   - e. 施測方式
   - f. 計時或計分的方式
3. 試題
   - a. 試題來源或編擬方式
   - b. 預試項數與保留題數
4. 基本市場
   - a. 可能的使用者
   - b. 施測和解釋此測驗者應具備之資格
   - c. 使用者之類型
5. 競爭對象

a. 類似此測驗的競爭性產品的數量及價格
b. 對此新測驗的需求
c. 與其他競爭性產品相比較下之優點

6. 研究摘要
a. 先前嘗試性研究
b. 特殊性研究
c. 計畫中的附加研究

7. 評論
a. 專家對此計畫的評論
b. 其他出版商對此計畫的意見

8. 編製者的學、經歷證明

雖然編製者提出計畫，但最後所出版的測驗卻常與原來計畫的不盡相同，這是因爲出版商會提供其他的主意及有關市場的資料。出版商在決定是否發展某一測驗時，是非常愼重的，因爲它攸關出版商的專業聲譽及財務盈虧。

二、評估測驗發展計畫

出版商接到編製者所提的計畫後，則依下列項目加以評估：

1. 產品的完整性
a. 理論基礎之完備
b. 容易使用程度
c. 原創性
d. 編製者之聲譽
e. 效果的保證
f. 產品的壽命

2. 行銷市場
a. 市場的大小
b. 市場的需求量
c. 市場競爭程度
d. 推銷過程所需之消耗品及服務
e. 推出之價格與逐年遞降比例
f. 維繫現有客戶和發新客戶的可能性
g. 對現有產品銷售的影響
h. 訂貨的型式（材料的組合、最低數量）

3. 本出版社研究發展能力
a. 研人員的知識
b. 研發人員的經驗
c. 所需的外界諮詢服務
d. 此計畫的重要性
e. 發展此測驗所需時間

f. 所需取用的外界資源
g.編製者的協調與配合
h.管理上的複雜性
4.本出版社的生產能力
a. 工作人員的知識與經驗
b.產品項目的複雜性
c. 需要裝配的程度
d.現有的銷售能力
e 工作需要外包的程度

三、簽合約與編訂工作進度表

若計畫被拒絕了，編製者可以另尋其他出版社再嘗試看看；若被接受了，則進一步要簽合約明確規定編製者與出版商的責任。這種合約通常包括了版稅、發展成本的分擔，給編製者的預付款，日後修訂版本時的權利義務等等；有些出版商更進一步要求定出明確的工作進度，以確定測驗能如期發行。至於經費與控制則大多由出版商負責，以求不超出預估的成本。

四、內容的編擬

㈠編寫成就測驗的試題

成就測驗是用來評量學生在某一課程領域的學習結果。因此，若同一課程在不同的學區中有不同的教科書，那在編寫前就應收集齊這些教材。除此之外，還要擬定此測驗內容的藍圖，即所謂的「雙向細目表」或「規格明細表」。所謂雙向細目表是由兩個向度組成的表格，一個向度代表此一課程中各種不同的內容領域，另一個向度則代表認知歷程的過程目標，例如：再認、回憶、定義、應用、分析、綜合、評鑑等認知活動。

在編寫成就測驗的雙向細目表時，編製者常會參考 Bloom 的教育目標分類典，尤其是第一册認知領域的部分； Bloom 把認知領域分成六大領域，即知識、理解、應用、分析、綜合、評鑑。 Bloom 的貢獻在於他促使教師及測驗編製者注意到較高級的心理歷程在教育與測量成果上的重要性，因爲過去的教學與評量實在太偏重於對事實與訊息的回憶與再認了。

有了雙向細目表即可開始編寫試題了。編寫試題是項很耗費心力的工作，並且該在有經驗的試題編寫者的督導下進行。此外，在很多心理與教育測驗的教科書中也都列舉了編寫各類試題應遵守的原則，可以作爲參考。不管這些原則有多少，編寫試題可算是種藝術工作，它需要有良好的寫作技巧、創造力、及有關寫出良好試題應循規則的知識。不管編題技巧如何，這些試題在實際進行預試之前，都要經過他人多次的檢查和修改。

㈡編寫能力測驗的試題

不像成就測驗那樣，性向或能力測驗並不是依據某一特定課程來編擬試題，它是依據測驗編製者對某一心理特質的理論觀點來編寫的。舉個例說，若一心理學家想編一智力測驗，那試題的型式和測驗的格式將會隨著他是採取單因素理論（即G因子），或是多因素理論而有所不同。

不管編製者的理論取向如何，在預試之後，在正式標準化之前，編製者對此能力測驗的概念仍可能

變更或是更明確化，在這過程中，因素分析法或是各種相關或迴歸分析法就是常用來建構和界定能力測驗的工具，當然，標準的試題分析程序也會用來確保各個試題的品質。

編擬能力測驗的試題時，可作爲參考的不是課程內容的雙向細目表，而通常是Guildford（ 1967 ）所提出來的智慧結構模式（ SIM ）。此模式像三個向度的立方體，它是由四種內容、五種運作、及六種產品相結合而成的一二〇個細格的模式。每一個細格可以編擬出一種類型的題目。然而Guildford的模式其理論意義大於其實用意義，一二〇個細格中至今仍有些細格未能編擬出試題來。

另外有人對常用的能力測驗的題目類型作調查，發現在測驗內容上大致可分成三大類，即語文（字、句子）、符號（字母、數字）和圖形（幾何圖形、抽象），而在思考運作上，類推則是最常用的形式，尤其在測量一般能力時，更常用語文和圖形的類推，以數目或字母作類推雖然較少見，但仍偶爾出現。讀者對各種類型的題目若有興趣，可以參看 Thorndike （ 1982 ）出的 Applied Psychometrics 一書，它提供了各種題型的範例。

能力測驗的試題編得好壞並不像成就測驗那樣有一堆的編題原則可加以判斷，它完全由所測量的特質、所用的統計方法及編製者的判斷和創意來決定。

五、題數的決定與試題的審查、校閱

常有人問到在預試時應準備多少題目？然而這問題永遠得不到一個直接的答案，在此，我們應考慮的因素是⑴這測驗所涵蓋的年齡或年級，⑵測驗的版本數或測驗的長度，⑶編寫試題者的功力，⑷試題類型新穎的程度，⑸發展測驗的成本。通常測驗適用對象的年齡範圍愈大，所需準備的題目就愈多；而同時要編兩、三個版本所需的題數自然多於只編一種版本的。編寫者的知識、經驗愈豐富，所編寫的試題品質愈佳，自然不必像無經驗者準備那麼多題目以備淘汰之用。至於新穎的題型，因題目品質難以事先預估，更需要多準備一些，以預防淘汰後題數不足。編製測驗的成本是出版商最關心的，若所編試題需要製作昂貴的圖表或作美工設計時，顯然這種題目不能隨意增加。

一般說來，預試的題數要比正式測驗的題數多出 25 %到 50 %，而上列五項因素才是決定要有多少題目的主要考慮因素。

題目在預試之前要經過編輯和內容審查。在成就測驗的例子中，內容審查是要找出題目內容不正確或語意含糊的地方；而能力測驗的審查則是要檢閱每個誘答項是否具有適度的似眞性。至於編輯校閱則是要徹底檢查在打印過程中是否有錯、漏字。如果內容和編輯方面的審查作得好，將可事先找出品質不佳的試題，將它予以放棄或修改，如此，將可降低很多預試成本。

另外，有些出版社爲了避免測驗偏差（ testbias ），也邀請不同民族或少數團體的成員來共同審查，以確保測驗內容的公平性。

六、預　　試

在試題的編寫、審查、編輯校閱都完成後，即可進行預試，在預試之前通常要考慮下列幾個因素：

㈠預試可動用的經費

出版商所能提供的經費究竟有限，編製者要預估每個受試者的平均施測成本。個別測驗、多元測驗組合、及一節課即可完成的團體測驗，它們的平均測施成本差異頗大，施測前不能不仔細籌畫。

㈡受試者的年齡或年級範圍

預試時，所取樣本的年齡或年級範圍不但要涵蓋原計畫要使用的年齡或年級範圍，還應向上、向下延伸，以便取得充分的資料來決定各試題的特性。在團體測驗中，試題可能出現在不同水準的測驗中，這樣才可以求出不同層次的學生在同一題目上的表現。在適性測驗（ tailor ed testing ）的編製中，每一個試題都要所有的年齡或年級水準的學生作才可以。

㈢預試樣本的大小

爲了節省經費，樣本要愈少愈好；但爲了在試題分析時，求得試題統計量的穩定的估計值，樣本卻要愈多愈好。大部分的學者認爲在求古典的難易度與鑑別力指數時，樣本至少要在 200 到 500 人之間才能求得穩定的估計值。當然除了大小之外，也要考慮到樣本的代表性，這也就是說樣本的特徵要符合未來使用對象母群的特徵。如果要進行試題偏差（ item bias ）之研究，那麼對於要研究的少數民族或特殊群體的取樣要多取一些，以便符合研究上對樣本大小的要求。

㈣測試方式

在施測時要注意下列一些細節：⑴若因配合學校行政措施只能在有限的時間施測，那就要注意試題題數的配合，最好能讓 90 %以上的學生作完所有的題目；若題數太多，每份試卷後面未作的試題太多，將無法求得正確的試題統計量。⑵每次測驗時間約在一小時或一個半小時，要依學生年齡而定。⑶如果是成就測驗要注意施測日期的安排，因爲在上學期或下學期，在學期初或學期末都可能造成不同的測驗結果，因此日期的選定要配合未來此測驗的使用目的。⑷施測指導語、答案紙、記錄紙要清晰且有組織，以便在各樣條件下都能順利進行，確保施測情境的標準化。

七、試題分析

所謂「試題分析」是把受試者對試題的反應資料拿去作統計分析，以了解各試題的特徵。進而判定各試題的品質及是否適用於某一測驗中。試題分析方法有兩種，一種是已經用了將近六十年的「古典試題分析法」，另一種是最近廿五年才逐漸被採用的「試題作答理論」（ Item Respones Theory ），它包含幾種潛在特質模式，於個別測驗的試題校準很有幫助。

㈠古典試題分析法

在古典試題分析法中，是將預試後的測驗資料拿來計算通過百分比、各試題與總分的相關，及每個誘答被選的百分比（指選擇題）。

在應用上它又分成⑴組內的應用，及⑵組間的應用：

⑴組內的應用－－

若測驗將來只用於單一年齡或年級組，那預試時只要取單一群體，例如用於高三學生的成就測驗，這時進行試題分析就很單純，不需要有其他的參照群體。

⑵組間的應用－－

當測驗要適於用很多個年齡或年級組時，就要取多個參照群體，而且通常是連續的幾個年齡或年級組。在這種測驗中，試題不只要適用於一個年齡組，而是要適用於多個年齡組。例如，分析智力測驗中的題目時，應該計算每個年齡組通過此題目的百分比，並且要檢查這些通過百分比是否隨年齡增加而增加，換句話說，要畫出該題答對百分比對年齡的迴歸線。

㈡試題作答理論（ IRT ）

由於電腦的高速計算能力加速了 IRT 的發展，IRT 的應用又被稱爲「試題校準」（ item calibration ）、「潛在特質分析」（ latent-trait analysis ）、「試題特徵曲線理論」（ item characteristic curve theory ）。試題作答理論有兩個優點，一是 ITR 基本上是一種迴歸方法，它求得的試題參數不會隨著樣

本的變動而變動，此亦即所謂的「不受樣本影響」的特性。另一優點是IRT是一種試題與測驗分數間的迴歸，它允許個別的試題依據測驗所測的潛在特質來加以校準。

總而言之，不論是古典的或是IRT的方法，都能對個別的試題提供有價值的訊息，而應該選擇那種方法則應看編製者需要那種訊息以及所測量的特質其性質如何而定。

八、編組正式版本

試題分析完後，這些試題大概可以分成三類，即「捨棄」、「修改後可用」、和「可用」。然後先取用「可用」的題目編入一個或幾個測驗版本中，若不夠再從「修改後可用」的題目中取用，如果這兩類仍不夠用，那就只好重新再編擬試題並重新預試一次。爲了避免浪費時間及增加編製成本，編製者應準備較充裕的試題來進行預試。

若同時編製多種版本的測驗時，應該使各版本間的涵蓋內容、試題難易度，試題間交互相關儘可能相同。使用試題統計量可以協助編製者達到平行測驗版本在統計上的要求，例如，它們之間有相同平均數、相同變異量及相同的交互相關。但在實際上，這種不同版本在統計量上的配對是不可能完美的，因此，編製者只能儘可能接近這個理想。

在編好了題本之後，還要準備其他輔助性材料，例如，施測的指導語，答案紙或記錄紙。就個別測驗來說，可能還有特殊的操作材料需要設計和製作。

由於不可能在收集常模資料之後，再變更指導語或答案紙，所以在準備這些材料時要非常仔細、小心，最好有多人共同審閱、討論。

九、取樣與施測

由於從常模樣本取得的測驗資料常被推論到整個國家的母群體，因此抽樣過程要非常注意其代表性。抽樣過程要考慮多種因素，比如，性別、種族、區域、社經地位、學校性質等，並使樣本在這些變項上的分配比例符合人口普查的分配比例。

每個建立常模的計畫常在「嚴格的科學現實」及「實施方便性」兩端中擺盪，編製者通常只能在有限的施測經費下儘量求得常模樣本的代表性了。在實施上，編製者最好先編製好常模樣本分配表，以作爲實際取樣時的依據。

㈠團體測驗的取樣

當測驗是以班級爲單位來取樣時，很明顯的一個好處是每個受測者的平均施測成本比個別測驗少很多，另一個好處則是省時又方便。在以班級來進行取樣時，編製者常以「多階段的分層隨機群聚抽樣法」進行取樣，這種取樣法有三個好處：⑴它可減少編製者因個人主觀偏好或貪圖方便而抽取的樣本，⑵樣本中每一個體被抽中的機率都已事先知道，⑶因事先知道抽中機率，所以可以計算加權量以便校正某類樣本過多或過少時所造成的偏差。

因爲使用多階段的群聚抽樣並配合可降低施測成本的團體施測法，結果團體測驗的標準化樣本都比個別測驗來得大。雖然樣本大並不保證常模就比較精確，但在有良好設計的取樣計畫下，樣本愈多卻有代表性。

在團體取樣的過程中尚要考慮學校配合意願的問題，由於某些學校拒絕接受測驗，編製者須連絡配合意願較高的學校，所以最後實際受測的對象是符合抽樣條件的志願受測學校，這與所謂的「隨機」就

有點差距了。

㈡個別測驗的取樣

由於要個別施測，使得個別測驗在建立常模時不能用機率取樣，只能用「配額取樣」（quota sampling ），在這方法下，每一個參與建立常模工作的測驗中心都分配到一定的配額，然後再由各中心的督導人員訓練施測人員及找尋合乎配額所指定條件的受試者。由於受試者同團體測驗的樣本一樣都是自願的受試者（學童要得其父母同意，成人要其本人自願），編製者無法指出以此法建立的常模精確到什麼程度。

有幾個方法可以減少取樣上的誤差，而較常用的是先用簡單問卷調查受測對象的基本資料及其是否願接受測驗，然後再利用電腦從自願受測的且又合乎取樣條件的人中隨機選取，此法雖不算機率取樣，但它減少了主觀判斷的影響。經驗告訴我們，由各測驗中心的人員去選取樣本時，最常見的誤差是把受測者的社經地位歸類錯誤。以電腦隨機取樣的方式不但減少了各測驗中心取樣上的工作負擔，也避免了取樣的錯誤；同時，如果受測者因故不能受測，也可以很快地找出合適的替代者，不必浪費施測者的時間。雖然這個方法是多了個問卷調查與建檔的工作，但對整個取樣施測工作正確性與效率上的提升卻是大有幫助。

一般而言，個別測驗的常模在每一年級或每一年齡組上至少要有一〇〇名受測者；這與團體測驗每一年級有數千人的常模當然是難以相提並論的。即使經過仔細的抽樣，要以一〇〇多人的資料去推論全國同年級學童的表現，仍是件很冒險的事，但是編製者在有限的經費與時間下，只能藉著最新的全國人口普查資料，按著性別、年齡、地區、社經地位、種族等變項的比例去取樣，以求得最具有代表性的樣本。

個別測驗的取樣工作已很複雜了，但最重要的還是施測者和受測者的合作，在個別測驗過程中會有很多無法預見的意外狀況發生，若受試者故意不合作或施測者怕麻煩而隱瞞事實，那所得測驗結果將不再正確，而常模自會有所偏差。

㈢輔助性研究

在建常模的同時，亦要進行一些特定的研究，以支持此測驗的功效。例如，以同一版本或複本來進行重測信度研究是最常見的。另外，如新、舊版本間的等化研究，不同版本或水準間的等化研究，求效標效度（測驗分數與另一標準化測驗分數、學業成績或工作考績之間的相關）。

這輔助性研究的樣本可能來自原先的常模樣本，也可能是另外施測的。至於選擇此研究樣本的標準則不像建常模時那麼嚴格，除非這研究是要以特殊對象（如聾生）作研究。在作信、效度研究時，因常以相關係數來表示，所以要考慮到取樣時受試者能力的分佈情形，以免取到同質性太高的團體，使能力分佈變小，進而削弱了相關係數。《教育與心理測驗標準》一書（ AERA,APA,&NCME,1985 ）（國內中國行爲科學社有翻譯本）提出一些指引，說明測驗正式出版時應該具備那些附加資料，頗值得編製者參考。

十、建立常模

㈠檢查答案紙與基本資料編號

受試者寫在答案紙上的基本資料，例如，性別、年級、出生年月、地區等資料，在進行電腦建檔之前都要給它們編上數字代號。而這編碼系統要事先設計好，並印成表格交給負責編碼或統計分析的人隨時參考。

在團體測驗中，大多採用機器計分（光學掃瞄式）的答案紙，這時只要檢查答案紙上所畫記號是否

太輕、是否超出規定位置而會使機器誤讀，否則不太需要人工檢查。

在個別測驗中，每份記分紙的基本資料都要加以檢查，看它是否符合分配給各測驗中心的配額的條件。除此之外，偶爾也要把記分紙上的內容用人工計分賦與分數或轉換成數字代碼代表對錯。

㈡資料建檔與電腦計分

如果答案紙無法以機器閱讀，則經編碼的基本資料及試題原始答案就得以人工鍵入磁片中，再用電腦進行計分及統計分析。在電腦或輸入人員不足的情況下，也可以採取先由多人同時用答案卡進行核計總分的工作，而每份試卷只輸入基本資料和總分即可。

可用機器掃瞄的答案紙，是經過精密設計的，其紙張的厚薄及透光率也有一定的規格，有些更精密的設計可使正反兩面作答位置不重疊，而同時讀取兩面的答案。答案紙一但經機器掃瞄之後，原始答案就存入電腦記憶體及磁片上，這些答案可在電腦中與標準答案相比對並算出個人原始總分，並可進一步再作統計分析。

㈢統計分析

當求得個人的原始分數後，即可進行統計分析，這些分析通常先求每個分測驗的基本訊息，例如原始分數的次數分配、累積百分比、平均數和標準差。這些描述性資料可以說明常模樣本在此測驗上的表現如何，若有必要亦應以年級、性別分組，分別求取這些資料。

若是有多種分數的測驗，各分測驗間的交互相關也應同時求出來。辨認受試者身份的基本資料也要計算，以便製作常模樣本分配表。

在上述的基本分析之後，接下來可能作母群與下層團體間的比較、各分測驗與總分之相關，及因素分析等。若常模表已建立好，常模樣本的原始分數通常會轉換成衍生分數，編製者可以拿這種常模分數作進一步的分析。

㈣編製常模表（分數換算表）

常模表是供測驗使用者以原始分數查出衍生分數（百分等級或標準分數）的分數對照表。它可用來解釋個人在某測驗上的表現，通常是印在測驗指導手冊的後面。在常模參照測驗中，它是手冊上必備的部分。

常模可以分成兩大類，一是「發展性」常模，另一是「組內比較式」常模。發展性常模中最常用的是「年級當量」（ grade equivalents ），和「年齡當量」（ age equivalents ），它們是把個人的表現與不同年級（年齡）組的平均表現去比較，然後說此人的表現相當於幾年級（歲）兒童的表現。由於發展性常模並非適用於所有學科或整個年齡範圍，以及它換算出來的結果不能作進一步統計分析，所以逐漸被廢棄，改採「組內比較」的方式來解釋測驗分數。

組內比較式的常模其分數換算方式很多，最常見的是「百分等級」、「標準分數」、「標準九」、和「T分數」等。由於它是以單一團體的表現爲參照，所以常以年級或年齡來區分團體，而所建立之常模亦稱爲年級常模或年齡常模。由於此形式的常模能更精確地指出個人表現在某團體表現中的相對地位，又便於進一步作統計分析，所以使用得很廣泛。

製作常模表既是藝術也是科學，在以往編製常模表常要以繪圖法靠著視覺與手將次數分配圖的資料加以修勻，而現在則利用電腦程式來繪圖和修勻。

出現在指導手冊上的常模表是經過很多複雜的調整、修正的步驟，在這過程中，常模編造者需要有充分的統計學知識及處理資料的能力；另外，他要有技巧地補充或調整異常的資料同時又要保持這資料基本上的完整。

如果所收集的試題都使用一共同的IRT模式來校準，而且其潛在特質量尺是以一個適當的群體加以標準化；那麼新試題或測驗版本若以原來那些試題來加以校準，則這些新試題或新測驗版本也可以使用原來的常模。這種應用是以IRT進行題庫（ item banks ）的標準化，它將所有的試題以一個共同的試題

難度量尺連結在一起，它可以從題庫中抽取題目構成不同版本的測驗，但仍使用同一個常模。

㈤附帶的研究

在建立常模的同時，編製者也要收集支持此一測驗的信度、效度資料。在某些情況下，在建立常模之前就得先進行不同測驗版本間的等化（ equating ）工作，這點在能用於多個年級的多層次測驗組合中尤其需要。

十一、擬定出版進度表

已編製好的測驗如未能如期出版，不但造成出版商資金的積壓，也會使參與編製工作的人員深感挫折。出版進度表通常分指導手册，答案紙、題本三類分別訂定編印進度表。例如：在指導手册方面，其工作項目依序爲：編製常模表、撰寫手册內容（一般說明、施測及計分方法、技術性資料、結果解釋方法及範例等）、編輯、打字、校對、改正，再校、印刷、裝訂、交貨入庫。在這些工作項目中，最容易耽誤時間的是撰寫手册的工作了，因此能夠愈早開始愈好。

十二、印製過程

印製測驗材料需要有出版業的各種專業人力的配合，由於電腦的應用，所謂「手稿」大都早已打在電腦磁片上，而編輯排版工作也已改用電腦文書排版系統直接在螢幕上進行。電腦也可以把用統計式算出來的常模表直接轉到文書處理系統中，免去了抄寫或拷貝時可能產生的錯誤和麻煩。

手册上的圖表愈多，製作程序愈複雜，出差錯的機會也愈多。如果答案紙是用機器計分的，那答案紙的設計要找熟悉該機器掃瞄規格的專家來設計，而不是請一般美工人員來設計。個別測驗所用的器材也需看材料性質請不同專業人員來製作。

十三、銷　　售

測驗和其他出版品一樣要以促銷，通常其促銷方式有，一以廣告信函直接通告可能的客戶：二在專業期刊或相關雜誌上打廣告：三派經過訓練推銷員登門推銷：四在專業研習會中展示產品。

十四、發行後的研究

某些測驗在出版後仍需繼續收集資料，尤其是該測驗在實際應用上的效度資料。大部分的新型測驗都是缺乏效度驗證的資料。出版商應留意客戶對其產品的反應，決定是否增印其他的參考資料；而指導手册也應定期更新或添加新資料，編製者應把他人對該測驗所作的研究結果摘錄進其修訂的指導手册中。

十五、對整個測驗發展工作的事後評鑑

出版測驗就和發展其他產品一樣，有成功也有失敗。發展過程中有些工作能順利進行，有些則問題百出，評鑑的目的就在於找出成功的原因或問題的所在，評鑑不是在追究工作人員的責任，而是提高員工對問題的覺察及如何在下次的測驗出版工作上解決這些問題。除非員工能對評鑑的目的有正確了解，且願以積極的態度面對評鑑，否則評鑑並無益處。

十六、結　　語

本文簡單地介紹了整個測驗出版商發展標準化測驗的過程，雖然實際的程序可能在各出版商之間有所不同，但本文所提示的程序都是最基本的歷程。至於各階段中，某些細節只能點到爲止，無法深入探討，讀者若有興趣，可以自行閱讀教育與心理測驗的專書。