

國家考選的策略與可行作法

朱錦鳳*

摘要

台灣無論在學校教育或人才選拔制度上都是一個測驗導向的國家，始終以來考選部都肩負著國家考試的重責大任，不僅在公務人員考選技術及評量方式，隨時代潮流不斷積極求進步，在專技人員證照考試方面也廣納建言，努力與國際接軌，深獲社會信任。近年來無論在電腦化測驗的建置、心理測驗於國家考試的研發、口試技術的精進、核心職能為概念出發的實作評量、情境式試題的發展等都有蓄勢待發的趨勢，實應給予高度的肯定與鼓勵。本文主要針對各評量種類及用途作一簡單介紹，再針對公務人員考試應具備的特色功能作一歸納性建議，如多元化評量、分階段淘汰、電腦化測驗、心理測驗、測驗效度、核心職能等，最後結合教考訓用之分工目標，並著重於國家考試的考選流程，作一整體性的策略建議。

關鍵字：測驗效度、多元化測驗、分階段淘汰、考選流程

* 東吳大學心理系副教授 (Associate Professor, Department of Psychology, Soochow University)

The Function and Strategy of National Examination in Taiwan

Gin-Fon Ju

Abstract

Taiwan is a test-oriented country both in education entrance and government employment. The Ministry of Examination is the principal administrative authority responsible for the conduct of national examinations in Taiwan which has always been an important duty. The Ministry of Examination has been pursuing the government employment's entranced examination according to the world trend, not only in the techniques and methodology for selection but also for the license and certification for profession and technician. It shall have improved many national examinations and has earned a great trust from the society. In recent years, many revolutions such as in building computerized testing, developing psychological tests, structuring oral techniques, emphasizing performance measures from job analysis, conducting situational exam and so on, have indeed obtained a high degree of affirmation and encouragement. The purpose of study, firstly, introduces the types and applications of some evaluation methods, and then proposes some features for an optimized national examination, such as multi-functional evaluation, sequential strategy of selection, computerized testing, psychological test, test validity, job analysis, and so on. Finally, the study integrates educating, examining, training and employment all together, and aims to provide a suggestion for the selection procedure of government employment in national examination.

Keywords: Test Validity, Multi-Functional Testing, Sequential Selection Strategy, Selection Procedure

壹 前言

以考試作為國家選才的制度一直是我國自古以來的傳統，沿襲至今國家考試也隨著時代轉變而做了許多改進。目前台灣是一個測驗導向的國家，而筆試仍是國家考試的主要方式。筆試其實分很多種類，如何以多元化方式發揮不同

測驗種類的功能，以達到成功選才的目的，是國家考試的努力目標。工欲善其事，必先利其器。首先，本文先對相關的測驗或評量之種類、定位及其可發揮的功能、目的作一簡單的介紹。

貳 相關測驗之定義及屬性

測驗以其題型來分不外乎兩大類：認知能力測驗和情意測驗。認知能力測驗的試題有正確對錯，主要測量應試者的認知能力，又稱測驗題，其題型大多以客觀計分的選擇題為主，但也有主觀計分的申論題或繪圖設計的題型。而情意測驗是反映一種主觀作答傾向、沒有正確對錯，主要測量應試者的人格特質、態度、喜好或感覺的狀態，最常見的題型為客觀計分的李克式量表（Likert Scale），但也有主觀計分的投射技術。

一、認知能力測驗

認知能力測驗主要測量個體的能力，依其功能及目的又可細分為兩大類：成就測驗及性向測驗。成就測驗主要應用於學校教育，是評量應試者專業

知識的學習精熟程度，可經不斷反覆練習而增加其能力。而性向測驗，類似智力測驗，主要測量應試者可學習、培訓的潛能，主要目的在預測其未來學業表現或工作績效，理論上，不致因短時間有太大改變。

（一）成就測驗

成就測驗主要測量個體的學習成效，若以其特性及分數解釋的差別來分，又可細分為一般成就測驗及標準化成就測驗。一般成就測驗的及格標準視該學科知識精熟水準而定，通常又為命題者對知識精熟的要求，故又稱為效標參照測驗（Domain-Referenced Test or Criterion-Referenced Test）。而標準化成就測驗也是成就測驗的一種，但差別是標準化成就測驗是心理測驗的屬性，具備標準化程序、信效度及常模，因此分

數經常模轉換後可供跨區域或跨年度的應試者相互比較用，故又稱常模參照測驗（Norm-Referenced Test）。

一般成就測驗與標準化成就測驗在題目內容、施測對象、分數解釋等方面都有相當大的差異（Anastasi & Urbina, 1997）。基本上，一般成就測驗的試題內容是針對特殊教材或特別課程所設計的，屬內容導向的測驗，主要的施測對象是針對一群特定修過課程或讀過教材的學習者，其分數好壞解釋與教材內容和課程目標息息相關，較偏向命題者主觀標準。一般成就測驗其包含範圍非常廣，從教師自編的小考、期中考、期末考，到國家證照考試或聯考等都可能都是這類型測驗。而標準化成就測驗其試題內容則偏重藉由生活常識作為命題媒介，範圍不限定於特定教材或教學而設計，針對特定的母群施測，如國中生、高中生或大學生等母群，重視預測效度，且其分數好壞反映的是應試者在該特定母群的相對表現。

通常，授課教師若欲了解班上學生的學習狀況，會應用一般成就測驗，但若為大型之國家考試則建議應用標準化成就測驗。我國學校目前幾乎都用一般成就測驗評量學生學習成效，但美國的學校則會用標準化成就測驗作取代，其主要原因是一般成就測驗其內容及品質易流於主觀的標準，並且無常模可作跨校或跨年度的比較分析。

（二）性向測驗

心理大師 Ann Anastasi（Anastasi & Urbina, 1997）曾指出性向測驗（Aptitude Test）是繼智力測驗後，以篩選及分類為目的所發展之分能力的類智力測驗，有別於一般所指之與學術（academic）或訓練（practical）相關的學業性向或特殊職業性向測驗。人類的智力經因素分析後大致分為幾大能力，如語文推理、數字應用、抽象推理、空間關係、知覺速度、機械推理、語文運用等。換言之，性向測驗主要是測量應試者的多元認知潛能。許多人會誤以為性向測驗是測量個人興趣、喜好的測驗，其實不然；性向測驗是有正確對錯的認知能力測驗，一般以選擇題命題、團體施測，主要測量人們各方面的潛能和天份。與智力測驗不同的是，性向測驗著重在應試者各分能力表現的強弱分佈狀態而非總分高低。

通常性向測驗是輔助決策的最佳工具，因為可團體施測、甚至電腦化，因此可處理大型團體人員的甄選作業。這正是性向測驗繼智力測驗發展後的價值及定位。性向測驗的用途非常廣泛。當性向測驗以學業成績當效標驗證效度時，其目的是預測學生之學業成就表現；當以職業種類當效標時，其目的是預測職業成就；當以工作績效為效標時，其目的是預測其職能表現。當然，工作績效並非全然以能力或性向測驗即可決定，人們對該工作的喜好程度及熟

悉程度亦能影響其就業之穩定度與成就感，因而間接影響其工作績效。

二、情意測驗

(一) 人格測驗

人格測驗屬於情意測驗的一種，主要測量應試者之行為特質傾向或態度喜好程度，答案沒有正確對錯。測量人格特質的方式可分為兩種：自陳式量表和投射技術。自陳式量表是人格測驗最普遍的作答方式，而李克式量表又是自陳式人格測驗最常使用的一種作答題型。通常，自陳式人格量表的題目內容是直接敘述所要測的行為特質，應試者依據題意選擇一個最貼近自己感覺的選項，來顯示自己的情意傾向，其可能導致應試者作假或符合社會期許的現象，而造成測驗結果失真的問題。因此，人格測驗特別重視檢核量尺的設計，如測謊量尺等，就是為了避免受試者作假或符合社會期許的檢核機制（朱錦鳳，2010, pp.406-408）。

由於不當的題目敘述可能會直接影響人格量表分數結果的精確性及客觀

性。因此在編製人格量表時，題目內容本身的敘述顯得格外關鍵。題意敘述適當與否，可能直接影響一份人格量表的效度。為避免不同個體對相同的題意產生主觀知覺或解讀的差異，因此題目敘述應謹慎處理下列情況：符合社會期許的作答反應、慣性作答反應誤差、反向敘述的問題、負向情緒字眼、雙否定敘述句、模稜兩可的詞彙等。

(二) 投射測驗

投射測驗又稱為投射技術，主要是以非結構性的模糊刺激，以受試者為中心來測量人們的潛意識需求、行為動機、內在衝突及情緒狀態。投射測驗大多為個別化施測，應用在臨床個案居多，可整體性地了解個體的內心世界。不像其他文字敘述的人格量表，投射測驗的題目通常是非結構性的圖形或語文的刺激，因此最大的優點是可有效避免受試者的作假及自我防衛。投射技術大致可分墨漬刺激、故事統覺、語句完成、繪畫表達等四種型態。本文重點不為投射測驗，在此不贅言。

參

效度種類

當測驗應用在人事決策用途，測驗效度是決策者最要重視的部分；而公職人員任用的國家考試應更要重視考試效度的評估（彭錦鵬，2005）。測驗效度

是指一份測驗能有效測量到所欲測量行為特質的程度。測驗效度又分三大類：內容效度、建構效度及效標關聯效度（Anastasi & Urbina, 1997）。

一、內容效度

內容效度 (Content Validity) 是指對測驗內容作系統化的檢驗，以確定其試題涵蓋該測驗所欲測量之能力特質的代表性內容。驗證內容效度的方法主要可分為質化與量化分析兩種。質化分析主要是以內容主題專家來判斷內容的品質。所謂內容主題專家，就是熟悉測驗所測之能力特質的專業人員。內容主題專家可能是該測驗所測之能力特質的專家，也可能是編製測驗的專家。

內容效度的量化分析是指項目分析 (Item Analysis)。簡單地說，項目分析主要是針對測驗所包含的各個題目，分析其難易度、鑑別度、一致性或社會期許等，甚至選擇題的誘答力分析也是其中一種。項目分析的主要目的是刪去品質不好而留下品質較佳的題目，藉此提升該測驗之效度。所謂品質佳的題目著重在鑑別度良好的題目，也就是能夠正確分辨受試者能力好壞的題目。

二、建構效度

建構效度 (Construct Validity) 是指測驗的編製過程中，以客觀、量化方法來驗證一份測驗的結構是否符合其理論假設，以反映該測驗在建置結構的有效程度。尤其針對測量抽象特質的測驗，如智力、幸福感等，就應特別重視建構效度的驗證，以區辨所測特質之具體行為表徵。驗證建構效度的方法有許多種，舉凡因素分析、結構方程模式、分

測驗間的相關分析、聚斂效度、試題反應理論等皆是。

三、效標關聯效度

效標關聯效度就是測驗分數能預測到外在效標的有效程度，針對測驗所要測量的能力特質，所取得之量化外顯行為指標稱為外在效標 (Criterion)。例如，成績、工作績效、考評、診斷或其他相關的測驗等，都是驗證效標關聯效度時經常使用的外在效標。換言之，效標關聯效度就是測驗分數與量化效標進行相關分析而得，通常以統計 r 表示。效標關聯效度若以效標取得的時間來分，又可分為同時效度和預測效度。

心理測驗的編製過程中，效標關聯效度是三種效度分析方法中，最後一個階段且必須持續進行的效度驗證。效標關聯效度對篩選決策而言，也是最重要的效度驗證方法，因為它關係著該測驗工具能預測其外在效標的有效程度。而效標關聯效度係數的平方值，就是該測驗能有效預測其外在效標的百分比。通常，我們會將要預測的行為特質，如工作績效，當作該測驗的效標進行相關分析，理論上應達顯著正相關，再以其相關係數的平方值計算該測驗能測量其效標的預測量。換言之，一份測驗若沒有提供效標關聯效度的相關資訊，則該測驗很難作為一個決策的輔助工具。

肆 職能評量種類

近年來許多學者應用企業人事甄選及職能分析的角度切入國家考試（吳三靈、王崇斌，2006；彭錦鵬，2010；熊忠勇，2010；蔡秀涓，2010；鄭夙珍、林邦傑，2009），這種概念類似工商評量（Industrial & Organizational Assessment，I/O Assessment）又稱職能評量的方法。通常，職能評量會針對不同公司、或不同職位之工作性質而進行工作分析，以協助企業在人事甄選、訓練、升遷及考核時能達到最高的決策效能。職能評量的最大特色是必須以工作分析為基礎，且以工作績效為效標。所謂工作分析（Job Analysis）就是針對該公司或職務特性分析其工作內容，並得到一個條列式、有系統、有組織、符合企業要求的資格、能力或特質等項目。所謂職能評量就是以評量方式預測到這些職能項目，其評量方式可採筆試測驗或實境方式進行。

職能評量必須非常重視預測效度的驗證。理論上，發展職能評量最恰當的人選應該是該機構人事部門的研發人員。因為他們對公司內部的組織、結構、運作、需求及財務狀況等各方面的了解最為透澈，但他們不一定對測驗編製、評量技術及信效度驗證有充分的專業知識。因此，目前國內職能評量在使

用上仍然有偏重內容效度而忽略預測效度的情形。未驗證預測效度的職能評量應用於人事決策時是非常冒險的。一份依據工作分析而編製的職能評量，除了須克服要將其有效轉化為試題敘述的困難外，還必須進行項目分析、信效度分析，以及結合決策理論。然而，令人氣餒的是，至今成功職能評量的例子仍非常少見，尤其採紙筆方式的職能評量更難有效克服文字所能測得的侷限，因此將職能帶入評量的目標目前仍停留於理論。

職能評量的種類五花八門，從最簡單的履歷表到性向測驗、人格量表、情境模擬、面試、試用等，都可能是企業選用的方式（Gregory, 2007）。不同種類的職能評量各有其優勢及適用情況，各企業通常根據其需求和預算，選擇單一或綜合的方式輔助員工甄選決策，以達到最大成本效益。因此，我們先對職能評量的種類、功能及定位作一簡單介紹。

一、履歷表

履歷表是提供個人背景及學經歷資料的最基本依據。尤其像性別、年齡、學歷、證照、成績單或婚姻狀況等客觀條件，很可能是企業機構優先考量的因

素之一。履歷表通常有助決策者預測應試者未來的工作能力、職業穩定度及適應狀況等。理論上，履歷表呈現的是個人背景條件及學經歷事實，因此應徵者很難作假或浮誇，這是履歷表的最大優勢。

舉例而言，Cascio (1976) 的研究曾發現履歷表中各工作年資及更換頻率，可有效預測員工未來的就職年限。換言之，履歷表和員工的工作穩定度有顯著的正相關。另一研究則發現，履歷表對員工未來工作表現的預測效度和性向測驗一樣好；但由於履歷表預測的效標可能與性向測驗有所不同，因此專家建議履歷表與性向測驗若能兩者並用，對預測應試者未來工作表現有加成效果 (Drakeley et al., 1988)。

二、性向測驗

性向測驗的目的是篩選具備特殊潛能的應試者，通常應用在需要特殊技能的工作職位上。由於性向測驗若以電腦化測驗取代，其施測、計分、結果都可自動化以降低成本，因此經常安排在員工甄選的第一階段進行；如果再和履歷表合併使用，能提供更大的加成效果。一般而言，區辨效度高的性向測驗是人事決策者的首選，對於員工篩選非常有幫助。經常被使用在人事決策的性向測驗通常包含語文推理測驗、數字應用測驗、機械推理測驗、空間關係測驗、知覺速度測驗等。雖然一般智力比任何

單一的性向測驗，對應試者未來工作表現具有更好的預測效度 (Outtz, 2002; Robertson & Smith, 2001)，但一些特殊職位則除外。目前在美國經常被使用的特殊性向測驗，有機械性向測驗、知覺速度測驗等 (Ghiselli, 1966, 1973)。

三、人格測驗

有些特殊的人格特質是特定職務或企業需要的基本條件，如主動、外向、領導、表現等。在國內，某些正向的人格特質也是大部分工商機構受歡迎的加分條件，如正直誠實、勤勞穩定、合作順從等。因此，人格測驗也是企業機構人事甄選普遍使用的職能評量工具。雖然許多研究指出，大部分的人格特質對於員工之工作績效及表現相關都不高 (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990)，但仍然有某些人格特質曾被研究發現可以有效預測應試者的某些工作表現，例如說謊、偷竊、曠職、拖拉、吸毒、犯罪等；但這些人格特質通常不容易獲得真實的資料 (O' Bannon, Goldinger, & Appleby, 1989)。在美國，評估這類人格特質的測驗統稱為廉貞測驗 (Overt Integrity Test)。

四、情境測驗

有些技術性的工作很難以筆試的方式準確測量其工作表現，因此需要工作情境或模擬的方式，來評估應試者隨機應變及問題解決的能力。例如，評估

處理行政作業的打字速度測驗、機械維修的機械組合測驗、交通道路管理的地圖辨識測驗、縫紉機師的工具靈活度測驗、雜誌編輯的標題版面及故事組織測驗、公關的語文表達測驗、通訊業務的角色扮演測驗、企業管理的商業書信測驗等，都曾被研究發現是具有良好預測效度的情境測驗（Asher & Sciarrino, 1974; Muchinsky, 2003）。

五、工作模擬

工作模擬是實際工作職務的微縮版，它不須包含完整的工作面向，只需著重在較困難且區辨效度較高的某些技巧元素上。工作模擬尤其對高成本、高代價、不容許失誤的工作職務篩選非常重要。換言之，工作模擬就是為了避免新進員工不適任所造成的賠償負擔，而以實際工作執行或操作程序的關鍵步驟作為篩選員工的資格測試。如飛行員、醫護人員、消防人員、軍人、警員等，就是以工作模擬的方式來確認應試者是否能立即勝任該職務。有些管理決策性質的職務，也會以工作模擬的方式測試應試者的決策判斷、執行經驗、人際互動及問題解決能力。

六、口試

由於口試對企業而言所耗費的成本較大，通常決策者會將面試安排在第二或最後階段進行。研究顯示，口試可能造成主觀偏見或月暈效應（Mayfield,

1964; Ulrich & Trumbo, 1965; Wagner, 1949），因此不建議當成唯一的決策工具。研究也發現，以一組決策委員的口試方式會比一人口試的方式明顯提升其結果的一致性，且結構性口試的效度是非結構性口試的兩倍（胡悅倫、陳皎眉、洪光宗，2009）。除非口試與其他評量工具（如認知能力測驗或人格測驗）同時並用，否則口試本身在決策研究中大多不被支持及看好，但仍有研究發現，結構性加上情境式的面談方式可提高口試的預測效度（Borman, Hanson, & Hedge, 1997; Maurer & Fay, 1988; Schmitt & Robertson, 1990）。

七、試用

試用是企業機構經常用來取代情境測驗或工作模擬的一種方式。由於情境測驗或工作模擬都必須先研發一個微縮的工作模擬情境，再驗證該情境、模擬的表現是否具有預測效度，這通常需要高專業人力及高成本才能達成。因此，一般企業為省時省力，最快速的方式就是讓應徵者實際進入機構一段時間，親身經歷真實的工作情境，再評估其工作表現。通常試用時間以一至三個月為原則。試用期可獲取許多應徵者相關的重要資訊，如人格特質、適應狀況、工作能力表現及其他各方面與公司的契合度等。試用期的最大缺點是企業無法接納太多的試用人員，通常試用人員不會大於預備錄取的正式員工數。由於試用

階段需要耗費許多人力、時間及成本訓練，因此若選擇以試用作為人事決策的

方式，建議將試用安排在口試之後進行。

伍 公務人員考試應具備之特色

公務人員除了要重視其專業知識之精熟能力外，基本的判斷、邏輯思考能力及語文能力也應列為基本門檻，實務職能更是工作不可或缺的能力，此外，態度嚴謹、情緒穩定、接受度高、廉潔忠貞、正直誠實等也都是優質公務人員必備的條件。因此，目前公務人員考試除以筆試測量專業知識外，還應輔以一般智力或性向測驗測量基本邏輯判斷能力，並以人格測驗或廉貞測驗測量倫理準則，最後再以試用方式培訓實務職能。畢竟公部門與民間企業的經營訴求有別，公務倫理及行為準則才是鞏固人民對政府信任的核心價值（陳皎眉、胡悅倫、洪光宗，2010；詹靜芬，2010）。以下是本文針對公務人員考試應具備的功能與特色提出建議。

一、一般智力評量

過去美國許多企業機構經常以智力測驗作為人事篩選的工具（Brigham, 1923），且近年來有些學者也認為包含智力、特質的核心職能更能預測員工之工作績效（Boyatzis, 1982; Spencer & Spencer, 1993）。因此，公務人員考試除須評量應試者的專業知識及技能外，

其一般智力，如邏輯力、判斷力、學習力、推理思考等，也是其必備的基本條件。

Sternberg（1994）的三元智力理論，是近代最能被理解且最具說服力的智力理論，也是目前最能落實在真實世界的人類智能理論。他認為所謂成功的智力，應該是一種調適、塑造及選擇環境的能力，以實現自己的目標及社會文化所扮演的角色。Sternberg提出的三元智力理論（Triarchic Intelligence Theory）指的是分析智能（Componential/Analytical Intelligence）、經驗智能（Experiential/Creative Intelligence）和應用智能（Contextual/Practical Intelligence）三種面向的智力。分析智能是傳統智力測驗所測到的認知能力，包含類化、執行、推理、知識等能力；經驗智能則是個人面對新環境時，結合舊經驗與新經驗，所表現出的自動化行為歷程；應用智能是指個人適應環境的綜合能力，包括適應環境的能力、改變環境的能力及選擇環境的能力。當每一個人在這三方面的智力結構強弱不同時，即代表每一個人的獨特性。換言之，會考試的人通常認知能力較佳，但不必然與其應用、經

驗、適應能力或工作表現有高相關。因此建議公務人員考試可增加性向測驗測量應試者之一般智力。

二、多元評量

不同測驗種類所發揮之功能不一，例如，成就測驗測量專業科目之精熟能力、性向測驗測量可被訓練的潛能優勢、人格測驗測量適合之個性特質等。不同評量方法所具備的特性也不一，例如筆試可評量專業知識、口試可評量儀表應對、模擬可評量實務操作、試用可評量職能及組織適應能力等。所謂多元評量就是運用多種測驗及多種評量方式於人員篩選決策中，使其達到最大決策效益。多元評量強調的是以多種不同性質的測驗及評量方法，最好再配合分階段篩選方式，區隔每一階段之考試功能，以發揮人員甄選之最大效能。換言之，國家考試除筆試外還可增加口試、

實務訓練、人格測驗或性向測驗等項目（陳執中，2005；楊戊龍，2010）。

三、分階段考試

分階段考試策略（Sequential Decision Strategy）就是以分階段考試篩選方式，且配合淘汰制度來提升決策效能。為方便解釋其決策程序，圖1以最簡模式：兩階段決策作說明。兩階段決策就是應試者經第一階段考試通過的人，再挑戰第二階段的考試。通常，這兩階段的考試會運用不同功能的評量方式分階段淘汰。例如，第一階段是測量專業能力的筆試測驗，先淘汰不合資格的人，而第二階段是人格測驗或口試等再淘汰不適任的人。一般而言，兩階段的決策會比單一次的決策效能好，而多階段決策又比兩階段的決策效能好，但須考量其成本效益。劉坤億（2009）也曾建議分階段考試是提高考選效度的關鍵作

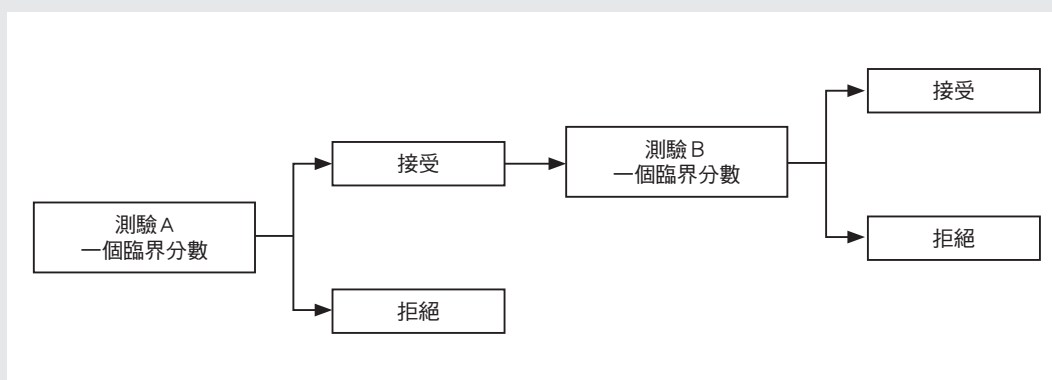


圖1 兩階段決策流程

為。彭錦鵬（2009）也指出我國考選制度目前仍未能以分階段分試淘汰降低考選成本。

四、電腦化測驗

隨著科技進步，電腦化測驗將成為未來趨勢。針對像學校、軍隊、政府、工商機構等大團體的施測，電腦化測驗無疑是一大福音。電腦自動化過程不僅可節省施測及計分的人力、加強標準化程序、還可立即獲得測驗結果，以便有效幫助決策者作篩選。但電腦化測驗也有其侷限，目前僅限應用於有正確對錯的選擇題型，且須一人一機作答。

基本上，電腦化測驗可分電腦測驗（Computer-Based Testing, CBT）與適性測驗（Computerized Adaptive Testing, CAT）兩種。CBT的題目內容與傳統紙筆測驗完全相同，唯一的差別在於各題目是依序單獨呈現於電腦螢幕，但即使看似相同，仍有可能對未曾使用過電腦的受試者造成負擔而影響其測驗表現（朱錦鳳，2002a）。而CAT主要以項目反應理論（Item Response Theory, IRT）為基礎，針對不同能力程度的受試者，給予不同難易度及鑑別度的試題，每位受試者作答最適合自己的題目及題數，電腦以最短的時間準確地估計其能力。適性測驗的三大條件是：大型化題庫、智慧性選題及精確的能力評估（朱錦鳳，2002b）。經過效度驗證的CAT一直被視為是一種優於傳統紙筆測驗的

評量方式。但陳榮坤（2009）曾指出目前應用於國家考試的電腦化測驗試題太難，且質疑全部為測驗題是否能滿足專業職能的評估。

五、心理測驗

心理測驗包含的種類很廣，舉凡智力、性向、教育、職業、臨床、人格、投射等測驗都是。不同種類的心理測驗各有其用途及功能，性向測驗及人格測驗是最普遍應用於美國公務人員考試的心理測驗。人格測驗以用途來分主要可分為兩大類：測量正向人格特質及負向情緒問題的測驗。通常，測量正向人格特質的測驗應試者容易符合社會期許作答，而測量負向情緒問題的測驗應試者容易隱瞞作假。因此，本文建議公務人員考試可增加廉貞測驗及人格測驗，但皆須注意社會期許或作假的問題。鄭夙珍、林邦傑（2009）也曾指出若將人格測驗列為國家考試應發展完備之測謊機制。

陳皎眉、黃富源、孫旻暉、李睿杰（2011）曾提及國家考試的筆試大多是成就測驗，無法測出應試者的口語表達能力、社交能力、危機應變能力等特殊能力，而心理測驗是人事甄選的最佳決策工具，但須具備良好的信效度。陳執中（2005）也強調在專業科目知識豐富不代表其人格特質符合工作所需。陳皎眉、胡悅倫、洪光宗（2010）建議人格測驗應在第二階段配合口試使用，除要

注意其信效度的驗證外，還要處理應試者可能說謊的問題，其結果也不應作為決定應試者是否錄取的決斷工具。

六、注意紙筆測驗的侷限

測驗並非萬能，尤其是紙筆測驗，以文字方式預測人類的能力或特質仍有其侷限，不同種類的測驗又有其不同的侷限。舉例而言，自陳式人格測驗由於題目敘述過於直接或假設性，因此應試者容易產生作假或社會期許的現象（朱錦鳳，2011）。社會期許的作答誤差至今仍是國內、外測驗專家的一大挑戰，雖然投射測驗可有效克服作假的問題（朱錦鳳，2009），但無論如何，當涉及國家考試之就業公平性時，仍必須有效評估其測驗效度。

林文瑛（2009）曾一語道出各類測驗的侷限。她說國家考試多半依賴學科知識紙筆測驗，但知識不等於工作能力，紙筆測驗也無法測量個人操守、工作熱誠、人格特質等。然而，人格測驗與情境測驗的發展也還未成熟到足以作為國家選才的工具。她強調紙筆測驗有時而窮，關鍵在於它依賴語言，而語言表達經常不是真相，像自陳式人格測驗就容易作假，但情境測驗又如何呢？誰能說領導能力、責任感、溝通技巧等等有固定的標準答案呢？

又如，看似重要的國文科考試，是否真能預測應試者的工作表現，必須進一步分析其預測效度（工作表現），而

不能僅以專家主觀判斷定論。最近被大家看好的情境式國文科試題（何寄澎，2011），究竟測量的是應試者的語文能力、閱讀能力、還是邏輯能力，亦需透過效標關聯效度的實證分析，方可充分瞭解。此外，電腦化測驗因受限於有正確對錯的選擇題型，比較適合評量應試者有關分辨及判斷的認知能力，但針對測量較高層次之組織、整合能力的申論題型而言，至今電腦化測驗仍無法克服其作答及計分的問題（陳榮坤，2009）。由此可見，不同種類的測驗工具各有其重視的效度問題，如何善用各種評量方式各司其職，以發揮最大決策效益是國家考試的努力目標。

七、以培訓建立核心職能

當紙筆測驗不如預期成效，許多學者期許國家考試應以職能理論切入（宋餘俠、陳仕煥，2005；吳三靈、王崇斌，2006；彭錦鵬，2010；熊忠勇，2010；蔡秀涓，2010；鄭夙珍、林邦傑，2009），但以文字方式測量工作績效至今仍困難重重，即使是情境測驗也很難克服主觀評分的問題，其效度分析更是一大挑戰。目前能將職能概念落實在考試的成功例子非常有限。理論上，核心職能的概念落實在公務人員的培訓階段可行性較高。沈昆興（2006）就曾提及優質的公務人力源自學校教育所培育，如何從工作分析將公務人員教考訓用連貫起來、相輔相成是當前重要議

題，期使達到適才適所、人盡其才的目標。陳志瑋和余致力（2005）也曾指出除考選制度外，教育訓練系統可落實培訓的功能。由此可見，核心職能的概念

雖然重要，但不一定能成功落實在國家考試中，本文建議核心職能可有效設計在教育或培訓階段。

陸 考選策略與作法建議

國家考試制度的主要目的為選拔優秀公務人力與專業人才，實扮演國家競爭力之推進者，而綜上所述，不同方式、不同種類的評量工具，可對不同階段的人員甄選發揮不同的評量功效。為使決策達到最大效益，本研究建議應以多元評量方式彰顯其功能，再配合分階段淘汰策略發揮其優質篩選的決策效能。

教考訓用的整合就是期望各機構發揮其分工合作、相輔相成的功能，而不致重複浪費資源。例如，學校機構負責專業知識、技能及人格養成之教育功能，同時也負責專業知識的評量，如成績、學位等。而國家考試的重點則在任職前的資格考試功能，因此除應參採應試者之在校成績進行初步篩選外，還應增加電腦化測驗進行一般智能的門檻篩選。此外，應再增加一些跨校、跨年度的標準化成就測評機制，及一些人格評量機制，如廉貞、情緒管理及適應量表等。取得資格後的職業培訓建議以核心職能為重點加強其工作績效，使未來公務人員之任用適才適所，以發揮教考訓

用的最佳分工狀態。

本文針對教考訓用的連結及其應發揮之功能，以流程圖方式呈現於圖2，並以公務人員考試為主軸作探討，著重多元分試的精神，將公職考試極盡最大選才功能。以下詳述公務人員考試各階段的運作模式。

一、第一階段：智力性向電腦化門檻測驗

目前公務人員考試之錄取率偏低，因此應善用多元評量工具，並配合淘汰策略使決策效益達到最佳。首先，建議考選部參採應試者的教育相關資料，如相關學位、前半排名成績等客觀考評標準作初篩，以減少報考人數。其次，公務人員需要基本智能及學習力，建議以電腦化測驗測量一般智能或特殊性向，較符合成本效益。試題內容著重在應考者之邏輯、推理、分析、語文能力的評量，或如機械、核對等特殊性向測驗，以確定錄取者具備整齊優秀的智能。第一階段的電腦化測驗其主要目的為一門門檻測驗，因此應建立淘汰機制。以現階

段報考人數及錄取率，並考量第二階段申論題的閱卷人力負擔，建議第一階段考試淘汰約三分之二的報考人數。剩下三分之一的人數將有機會進入第二階段的專業科目考試。

二、第二階段：專業知識標準化成就測驗

理論上，在教考訓用的完整結合下，教育訓練的修業科目和成績應作為專業知識的最佳指標，不需再重複測驗，但由於各校的課程設計和成績標準

不同，雖無法跨校比較優劣，但絕對可作為書面審查或口試的參考依據。第二階段的專業科目考試建議以申論題方式命題，主要目的在測量應試者之更高層次的組織、整合能力。第二階段的考試其主要目的為一精熟測驗，試題內容以專業知識為主，且最好設計為一標準化成就測驗，以分主題方式建立題庫，如此才能確保測驗信、效度品質及每年題目內容的涵蓋性、難易度、鑑別度等特性的一致性。建立題庫之命題者建議來自多元管道、多元思維命題，如學校開

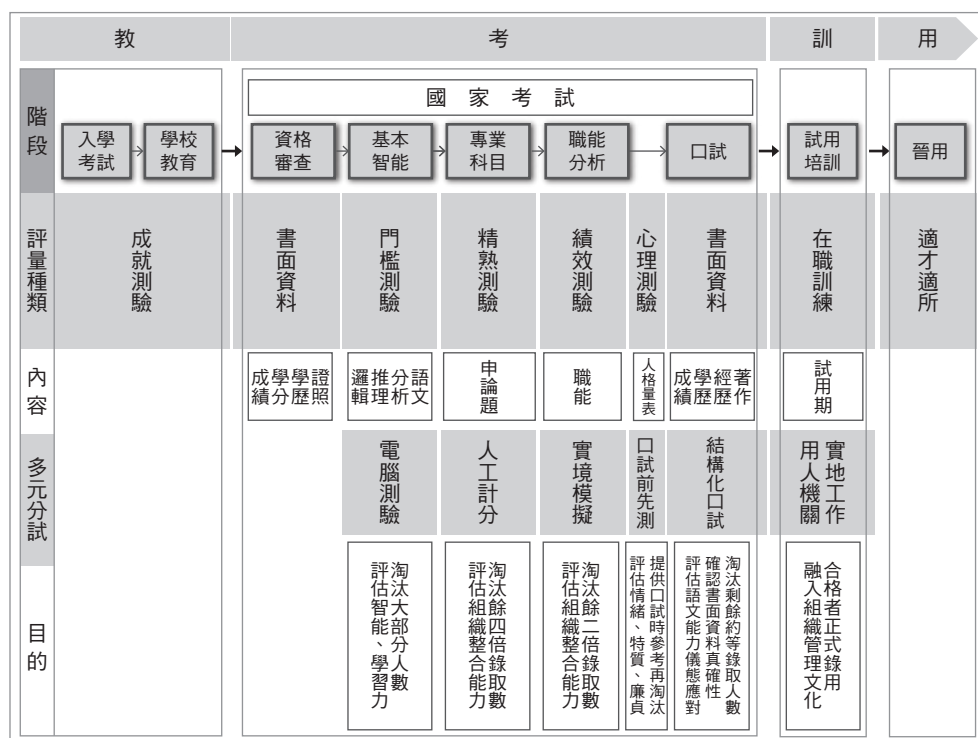


圖2 理想的考選流程

課之教師、具證照之專業人員、公部門資深優秀之高階文官等。試題兼顧理論與實務並重，並以分領域或分章節方式設計題庫叢集，逐年累積增加。每次考題從各叢集題庫隨機抽取固定比例的題數，並包含定錨試題，故可建立跨年比較的固定參照團體常模，以達跨年、跨區比較的功能。由於申論題須人工計分，為提升目前閱卷品質，應考人數不宜太多，且考量下一階段為實務操作的績效測驗，非常耗費人力成本，因此建議本階段考試再淘汰約三分之二的參試人數，剩下約11%的人數將有機會進入第三階段的績效考試。

三、第三階段：核心職能實地演練績效測驗

第三階段的績效考試主要以職能理論為基礎，各職務應以工作分析方式建立其核心職能，由各用人機關參與具體明列並決定其最重要之關鍵能力。績效考試建議以實地模擬方式進行，因為核心職能不容易以文字方式評量精確。實地模擬的考試方式需要專業人員研擬重要關鍵的模擬元素及量化客觀的評量標準，其主要目的在評量應試者的實務操作能力，且其能力需具備有效預測未來工作表現的功能。發展一套具有預測效度的關鍵模擬作業並不容易，這也是目前模擬測驗未普遍使用的主要原因，但針對特殊工作如飛行員、治療師、工程師、記者或高階文官等，核心職能實地

模擬的績效測驗仍不失為最佳方法。本階段考試建議再淘汰一半的參試人數，剩下約5%的原始報考人數將有機會進入最後口試階段。

四、第四階段：表達應對結構性口試測驗

口試透過技巧性提問不僅可確認書面資料的真確性，還可面對面看見應試者儀容、態度、肢體語言、人際應對等筆試無法獲得的重要資訊。結構化口試更可有效評量應試者的語言表達、社會互動、人格特質、甚至核心能力。因此，若透過專業訓練的命題及評分，建議口試可與前階段的實務模擬績效考試合併運用。換言之，從結構化口試過程同時評估應試者是否具備高階文官的核心職能，或從實務模擬考試同時評估飛行員、治療師的口試項目，都是兩階段考試相輔相成的結合性運用。無論是實地模擬或結構化口試，考選部需要投入的人力及訓練都很吃重，因此參試人數必須有效掌控，以落實普及口試的目標。倘若要再加人格測驗的評量，建議可在口試前進行，結果可供口試時參考。

對政府而言，公務人力素質是影響政府績效的主要因素，進用並留任優秀人才，是政府組織提昇整體競爭能力的主要目標。若目前國家考試之筆試與學校學科過於密切，所存在的潛在問題及應有的內涵及定位可能需要更深入探討（呂育誠，2005）。目前考試無論甄試、測驗或檢定皆仍以筆試為主，未來應更重視專業技術及實務經驗的落實，尤其是專技人員的證照考試（楊戊龍，2005）。台灣是一個測驗導向的國家，考試將主導學習，若我國考試制度皆以筆試為主，將會使學生學習偏重專業知識而忽略實務的重要性（賴其萬、曾慧敏，2005）。同時，筆試也會助長補習班盛行，促使考生偏重知識性的學習而忽略解決問題之思考能力，考生也會以考取為目標而忽視人格及倫理道德之培養。目前國家考試制度有兩個主要問題：(1)錄取率低造成教育資源浪費，(2)以筆試為主的考試方式未能確保人力素質（劉奕權，2007）。

由此可見，如何使國家考試多元化，以發揮國家考試的最大決策效益，是考選部的努力目標。許多先進國家都設置有國家層級的評鑑中心，負責評鑑構面的確定、評鑑工具的選擇與研發、評鑑員的遴選與訓練、評鑑任務的實

施、評鑑結果的報告與回饋等任務。本文也建議國內能設置評鑑中心，集結專業人員持續推動國家考試（吳復新，2005）。簡言之，國家考試應以多元評量方式區隔其功能，再配合分階段淘汰策略發揮其優質篩選的決策效能。

依據目前錄取率過低的公務人員考試而言，建議先以書面審查應試者的資格、成績進行初篩。爾後將公務人員考試分三階段完成：第一階段建議以電腦化測驗方式測量應考者之一般智能或特殊性向，淘汰大部分的報考者；第二階段為專業科目考試，建議以申論題方式測量應試者之專業知識，並以標準化成就測驗為目標，持續驗證測驗效度，建立題庫及常模，再淘汰大部分的參試者。第三階段為結構性口試測驗，透過技巧性提問甚至可以發揮實地模擬的功效，從結構化口試過程中評估應試者是否具備核心職能。結構化口試需要投入更多人力及訓練，因此參試人數應控制在錄取人數的兩倍數量，以落實普及口試的目標。最後錄取者應結合用人機構建立試用及培訓制度，以核心職能為基礎加強錄取者的工作績效及組織適應，以達公務人員適才適所之目標。

參考文獻

中文部分

- 朱錦鳳 (2002a)。傳統紙筆測驗與電腦化測驗的比較。測驗年刊，49(1)，75-87。
- 朱錦鳳 (2002b)。國中生物科適性測驗的建構歷程-IRT理論的應用。教育與心理研究期刊，25(4)，615-627。
- 朱錦鳳 (2009)。「畫人測驗」在人格投射的應用。2009年第48屆台灣心理學研討會。台北，台灣。
- 朱錦鳳 (2010)。心理測驗--理論與應用。台北：雙葉出版社。
- 朱錦鳳 (2011)。社會期許測量誤差對量表效度的影響。2011年中國測驗學會研討會。台北，台灣。
- 何寄澎 (2011)。國家考試國文科命題的新思維與新方向。國家菁英季刊，7(3)，1-16。
- 吳三靈、王崇斌 (2006)。公務人力資源管理導入核心能力之研究。國家菁英季刊，2(4)，25-42。
- 吳復新 (2005)。評鑑中心法及其在人力資源管理的應用。國家菁英季刊，1(1)，175-204。
- 呂育誠 (2005)。財稅文官考選制度的興革。國家菁英季刊，1(2)，69-86。
- 宋餘俠、陳仕煥 (2005)。淺談文官核心能力與國家競爭力。國家菁英季刊，1(2)，21-32。
- 沈昆興 (2006)。公務人員教考訓用配合制度之析探。國家菁英季刊，2(4)，1-24。
- 林文瑛 (2009)。國家選才還能有什麼錦囊妙計嗎？國家菁英季刊，5(3)，19-25。
- 胡悅倫、陳皎眉、洪光宗 (2009)。國家考試口試之命題與評分。國家菁英季刊，5(4)，35-56。
- 陳志瑋、余致力 (2005)。公務人員考試錄取人員訓練制度改進之研究。國家菁英季刊，1(1)，99-116。
- 陳執中 (2005)。全球化下的涉外人員考試。國家菁英季刊，1(4)，71-80。
- 陳皎眉、胡悅倫、洪光宗 (2010)。人格測驗在國家選才上之使用與發展。國家菁英季刊，6(4)，1-12。
- 陳皎眉、黃富源、孫旻暉、李睿杰 (2011)。論心理測驗與國家考試。國家菁英季刊，7(1)，15-31。
- 陳榮坤 (2009)。專技航海人員考試電腦化題庫運作之探討。國家菁英季刊，5(4)，149-166。
- 彭錦鵬 (2005)。公務人員考試方法技術檢討與改進之研究。「94年度考選制度研討會系列六『公務人員考試法制與人才選拔研討會』」論文。台北：考選部、國立臺北大學公共行政暨政策學系。
- 彭錦鵬 (2009)。考選制度的觀念革新—以簡併考試類科及考試及格人員地方歷練為例。國家菁英季刊，5(1)，49-67。
- 彭錦鵬 (2010)。公務人員考選制度的變革與未來展望。國家菁英季刊，6(1)，

17-40。

楊戊龍 (2005)。專技人員考試種類認定政策變遷之研究。國家菁英季刊，1 (3)，頁 83-98。

楊戊龍 (2010)。從職能評估角度談考試方式及應試科目設計。國家菁英季刊，6 (3)，37-55。

詹靜芬 (2010)。論公務人員之倫理養成途徑。國家菁英季刊，6 (2)，19-36。

熊忠勇 (2010)。從職能評估角度談考試方式及應試科目設計。國家菁英季刊，6 (3)，75-90。

劉奕權 (2007)。當前補習教育現況及其對國家考試之影響。國家菁英季刊，3 (1)，頁 47-64。

蔡秀涓 (2010)。從職能評估角度談考試方式及應試科目設計。國家菁英季刊，6 (3)，57-74。

鄭夙珍、林邦傑 (2009)。心理測驗納入國家考試之可行性探討。國家菁英季刊，4 (5)，73-84。

劉坤億 (2009)。從多元角度檢討分階段考試之利弊得失。國家菁英季刊，5 (1)，35-47。

賴其萬、曾慧敏 (2005)。從美國醫師考試制度談我國醫師考試之改進。國家菁英季刊，1 (3)，頁 167-180。

英文部分

Anastasi, A., & Urbina S. (1997). *Psychological testing* (7th ed.). New York: Macmillan.

Asher, J. J., & Sciarino, J. A. (1974). Realistic work samples: A review. *Personnel Psychology*, 27, 519-533.

Borman, W. C., Hanson, M., & Hedge, J. (1997). Personnel selection. *Annual Review Psychology*, 48, 299-337.

Boyatzis, R. E. (1982). *The Competence Manager: A Model for Effective Performance*. New York: John Wiley & Sons.

Brigham C. C. (1923). *A Study of American Intelligence*. Princeton: Princeton University Press.

Cascio, W. F. (1976). Turnover, biographical data, and fair employment practice. *Journal of Applied Psychology*, 61, 576-580.

Drakeley, R. J., Herriot, P., & Jones, A. (1988). Biographical data, training success and turnover. *Journal of Occupational Psychology*, 61, 145-152.

Ghiselli, E. E. (1966). *The validity of occupational aptitude tests*. New York: John Wiley.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, 26, 461-477.

Gregory, R. J. (2007). *Psychological testing: History, principles, and applications* (5th ed.). Boston: Allyn and Bacon.

Maurer, S. D., & Fay, C. (1988). Effect of situational interviews, conventional structured interviews, and training on interview rating agreement: An experimental analysis. *Personnel Psychology*, 41, 329-344.

Mayfield, E. C. (1964). The selection interview: A re-evaluation of published research. *Personnel Psychology*, 17, 239-260.

- Muchinsky, P. (2003). Psychology applied to work: An introduction to industrial and organizational psychology (7th ed.). Belmont, CA: Wadsworth.
- O'Bannon, R. M., Goldinger, L. A., & Appleby, G. S. (1989). Honesty and integrity testing. Atlanta, GA: Applied Information Resources.
- Outtz, J. (2002). The role of cognitive ability tests in employment selection. *Human Performance*, 15, 161-172.
- Robertson, I., & Smith, M. (2001). Personnel selection. *Journal of Occupational and Organizational Psychology*, 74, 441-472.
- Schmitt, N., & Robertson, I. (1990). Personnel selection. *Annual Review of Psychology*, 41, 289-320.
- Spencer, L. M., Jr. & Spencer, S. M. (1993). *Competence at work: Model for Superior Performance*. New York: John Wiley and Sons.
- Sternberg, R. J. (1985). *A triarchic theory of human intelligence*. New York: Cambridge University Press.
- Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin*, 63, 100-116.
- Wagner, R. (1949). The employment interview: A critical review. *Personnel Psychology*, 2, 17-46.