

出國報告

2012 年第八屆國際測驗年會報告

服務機關：考選部

姓名職稱：黃專門委員立賢、張專門委員麗雪

派赴國家：荷蘭

出國日期：101 年 6 月 30 日至 7 月 7 日

報告日期：101 年 10 月

目錄

壹、前言	1
一、參加緣起	2
二、行程紀要	3
三、2012年國際測驗委員會年會紀要	4
貳、實施電腦化測驗的指導準則	6
一、適當關注科技的議題	6
二、增進電腦或測驗/網路測驗的品質	10
三、對於電腦化/網路測驗提供適當的控制水準	15
四、對於電腦化/網路測驗提供適當的安全性和隱私防護的措施	18
參、建置測驗評量中心	22
一、歐美測驗評量中心建構的特色	23
二、建構測驗評量中心各階段的重點	24
三、科技對於測驗評量中心的影響	26
四、測驗評量中心的訓練議題	27
五、運用測驗評量中心所得資訊做決定	30
六、倫理、專業和合法的考量	32
七、監控評量結果	33
八、測驗評量中心完整的政策說明	34
肆、測驗的編譯—測驗國際化的挑戰	36
一、測驗翻譯或編譯	36
二、國際測驗委員會測驗編譯之指導原則	38
三、歐盟機構人事考選制度經驗分享	41
伍、測驗資訊科技運用的進展	48

一、電腦化測驗的挑戰	48
二、電腦測驗的創新--21 世紀技能與評量方式的 改變	50
陸、結語與建議	59
參考文獻	66
附錄一 第八屆國際測驗年會會議主題與議程.....	71
附錄二 第八屆國際測驗年會照片.....	86

表目錄

表 1-1 會議行程表.....	3
表 3-1 各項評量方法活動彙整表	23
表 3-2 測驗評量中心執行各階段工作彙整表	25
表 4-1 歐洲人事遴選辦公室考選人力之機構及其職員數	42
表 4-2 歐盟機構人力考選內容及時程	45
表 4-3 評鑑中心所評量的測驗項目及欲評量之核心職能	46

圖目錄

圖 5-1	Primum 臨床個案模擬測驗電腦測驗畫面	53
圖 5-2	世界數學測驗試題樣本	54
圖 5-3	虛擬表現評量計畫測驗畫面	55
圖 5-4	Genetics Lab 電腦問題解決模擬情境式智力測驗試 題架構	56
圖 5-5	Genetics Lab 電腦問題解決模擬情境式智力測驗 畫面	56

摘要

2012 年第八屆國際測驗年會於荷蘭舉行，本部派員參加主辦單位阿姆斯特丹大學所舉辦的工作坊及各項學術論文研討會，藉以獲取先進國家發展科學化測驗工具的經驗，以改進我國人才考選與精進評量方法。本次年會有五項議題，分別為一、測驗與數位科技；二、跨國測驗與異質測驗族群之評量；三、測驗理論的發展；四、測驗之專業性、指導方針與訓練；五、測驗相關政策及道德議題。

本報告就此次年會所參與之研討會內容選擇與我國辦理國家考試較為相關之議題，分別就實施電腦化測驗的指導準則、建置測驗評量中心、測驗的編譯—測驗國際化的挑戰、測驗資訊科技運用的進展，彙整所見心得及相關文獻，提出各項建議。各項建議如下：

一：電腦化測驗使用多媒體優勢，檢驗特殊效度；二、新增電腦化測驗科技功能，保持系統彈性；三、編寫發展電腦化測驗參考手冊；四、透過比對系統處理試題曝光議題；五、利用電腦化測驗蒐集的各項作答樣態與統計數字，提升未來試題品質，也可提供應考人回饋；六、釐清測驗評量專責機構的人員角色；七、在試題與評量程序中，加入模擬測驗的概念；八、加強核心職能、工作分析、應試科目與評量方法的連結；九、強化測驗結果解釋的能力；十、善用統計資料擴大決定之層次與精確性；十一、視考試性質，評估將具評鑑效度之考畢試題重新列入題庫運用之可行性；十二、驗證電腦測驗應用的介面影響效度。

關鍵字：國際測驗年會、國家考試、電腦化測驗、測驗評量中心

壹、前言

一、參加緣起

測驗是對一個人的知識、技術、能力、興趣、態度或其他某些特質進行評量的過程，在教育歷程、醫學臨床等職業能力之測試及職場僱用、升遷評鑑上，測驗的運用已然成為常態。為評量政府人員進用的考選制度歷史淵源長久，中國的科舉是最早大規模應用測驗進行人才考選的制度，也是一項對於社會發展具有長遠影響的政府人事政策，雖評量內容較為單一，或者隨著歷代文學與學術發展，應試科目有所改變，惟已有當代測驗評量的雛型。我國的國家考試大多採取筆試作為評量工具，大多偏向「分科成就測驗」，是否可有效區辨應考人的專業知能，而及格者所具備的核心職能是否又能完全符合用人機關的需求，測驗分數與未來工作的表現是否具有正相關，均值得進一步探討。測驗工具有選擇決定與安置決定的功能，為達到適才適所的理想，不斷提升測驗工具的信度、效度以及採用多元評量方法應為考選部的施政主軸。他山之石，可以攻錯，汲取先進國家發展科學化測驗工具的經驗，並考量我國的文化適應性後，將有助於改進我國人才考選與精進評量方法。

國際測驗委員會（International Test Commission）係由各國心理學會、測驗委員會及相關測驗機構所組成的非營利組織，主要功能在於推廣及評估測驗政策之制定、研發與實施。該會成立於1978年，總部設在英國，目前有20個正式會員（國家及專業心理協會）與46個副會員（研究機構、測驗出版商與其他測驗委員會）和超過150位個人會員（學生及從事測驗工作之個人）。約在10多年前，開始在世界各國輪流舉辦年會，每兩年辦理一次，提供國際性的平台，讓各國學者發表其研究成果，也提供發展測驗公司介紹研發測驗工具之機會。

本部為強化國際參與及擷取先進國家發展測驗工具之經驗，於2007年10月以副會員身分加入該會，並自2008年第6屆年會開始派員參加，

藉以瞭解國際間測驗理論與實務的研究發展趨勢，進而思考測驗科學應用於國家考試的可能性。2010年第7屆年會於2010年7月18日至21日在香港中文大學舉行，本部亦派員參加。

2012年第8屆年會於2012年7月2日至5日在荷蘭阿姆斯特丹市舉行，由荷蘭阿姆斯特丹大學心理學系主辦，共計有來自全球57個國家之測驗、教育機構及政府機關逾500名之專家、學者參與，國內學者與研究生參與者也不在少數。會議以測驗工作坊、研究論文口頭發表及專題講演方式進行。本報告謹就參加本次年會所見所聞，選擇部分與我國國家考試有關之重要主題，除說明其主要內容外，並提出心得與建議，以作為未來改進國家考試測驗工具與評量方法之參考。

第9屆國際測驗年會預定於2014年7月2日至5日在西班牙聖賽巴斯提安（San Sebastian）舉行¹，暫定討論主要議題為「評量實務典範的全球與區域性挑戰」，子議題為：（一）國際化測驗的挑戰與機會；（二）測驗的新科技與新測驗模型應用；（三）測驗實務典範的準則與方針；（四）多元語言及多元文化脈絡中之測驗。

¹詳細資料請參考：<http://www.itc2014ss.com/>

二、行程紀要

會議行程表

日期	時間、交通資訊	地點	會議及活動
6/30 (六)	22:45 中華航空 CI65	台北 至	啟程
7/1 (日)	9:15	阿姆斯特丹	
7/2 (一)	8:00-12:30 13:30-17:00	阿姆斯特丹 熱帶學研究 所	工作坊 2 工作坊 5 工作坊 7
7/3 (二)	8:00-12:30 13:30-17:15	阿姆斯特丹 熱帶學研究 所	1. 報到暨開幕式 2. 各場次專題研討會
7/4 (三)	8:45-12:30 13:45-17:15	阿姆斯特丹 熱帶學研究 所	各場次專題研討會
7/5 (四)	8:15-12:00 13:15-14:30	阿姆斯特丹 熱帶學研究 所	1. 各場次專題研討會 2. 閉幕
7/6 (五)	14:05 中華航空 CI66	阿姆斯特丹 至	返程
7/7 (六)	13:05	台北	

2012 年國際測驗委員會年會由本部題庫管理處黃專門委員立賢與資訊管理處張專門委員麗雪共同前往參與，除於會議召開前參加「測驗工作坊」外，因大會發表之研究論文超過 200 篇，專題演講也超過 30 場，僅能選擇與本部業務有關之議題參加。

三、2012 年國際測驗委員會年會紀要

第 8 屆國際測驗委員會年會於 2012 年 7 月 2 日至 5 日於荷蘭阿姆斯特丹市皇家熱帶學研究所舉行（原訂於阿姆斯特丹大學舉行，因該校預定舉辦會議場地於一年前遭逢祝融，未及於本次會議舉行前完成修復，由該校洽借皇家熱帶學研究所作為年會會場），由荷蘭阿姆斯特丹大學心理學系主辦，計有來自全球 57 個國家、超過 500 位學者、測驗公司人員和政府機構代表參加，活動內容分為兩部分，一為 8 場次的「測驗工作坊」，本部人員選擇參加之場次為：「評量之技能：心理學者及研究人員的基本概念及技能」、「現代臨床測驗與評量之心理評量原則應用 (Measurement matters: Applying psychological measurement principles in modern clinical testing and assessment)」、「測驗編譯：測驗等化的追求」；二為專題研討會，進行的形式包括專題演講、論文發表與研討、海報展示與詢答。本次年會的主題為「評量的現代進展：測驗與數位化科技、政策及指導方針 (Modern Advances in Assessment: Testing and Digital Technology, Policies, and Guidelines)」，會議共有五項議題如下：

- (一) 測驗與數位科技。
- (二) 跨國測驗與異質測驗族群之評量。
- (三) 測驗理論的發展。
- (四) 測驗之專業性、指導方針與訓練。
- (五) 測驗相關政策及道德議題。

在 3 天會議中，無論是專題演講或是論文研討會，與會人員就測驗領域在理論及實務運用上的發展，提出許多研究方向及經驗分享。惟會議期間每一時段同時進行 7 至 8 場次研討會，每一場次研討會發表 4 至 6 篇研究報告，報告均採口頭方式發表，部分研究內容及結論尚在發展階段；又本次會議場地多屬小型研討室，常發生因參與人數過多無法容納之情形致造成向隅；另因國情不同，國外編製測驗通常有大型測驗公

司的參與，測驗學者則負責開發與評估使用成效，我國則由政府機關辦理，僅有少部分引進測驗學者進行研究。本研究將就本次年會所參與之研討會內容與我國辦理國家考試較為相關之議題，分別就實施電腦化測驗的指導準則、建置測驗評量中心、測驗的編譯—測驗國際化的挑戰、測驗資訊科技運用的進展，彙整所見心得及相關文獻，提出各項建議。

貳、實施電腦化測驗的指導準則

國際測驗委員會致力於測驗發展的國際化、科學化與標準化，因此對於發展「測驗指導準則」(test guidelines)的議題在近十年來，一直努力進行各種研究與擬定執行的標準程序(The British Psychological Society, BPS, 2012)，在前一次年會曾有多篇論文與工作坊加以討論。本次年會則因先進國家電腦化測驗發展的經驗，又有許多大型測驗公司發展各種測驗評量的資料庫與軟體，有些測驗甚至利用網路進行即時性評量，對於電腦化測驗發展者、出版者以及使用者而言，均需要更明確的指導準則加以規範，以確保在技術與倫理層面的準確性與適當性。

近年來電腦科技快速進步，促使電腦化測驗或網路測驗面臨一些問題的挑戰，例如行政管理的標準、測驗的安全、測驗過程與測驗結果控制。國際測驗委員會自 2005 年開始著手建立有關電腦與網路測驗的參考規範，主要目的有二，一是為電腦化測驗(Computer-based test)與網路傳送試題方式的測驗(Internet-delivered test)，在實務運用過程中，發展出廣為國際認可遵循的準則；二是增加所有使用電腦或網路相關測驗人員對於執行規範的共識(ITC, 2005)。考選部自民國 93 年開發電腦化測驗應用於航海人員國家考試以來，無論系統面或應考人作答介面都曾不斷地改良，在試題派送方面也從個別考場派送，進到由國家考場端統一派送，已具網路測驗之雛形。雖然目前電腦化測驗尚未採取先進國家電腦化適性測驗之評量方法與題庫建置方式，但有關以電腦或網路進行測驗的形式，國際測驗委員會所建立的指導準則仍值得我們參考。以下的內容參考規範將分別從測驗發展者、測驗出版者與測驗使用者的角度分別闡述。

一、適當關注科技的議題

(一) 重視電腦硬體與軟體設備

1. 測驗發展者

- (1) 對於電腦化測驗的軟、硬體設備提供一個最低設備標準，利用網路進行測驗時則須具體指明瀏覽器的版本與規格。
- (2) 使用理想的傳送平台，以確保傳送和畫面呈現穩定性與一致性。
- (3) 使用適合的科技特徵強化其可用性，並且依循已建立的使用者圖解介面設計規格。譬如說，複雜的圖解和互動式的特徵可能降低軟體執行的速度或增加下載試題的時間。設計優良的試題以符合測驗目標與目的，至於使用多媒體特徵的試題則僅應用來對於特殊效度考驗的情況 (Esther & Patrick, 2012)。
- (4) 設計網路傳送試題的系統，必須考量不同試題型態傳送所需的時間。
- (5) 考量從不同應試地點連結系統的速度。

2. 測驗出版者

- (1) 從測驗編製過程的文件核對軟、硬體與瀏覽器的最低規格以確保測驗使用者的使用無礙。
- (2) 確認試題傳送平台可支持測驗系統運作順利。
- (3) 確保在軟、硬體的變動情況下，電腦化測驗仍然可以容易運作和維護。
- (4) 在測驗公開出版使用後，系統需保留彈性，讓新增特殊科技功能的試題能夠增加。

3. 測驗使用者

- (1) 確認使用者有足夠的軟、硬體電腦設備操作的知識與技能。
- (2) 確認系統在測驗使用過程中，可以證明其適用性。
- (3) 螢幕升級的改變在測驗使用者的作答過程中，應隨時配合軟、硬體及系統，充分告知使用者。
- (4) 改變系統或作答過程，則須讓測驗使用者了解此一衝擊程度以及如何應變。

(二) 考量電腦化測驗/網路測驗的各項設計完整性

1. 測驗發展者

- (1) 測試系統確保其完整性，使其能夠處理系統當機與操作者失誤時，仍然能夠在短時間修復，回復到正常使用狀態。
- (2) 當電腦化測驗/網路測驗系統測試時，須設計系統迅速回應指令並反映到螢幕上。
- (3) 若利用網路傳送試題進行測驗，須盡力縮小傳送阻礙、喪失網路連結以及下載過慢的衝擊。
- (4) 提供有關軟、硬體排除一般經常發生障礙的使用說明手冊。

2. 測驗出版者

- (1) 進行一系列適當的平台檢查，以確實增強系統的完整性與堅固性。
- (2) 對所有系統提供足夠的測試點，即使設備的連接點出現故障，各項進出的溝通訊息仍可暢行無阻。
- (3) 為避免測驗使用者產生錯誤引起測驗管理上的問題，應提供測驗使用者指導說明或一些聯繫的方法以有效排除故障。
- (4) 對於測驗使用者所需要與科技有關的文件，必須由足夠訓練的人員提供。

3. 測驗使用者

- (1) 開始進行電腦化測驗前，檢核提供完善測驗系統的證據與文件。
- (2) 在電腦化測驗的進行過程中，須特別重視如何回應問題的模式。
- (3) 使用測驗者在輸入訊息進入電腦中，須清楚地觀察反應與產生問題，以提供給測驗發展者與出版者。

(三) 使用電腦化或網路測驗呈現時須考量人的因素

電腦化測驗與網路測驗畢竟是提供給人使用，必須從使用者角度去考量，以下分別從不同角度探究：

1. 測驗發展者

- (1) 螢幕需有良好的解析度與色彩，螢幕擺放位置須顧及安全。
- (2) 螢幕呈現特殊重要訊息時須置於明顯位置，試題字母也須清楚端正，以便免擾亂應試者的思緒。

- (3) 提供測驗使用者有瀏覽試題以及回頭作答的功能，同時在展現試題時須有邏輯性與意義性。
- (4) 呈現測驗者的姓名、試題數量、試題的提醒語或是位於同一頁面中相同的位置給予提醒。
- (5) 在電腦化測驗螢幕上出現非鬧鈴式的時間顯示，如果程序執行錯誤也能清楚告知，也可讓應試者有改正錯誤的機會。

2. 測驗出版者

- (1) 電腦化測驗或網路測驗發展過程中須檢核螢幕介面設計與呈現，提供測驗發展者注意到細微問題。
- (2) 電腦化測驗的試題呈現需相當順暢無礙。
- (3) 當有需要時，適當檢核有意義的錯誤訊息。

3. 測驗使用者

- (1) 需熟悉螢幕介面設計的邏輯，以及確保使用系統的特色是適當及相容的。
 - (2) 熟悉電腦化測驗試題如何呈現以及應試者如何反應。
 - (3) 檢核出現操作錯誤的訊息，需使用非鈴聲提醒並且須告知如何排除。
- (四) 在電腦化測驗/網路測驗中提供協助、有用的資訊與練習的試題**

1. 測驗發展者

- (1) 提供清楚、正確和適當的書面或電子式文件協助，確保這些文件對於所有應試者是適當的。
- (2) 提供應試者在螢幕上容易得到指導說明，包括試題數量、時間限制、試題型態以及測驗程序。
- (3) 發展合適的個人式指導或練習試題，提供應試者機會熟悉電腦化測驗或網路測驗的操作。

2. 測驗出版者

- (1) 提供額外支持性服務給應試者，從電腦螢幕上就可以呈現，例如何時進入系統與離開系統。

- (2) 檢核電腦化測驗或網路測驗中適合的練習試題，所提出的指導說明或練習試題也無需全部看完才可回答正式試題。

3. 測驗使用者

- (1) 當有需要時，熟悉如何獲得操作電腦或測驗或網路測驗附加的訊息。
- (2) 提供應試者充足機會熟悉測驗的軟、硬體設備。
- (3) 告知應試者如何使用輸入設備回答練習試題，並從練習中得到回答正式試題的技巧。
- (4) 利用電腦設備蒐集應試者的反應回饋給測驗發展者，幫助他們了解應試者正向的經驗。

二、增進電腦或測驗/網路測驗的品質

(一) 確保使用電腦或測驗/網路測驗的知識、能力和適當的使用

1. 測驗發展者

- (1) 準備檢核清單測試與研究電腦化/網路測驗的運作模式是否依據測驗內容及充足技術以完成建構相關的測驗。
- (2) 確保所有測驗發展者，包括命題者、心理計量學家、軟體設計者都有充足的知識與能力發展電腦化/網路測驗。
- (3) 遵守有關電腦化/網路測驗合法、專業和倫理的作為和準則。
- (4) 在測驗與試題發展過程中，透過安全保密協定，保護測驗內容的安全。

2. 測驗出版者

- (1) 根據測驗目的所設計的測驗內容與提供足夠的技術，確保電腦化/網路測驗合用。
- (2) 提供測驗使用者清楚的指導說明，以利其正確存取和管理網路測驗資料。
- (3) 定期維護和更新與電腦化/網路測驗有關的文件，包括適當修正法規和政策。

(4) 了解網路測驗分數呈現與使用的限制性。

3. 測驗使用者

(1) 評估電腦化/網路測驗內容與技術的適當性，必須與測驗方法的選擇有關。告知應試者測驗目的，他們就可依據個人的情況做出較佳的決定。

(2) 為具備電腦化/網路測驗適當的知識與操作的模式，可增加適當的訓練課程以及閱讀有關的訓練材料。

(3) 基於專業倫理需告知應試者網路測驗的限制。

(4) 若進行網路測驗時，提供應試者聯繫的管道，例如電話和電子郵件信箱。

(二) 重視電腦化/網路測驗心理計量上的品質議題

1. 測驗發展者

(1) 電腦化/網路測驗的信度、效度與公平性的相關文件應予以保留及提供相關訊息。

(2) 確認目前最新心理計量標準，例如信度與效度，並且須注意應用於發展不同測驗時可能有所不同。

(3) 設計電腦化/網路測驗需注意測量的是應試者的專業知識與能力，而非其電腦技能。

(4) 在適性測驗的試題選擇和控制試題順序時，需要先描述其理論和實務上的方法。

(5) 試題內容改變時，重測和評估也必須改變。

2. 測驗出版者

(1) 提供電腦化/網路測驗適當的心理特質評估文件。

(2) 各類電腦化/網路測驗需要提供適合的心理計量證據。

(3) 各項測驗建構模式對測驗使用者而言都不相同，必須提供測驗使用者相關的解釋。

(4) 當測驗內容變動時，心理計量模式的適用狀況，必須重新確認與再

評估。

3. 測驗使用者

- (1) 合適的電腦化/網路測驗心理計量證據文件需先行確認。
- (2) 對於網路測驗而言，必須提供進行測驗的網站有效的心理計量測驗。
- (3) 使用適當的訓練計畫以及確認不間斷的專業發展。
- (4) 準備文件的資訊必須提供測驗試題或參數的改變情況，以及對測驗屬性之衝擊。

(三) 電腦化/網路測驗從紙筆測驗發展而來，須確保有相同效度的證據

1. 測驗發展者

- (1) 提供清楚證據的文件，說明介於電腦化/網路測驗和非電腦測驗有相同效度，在前述兩種不同形式的測驗應該比較效度、從信度評估中了解兩項不同形式測驗的相關，以及產出兩種測驗的平均數、標準差以比較其分數。
- (2) 對電腦化測驗或紙筆測驗形式，試題呈現的方法基本上是相等的。

2. 測驗出版者

- (1) 評估電腦化/網路測驗與紙筆測驗具有相同效度甚至更佳的證據，特別應從常模角度加以解釋。
- (2) 假如測驗發展者並沒有提供效度相同的證據，應該引入適當的效度研究加以證明。
- (3) 假如測驗發展者並沒有提供在不同情況下，測驗使用有效的相關證據，其他有關測驗等化或常態分配的情形也應詳盡說明。

3. 測驗使用者

- (1) 確認電腦化/網路測驗與紙筆測驗測驗效果相等的證據是否充足。
- (2) 假如常模來自於紙筆測驗，則進一步確認其平均數與標準差在其他測驗樣本上仍然相似。
- (3) 確認電腦化/網路測驗與紙筆測驗技術上的特徵，例如應試者的管控與試題的呈現，應接近於紙筆測驗。

(四) 電腦化/網路測驗計分和試題分析結果的正確性

1. 測驗發展者

- (1) 確保電腦化/網路測驗計分規則的正確性。
- (2) 提供有關電腦測驗效度和使用的詳細說明書。
- (3) 使用電腦化/網路測驗所做的計分報告須能將應試者的反應加以分類，例如「外向類型」、「具有銷售潛能」等等，完成計分手冊以便能完成以「電腦為基礎」的測驗解釋 (Computer-Based Test Interpretations, CBTI)。
- (4) 從特殊的分數或分數型態中敘述合理的電腦化測驗解釋。
- (5) 當測驗資料使用鍵盤輸入電腦時，必須設計檢核資料正確性的機制。

2. 測驗出版者

- (1) 在測驗使用前確認計分規則經過仔細的評估。
- (2) 告知測驗使用者有關電腦化/網路測驗的計分規則，例如不計分的試題、猜測試題答案的扣分方式。
- (3) 告知測驗使用者電腦化的測驗解釋是來自於那一種效度方法的解釋。

3. 測驗使用者

- (1) 檢查和了解電腦化/網路測驗計分的規則。
- (2) 在適當的時機告知應試者分數是如何產生的。
- (3) 了解電腦化測驗解釋的原理、方法與限制。
- (4) 確認以藉由鍵盤輸入電腦資料的正確性。

(五) 適當的解釋結果以及提供適當的回饋

1. 測驗發展者

- (1) 闡明電腦式的測驗解釋對於電腦化/網路測驗可能的限制。
- (2) 設計與確認個人化的電腦測驗報告原型給所有參與測驗過程的人員了解。
- (3) 闡明如何獲得不同的報告，特別需要考量媒介 (文字或圖形)、報告

複雜度、報告的架構、測驗的目的、可變動的程度、報告的形式和形態。

- (4) 提供回饋時須考量適當的指導，包括解釋電腦化測驗報告所需的訓練。

2. 測驗出版者

- (1) 使用電腦化測驗報告時特別需注意一般性資訊而非特殊測驗目的，例如對特殊的個體；基於測驗分數所做的解釋，係來自於輸入的資料所得到，但其它補充的資料雖然重要但不能列入考慮；對於開放式或控制性模式的網路測驗，應試者可能接受測驗係在非標準化、無人監督或多變的情況下。然而，分數的解釋應該是在受監督管理的情況下較為理想。
- (2) 評估電腦式測驗報告在電腦化/網路測驗系統的合適性，則需注意報告效度和實用性的證據、報告的涵蓋範圍、相類似資料在報告中的一致性、對於想閱讀報告者的可接受性、對於測驗使用者的時間、價值和長度以及系統誤差的自由度。
- (3) 給予測驗使用者如何有效地與應試者、其他有關的人分享電腦式測驗報告。
- (4) 告知測驗使用者與提供應試者電腦式測驗報告有關倫理與實務議題的資訊。

3. 測驗使用者

- (1) 當解釋電腦化測驗報告結果時，必須了解潛在的限制，例如分數解釋需要受到監督管理、在標準化狀況下的測驗必須在公開與控制情況下驗證效度；使用此種標準化測驗可以確保應試者真正的能力。
- (2) 對於應試者與想閱讀報告的人，選擇和使用最適當的電腦化測驗報告樣態。
- (3) 若有可能的話，編輯電腦化測驗報告時的相關資訊，應該須包括應試者的背景、行為、能力、態度和人格。

- (4) 確保網路測驗的報告須符合可理解性與有意義性的形式。
- (5) 考量對於使用網路得到回饋的倫理議題，有幾項值得注意的是提供負面回饋給應試者，卻乏應試者作答情緒狀態的知識；提供立即性回饋卻產生負面效應的困難。

三、對於電腦化/網路測驗提供適當的控制水準

(一) 詳述測驗狀況的控制水準

1. 測驗發展者

- (1) 證明提供的軟體、硬體以及所有程序上所需的設備有助於電腦化/網路測驗的管理。
- (2) 對於電腦化/網路測驗適當管理所需要的測驗狀況加以清楚描述。
- (3) 設計電腦化/網路測驗能與國家有關使用電腦的健康、安全、合法之一致性規則相容。

2. 測驗出版者

- (1) 當採取網路做為測驗媒介時，即需模擬應試者參加測驗的各種狀況。
- (2) 在電腦化/網路測驗建置過程中，必須告知測驗使用者需要考量健康與安全要素，例如網路測驗進行如果太長將會傷害健康。

3. 測驗使用者

- (1) 當管理測驗時，確保軟體、硬體以及各項週邊設備都能在測驗前順利運作，各項操作的細節也需記錄在操作手冊中。
- (2) 當電腦化/網路測驗在一個特定的測驗中心舉行時，希望使用測驗者能夠在該地點非常舒適。例如舒服的座位、所有的鍵盤和控制開關都能操作自如、有足夠的測驗空間。
- (3) 當藉由網路進行測驗時，提供測驗使用者參加測驗的具體且詳盡說明。
- (4) 對於參加電腦化/網路測驗需要較長時間時，需要安排適當的休息時間。

(二) 詳細描述電腦化/網路測驗適當的控制監督

1. 測驗發展者

- (1) 提供電腦化/網路測驗所需監督的證明文件，但必須是開放模式而非直接的個人監督模式；控制的模式內容必須讓應試者清楚了解；監督的方法應由應試者登錄並確認測驗被有效管理與正確完成所有試題的施測。
- (2) 提供電腦化/網路測驗設計的使用說明。

2. 測驗出版者

- (1) 陳述電腦化/網路測驗的監督程度。
- (2) 對於特定的測驗方案所建置的電腦化/網路測驗，詳細說明其功能與限制。

3. 測驗使用者

- (1) 對於電腦化/網路測驗管理者確認其監督層次。
- (2) 使用電腦化/網路測驗需在適當的測驗方案中才可能適當實施。

(三) 慎重考慮先前練習的控制和試題的曝光

1. 測驗發展者

- (1) 對於高風險的網路測驗需控制題庫試題的曝光率使其相等。
- (2) 在測驗進行中試題的分派須予以限制，將不需要曝光的試題比例降到最低。
- (3) 確定試題的數量夠大，足以在電腦適性化測驗中進行各項複本測驗，也可有效管理試題曝光比率。
- (4) 當複本測驗使用時，進行心理計量分析時需考量等化的問題。
- (5) 仔細考慮試題分派的議題，可能會受到測驗內容記憶的影響。
- (6) 若控制試題的曝光有其效果，將會降低應試者作弊的盛行率。

2. 測驗出版者

- (1) 若試題曝光受到控制，網路測驗可視為最大表現測驗。
- (2) 提供測驗使用者有效控制試題曝光的訓練。

- (3) 試題曝光控制得宜，提供應試者適當的練習，將不會危害試題的安全性。

3. 測驗使用者

- (1) 從測驗中對應試者說明清楚複本或多重測驗的等化問題。
- (2) 從先前的曝光試題保護電腦化/網路測驗，而非教導應試者熟悉真正的測驗內容。

(四) 考慮監控應試者的誠實或作弊

1. 測驗發展者

在電腦化/網路測驗系統中設計各項特色，例如進入作答路徑以及使用者的設定，都可使測驗出版者與使用者對於測驗系統有效管理與控制。

2. 測驗出版者

- (1) 基於實務操作面，詳細描述進入不同評量系統所需真實性的標準，當進入測驗系統時，無論是測驗使用者或是應試者都需要連續使用使用者姓名和密碼。
- (2) 對於普通或高風險測驗需要多階段的監控措施，提供資訊給測驗使用者降低應試者作弊的風險，進行測驗在開放或控制模式時，必須在受到測驗使用者監控的情形下進行連續性的效度測驗與比較分數。
- (3) 假使測驗不容易受到控制，對於測驗效度的威脅就會存在。
- (4) 沒有直接的人為的監控下，誠實和榮譽的應試政策在執行的過程中就需要設計更多的階段性程序加以確保。

3. 測驗使用者

- (1) 對於普通和高風險的測驗需要確認如何降低作弊的機會，例如電子式的作弊，需要測驗管理者有系統規劃各項防範措施。
- (2) 對於普通和高風險的測驗，例如各項工作招募和選擇人才，雖考量便利性，但測驗地點的安全性則更形重要。
- (3) 提供測驗使用者已經參加過各種防弊演練的監場人員，而且要他們

承諾會秉持公正執行測驗進行的監督工作。

四、對於電腦化/網路測驗提供適當的安全性和隱私防護的措施

(一) 考量測驗工具的安全性

1. 測驗發展者

- (1) 在電腦化/網路測驗中，設計一些能夠將試題曝光、計分關鍵、違法出版測驗解釋、下載及拷貝試題到另外一台電腦的風險降到最低的程式。譬如說軟體藉著目錄功能的選擇喪失控制瀏覽器的功能。
- (2) 在電腦化/網路測驗系統中設計一些防護措施，例如防火牆，以避免病毒或駭客入侵。

2. 測驗出版者

- (1) 電腦化/網路測驗系統中重要的智慧財產如計分規則、常模、計分解釋，必須儲存在主機伺服器上，只有測驗試題和輸出的結果出現在螢幕上。
- (2) 測驗使用者和測驗中心在電腦化/網路測驗使用的素材必須受到限制，例如當測驗以網路傳遞時，測驗使用者需要獲得密碼以取得應試者的作答資料或設定測驗進行的相關狀態。
- (3) 密碼只能授權給有資格的測驗使用者。
- (4) 確認和檢查電腦化/網路測驗避免駭客入侵和電腦病毒的機制；確認網路測驗避免未授權者及使用非法的方法入侵伺服器。
- (5) 確保網路測驗敏感性的內控機制，以防止資料被竊取拷貝，同時需監控網站是否有違法，以及相關防護過時的情形。
- (6) 採取保護測驗內容安全的步驟。
- (7) 採取適當的方法防止測驗報告通知書被竊取，以及對測驗計畫的衝擊。

3. 測驗使用者

- (1) 熟悉測驗發展過程中各項設計，降低未被授權者獲得測驗資料的方

法與程序。

- (2) 尊重電腦化/網路測驗發展者和出版者的智慧財產權。
- (3) 若沒有測驗版權者的書面同意，應保護電腦化/網路測驗資料不要被拷貝、印刷成書面及重製。
- (4) 測驗提供者與出版者應告知違反安全保護將會受到何種處罰。

(二) 考量應試者資料轉送到網路上的安全性

1. 測驗發展者

- (1) 當設計網路測驗時，需建構應試者安全應試的環境。
- (2) 確認另一台主機伺服器運作正常，而且所有的插座孔都配置妥當。
- (3) 設計良好的管理系統以確保使用者能夠存取、確認和刪除伺服器資料，也可保護伺服器中的資料不致破壞或竊取。
- (4) 對於所有蒐集應試者資料時應有備援系統，問題產生時也應有回復的功能。

2. 測驗出版者

- (1) 應試者資料藉由網路系統傳送時應予以加密。
- (2) 告知測驗使用者有關原始資料保護的權利與責任。
- (3) 管理蒐集資料的支援系統，應提供測驗使用者問題產生時詳細的回復計畫。

3. 測驗使用者

- (1) 對於管理電腦化/網路測驗最優先的要務，即為告知應試者在網路上安全的防護措施。
- (2) 確認電腦化/網路測驗的服務提供者可以提供備援資料。

(三) 保持應試者測驗結果機密性

1. 測驗發展者

- (1) 設計電腦化/網路測驗安全儲存資料在電腦伺服器與硬碟中，而且需要有安全的防護機制。
- (2) 藉由提供科技協助，將電腦化/網路測驗原始作答資料亂碼轉換成

可辨識的資訊結果時，具有完整的一致性。

(3) 設計加密措施和密碼以限制資料存取。

2. 測驗出版者

(1) 當測驗出版者需要儲存測驗資料時，具體說明資料安全的保護程序。

(2) 告知測驗使用者資料的存取方式與時限。

(3) 保護敏感性的個人資料，需在電腦硬碟或伺服器中得到堅實的保護。

3. 測驗使用者

(1) 當資料以電子形式儲存時，測驗使用者知道如何善加保存。

(2) 堅守國家訂定蒐集、使用和儲存個人化資料的各項法規。

(3) 當在測驗中心使用設備以電子形式儲存個人敏感性資料時，加密程序和使用密碼是必要的。

(4) 應用安全和值得信賴的備援系統儲存個人化資料。

小結

電腦化測驗/網路測驗所發展出來的國際通用指導準則，係從試題發展過程、電腦及網路的軟硬體設備、測驗評量方法的基礎、試題的曝光程度、資料保存與系統安全性，廣泛且細微地提出每個環節應該注意的地方。且從測驗發展者、測驗出版者與測驗使用者的角色分別羅列各種需要注意之處，雖然考選部在建構電腦化測驗試題、系統架構與作答介面時並沒有如此清楚的區隔，例如「測驗發展者的角色」主要由題庫管理處負責；而國外測驗出版者通常由測驗公司負責，但仔細檢視其內容，本部資訊管理處與軟、硬體設計廠商雖未負責出版，但工作內容近似；另國外對測驗使用者的角色定義通常在使用單位，且須有能力扮演監督者的角色，以確保測驗發展者與測驗出版者在編製試題與建構系統與作答環境均有科學化與完整性的設計。國外電腦化測驗/網路測驗建置的分工看似獨立，卻彼此需要密切合作與溝通，更需有檢查與回饋機制。

我國電腦化測驗發展歷史並不長，在軟、硬體的設計上也經過不少

摸索及漸進發展的過程。然而，實施電腦化測驗並非僅為作答介面的改變，而是藉由電腦科技的進步提高評量的信、效度，更容易管理試題、個人資料、確保安全以及對測驗結果進行完整的解釋。本章電腦化測驗/網路測驗指導準則可提供本部對於電腦化測驗對於多媒體試題設計、題庫建置、系統設計之彈性、試題曝光之處理、測驗結果之回饋完整的改進思考，相關具體建議將於第六章加以詳述。

參、建置測驗評量中心

此次赴荷蘭參加國際測驗年會，對於荷蘭測驗公司 CITO 總裁介紹電腦模擬測驗應用在醫學評量的印象深刻，尤其他的引言曾提到最早的測驗來自於中國隋唐時代的科舉制度，簡報中也有科舉制度的圖片，但測驗科學化也不過是上個世紀初的事，至今也不過 100 多年，仍有繼續進步的空間。在各國專家學者發表的論文中，有一個焦點是集中在介紹各國測驗發展的趨勢與文獻的回顧，例如美國、歐盟、英國、西班牙、德國等，在此架構下，對於各先進國家測驗評量中心的建置著墨甚多，主要角色在於扮演測驗發展者，雖非具有完全的官方色彩，但對各類人員甄選與測驗的發展，卻具有指導性與改進的功能。

測驗評量發展中心對於各專業工作領域人員的甄選確實發揮了功能，且得到廣泛的肯定，因為他們可以為用人單位特殊的需求透過能力與行為的確認，選擇具有發展潛力的應徵者。優良的測驗評量中心有以下的優點 (Psychological Testing Centre, 2012)：例如對於人員甄選具有綜合性的判斷資訊；對於工作計畫具有有效的決定力；藉由多重判斷增加選才公平性；有效覺察工作角色的核心職能；從組織運作過程中提升選才判斷效度；對於選才系統具有合法的地位；最重要的是可以預測工作表現。

以下對於各國測驗評量中心的創建與運作加以介紹，主要探討的內容包括藉由不同專家和行政人員的合作，藉由過去發展的經驗，如何做好人力甄選與管理。本部職司全國公務人員與專門職業及技術人員甄選，雖多數同仁屬人事行政職系，惟對於科學化的甄選與測驗評量仍較為陌生，又因人力受限，對於評量方法與試題研究仍受到極大的限制。爰此，目前有成立測驗評量專責機構研究之議。以下將就此次研討會所發表的各國發展測驗評量機構的經驗，再加上文獻的蒐集，綜合將相關的議題與運作經驗、方法與程序說明如後 (Geisinger, 2012; Lindley & Bartram, 2012; Carlson & Geisinger, 2012; Elosua & Iliescu, 2012;

Hagemeister, Kerstling, & Stemmler, 2012)。

一、歐美測驗評量中心建構的特色

(一) 測驗評量中心的特色

歐美測驗評量機構有一些共同的特色，例如有多重評量的過程，可藉由科技蒐集各種資訊確認所需特質；設計候選者或參與者 (candidate/participant) 共同參與甄選活動的團體，通常需要將各種甄選方法加以整合，最常用的是模擬測驗，希望透過真實情境的模擬了解應試者的真實能力 (詳如表 3-1)；藉由經訓練的評量人員觀察應試者與工作相關的行為，尤其是了解從事工作所需的核心職能，以及分析優秀工作者與表現不佳工作者之間的差別在那裡，尤其優秀者在性格、態度、特殊的工作技能隨時間過去有所成長；做各種甄選決定時需要運用各種資訊，在不同情況下各項資訊對評估者而言是相當有意義的，各項資訊或計分可以反映出應試者的優缺點。

表 3-1 各項評量方法活動彙整表

活動	描述
表達	對一群相關的團體進行模擬的簡報
團體討論	給予資訊進行團體互動的討論
一對一角色扮演	一對一的互動進行溝通與協商
籃中演練(e化)	從預先設計好的試題模擬角色以了解其反應適當性
寫作分析	對於與工作相關的議題進行寫作分析
面談	藉由結構化口試蒐集更多的資訊
心理特質評估	經由標準化測驗例如認知、人格、動機或興趣測驗深入了解受試者

(二) 測驗評量中心合格的標準

設立測驗評量中心有各種不同的設計與考量，要符合水準並發揮其功能，應該遵循以下的標準：

1. 對於工作職位所需招募的人員，必須在事前先有清楚的工作分析，以確認能力、特殊工作表現與工作分析之間的連結。
2. 確認測量能力的測驗方式是值得信賴的，即使重複測量，仍然具有相當高的可信度。
3. 評量受試者時至少需要兩種的模擬測驗。
4. 在評量活動的安排上彼此應該有清楚的區別。
5. 評量受試者的人員在進行觀察、紀錄、分類與評估時應該受過足夠的訓練。
6. 評量者在獨立完成各項評估工作時，需要具有報告撰寫的能力。
7. 在評量完成的過程中需要有整合性的結論，也能夠摘要性將獲得的行為資訊清楚地說明所代表的意義。
8. 各項測驗的回饋資料應協助給受試者以提供他們未來的成長。
9. 測驗評量發展機構須將各種測驗計畫與結果做完整的報告與出版，說明資料是如何獲得計分，對於其他人獲取資料的權利為何。
10. 對於評量方法所進行的各項活動限制應詳加敘述。

二、建構測驗評量中心各階段的重點

歐美先進國家測驗評量發展機構的建立，多數成立之初都是屬於非營利型組織，有些早期從設有測驗、心理與統計科系的大學附設機構發展開始，如美國的 BUROS。後來多數的測驗評量中心自行獨立，除從事各項學術性研究外，營利性的測驗公司紛紛成立，與測驗評量中心更強化了彼此的連結，雖然測驗公司開始也聘用相關研究人員，但為求所得評量結果有更科學化的理論依據，也經常會和此類研究機構合作。測驗評量發展機構能夠歷久不衰，需將組織目標明確化外，發展過程中各項組織重要的配套措施也須確保持續進行，在執行成果的發表時，也能將各項測驗評量發展的重點說明清楚。以下將測驗評量機構執行各項工作分為計畫前、發展過程中、真正的運作執行以及執行業務後的檢討，說

明各階段的重點，整理為表 3-2 (PTC, 2012)：

表 3-2 測驗評量中心執行各階段工作彙整表

階段一：執行前的計畫	
1. 確認需求	建立一個組織有共識的需求。
2. 組織承諾	與所有相關的成員，包括委員會成員、管理階層、評量者在執行過程中有共同的承諾。
3. 組織目標	建立在運作過程中清楚的目標，例如在評量方法、選擇決定、組織效能提升和發展。
4. 建立政策	為測驗評量中心擬定對於組織的發展政策。
階段二：建置過程中的發展	
1. 引導工作分析	使用精確的工作分析技巧，形成有關清楚分辨引導能力或行為的方法
2. 使用模擬測驗	使用工作分析的結果和進階的調查，確認和提出有關用人機關所需適當的模擬測驗活動。
3. 設計過程	建構整合評量明確能力表現程度的活動。
4. 設計各項出版物	準備各項出版物格式、時間表和後勤的各項支援。
5. 實施訓練	對於參與測驗評量中心的所有人（評量者、觀察者、角色扮演者和設計者）設計和執行訓練計畫。
階段三：執行業務的過程	
1. 實驗性作法/不斷的改進作法	假如可能的話，選擇不同需要評量的個人，確認各項評量方法依據時間表操作後的有效性與公平性。
2. 機構的運作	測驗評量中心在執行受試者或參與者的各項評估活動，需要包含過程中品質的檢核。
階段四：執行業務後的檢討	
1. 做決定	依據測驗評量中心的各項評估結果做決定。
2. 提供回饋	依據組織的需求，對於所有的受試者或參與者提供回饋。
3. 監控與追蹤	對於測驗評量中心的發展，可以設定瀏覽和監控的程序，包括對於評量與真正實際工作表現的相關。

三、科技對於測驗評量中心的影響

網路與其他科技的進步正在挑戰目前測驗評量機構對於評量方法與過程的方法，重要科技方法的應用可以用來管理行政上的負擔，進而對試題的反應與計分自動化。電腦科技應用在現代測驗評量機構除了使行政的負擔減少，還可傳送各項評量的活動，與傳統的評量方法一樣具有品質也符合倫理的標準。使用電腦設計各項評量活動除了可以容易複製工作上的環境，也更可以強調工作效度，但系統需將受試者的需求說明清楚，需要展現的能力為何，才能透過軟體的設計作為評量的工具。

自動化的計分系統在速度與信度上有較佳的優勢，經常使用在一般預測性的測驗中。計分支援系統可使評量者評分外，還可表現出受試者的反應型態，以下再深入說明使用科技在評量機構中的各項作為。

（一）工作分析

許多利用電腦提升工作分析效能的系統，對於能力的評估比傳統的方式更具優勢，優點如下（Markus, 2012）：

1. 電腦系統對於工作可提供一個平衡的觀點，將工作所需的能力與行為做適當的研究與充分的了解。
2. 以電子形式蒐集的資料，可以降低行政管理對於大規模樣本蒐集的負擔。
3. 以電子形式進行資料存檔，將使得資料更容易儲存與尋找。
4. 有效的電腦科技進行工作分析所得的結果，係依賴對於熟悉工作內容的人員以及反應所得。

（二）模擬測驗的電腦管理

電腦在近年來已漸漸使用其多媒體的優勢對模擬測驗進行計畫、展示與管理，但與評量效度無關的心理計量內容也無須置入，重要的內容應該包括與工作能力有關者、評量內容容易了解、對所有人具有同樣的公平性、容易預測未來的表現。

（三）記錄應試者表現的證據

評量者可能從使用電腦科技中獲益，工作行為的清單和需要特別注意的地方也可以使用筆記型電腦存取重要的評量過程與結果。

(四) 受試者反應的評量

電腦在評量過程的某些層面上有能力表現的很好，只要受試者的反應能夠進入電腦中，也可以被解釋。

(五) 報告的撰寫

對於測驗評量中心所完成的報告書，在給予應試者回饋或為某些目的做決定時，其實相當耗費時間和資源。以電腦為基礎的專家系統、陳述行為的清單、完善的設備都是提升可控制部分與降低評量困難度的方法。在完成報告書時特別要注意有意義的細節必須特別摘錄，而且對於特別分數所代表的意義或不尋常的分數剖面圖，都需要再三確認產出結果與解釋的適當性。

(六) 虛擬的測驗評量中心

虛擬的測驗評量中心藉由科技的模擬測驗正在起步中，若要有更大的進展，進步的科技設備可以讓評量者和受試者溝通無礙，也需建構安靜的標準化與保密環境。

(七) 利用新科技的潛在問題

雖然藉由新科技更容易評量，下列問題也必須注意：

1. 受試者較喜歡面對面的互動。
2. 藉由新科技評量可能會降低社會互動性的過程。
3. 組織非個人化的印象將被傳遞。
4. 有一些評量過程例如團體活動並不容易藉由科技實施。
5. 某些心理計量性的特質需要進一步的調查。

四、測驗評量中心的訓練議題

在測驗評量中心內有許多人員的角色需要進行訓練，包括評量者/觀察者、管理階層者、角色扮演者以及評量設計者。但前述的角色在實際上有時也不容易分辨清楚，例如評量者可能也是角色扮演者，但是不同的角色

需有不同的訓練。以下說明這些角色的內涵與訓練重點：

（一）評量者/觀察者 (Assessors/observers)

評量者被賦予在活動中評量受試者行為表現的責任，訓練時需要考量測驗評量中心的工作方式或應遵循的準則、經常使用的評量工具、實際工作上有用的技能、可能會過時與亟需更新的技能、回饋技巧、組織運作的程序、公平的評量機會、保守測驗機密。評量者需要了解測驗評量中心重要的運作規則、機制、政策以及標準，訓練的焦點應清楚著重在經常使用的測量工具所進行的各項活動，尤其在各項新工具提出時，或在不同的測驗機構工作時都需要接受進一步的訓練。

評量者需要在過程中發展有關觀察、記錄、分類和蒐集證據的技能，以使得做決定更具有正確性。評量者需要能夠從應試者不同背景中有效評估行為，這對評量機構的公平性和有效性是相當重要的。在訓練過程中強調公平機會的評量，以及了解評量受到那些因素的影響。評量者需要從回饋中訓練評量技能，他們應該能夠解釋不同的行為表現導致不同的評量結果，應考人從回饋中接受最後的結果僅能視為最小的目標，進而希望能夠認同與支持他們未來性的發展。

最後，無論是最初的選擇，或另外提升一種角色的再分派過程，評量者對於接受測驗評量中心做決定的個體所提出的疑問應該能有所回應，而要處理這些問題，每年必要的在職訓練應持續進行。

（二）管理階層者（提升效能者）(Facilitators)

對於測驗評量中心的管理階層人員而言，在日復一日的運作中，至少需要滿足兩種角色的功能期待，一是「品質控制」；二是「設定時間表與機構內的管理」。他們需要了解測驗評量中心的標準，而且需要知道如何建立與更新，這包括了測驗對於測驗評量中心各項指導準則中各項合適的規定與實務運作。管理者也需要知道如何掌握時間表以確保測驗評量中心運作的順暢性，雖然時間表可能是測驗評量中心設計者確立的，但有時候在一些偶發事件上也須適時作出一些調整。管理階層者也必要被訓練以有效管

理測驗評量中心的場所，例如辦公場所的分配與安排，以及機構內人員如何彼此有效的聯繫，尤其是管理階層與評量者共同的訓練與對話，一年至少需要一次，而且如何保存評量資料的安全性與機密性更是他們重要的責任。

（三）角色扮演者（Role player）

測驗評量中心的角色扮演者主要是與受試者互動使其產生需要評量的行為，這經常發生在一對一的評量情境中。角色扮演者經常被訓練用來了解受試者一般性的表達，並引發出其典型的行為表現，他們也必須熟悉角色扮演時的各項活動與操作的方式。

角色扮演者也需要知道如何嚴格遵守評量前準備好的「劇本」，而且須謹慎小心。他們的訓練應該包括檢查標準的一致性，同時週期性的定期檢驗與更新。角色扮演者需要清楚界定他們傳遞資訊的方式和正確性，而非對受試者進行不夠嚴謹的推論。

（四）設計者（Designer）

測驗評量中心的设计者意指能將工作計畫和機構內工作內容整合的人員，他們通常是職業心理學家，設計者的訓練應該包括工作分析的方法、選擇適當的評量活動、測驗評量中心的時間管控、評量活動的報告撰寫。實際上，某些測驗評量中心將工作分析視為一項獨立的工作，可以支持受試者表現的證據，某些測驗評量中心又將所有的評量活動交給外部人員執行。如此一來，雖然降低設計者的負擔，但他們仍然應該了解工作分析的原則和如何撰寫評量活動的結果。工作分析的結果應該使設計者能辨認分數代表各種人員能力的組型，也可反映出有效的表現，要探尋這些有關工作的豐富資訊的方法可能包括問卷、焦點團體、內容分析法、口試、批判事件的技巧等。

選擇評量活動適當的訓練聚焦於工作分析時，究竟需要何種適配的活動？評估每一種能力使用的評量方法均有不同，設計者經常能夠對相關內容提出適當的問題，但對於不同的受試者應一視同仁（Enrique &

Manuel, 2012)。

測驗評量中心設計時間管控表，係對機構內的事項經過縝密的規劃，對於前述三種人員例如評量者、角色扮演者和管理階層者，也提供彼此合作與回饋的機制，所訓練計畫中應包括不同角色間可以看到彼此的工作圖表，若時間不足難有充分討論，也會影響各種評量成果的品質。另外在評量活動書面報告的書寫訓練上，應該包括各項指導說明使用以及評量方案的標準，其中方案標準可能隱含了原先團體討論的衝突或日常活動摩擦妥協的結果，這些對角色扮演者有額外的資訊，同時可涵蓋工作分析後對於評量活動採取類型與內容的選擇。

另外，在使用標準化心理測量工具時，以英國使用為例，使用評量者應該得到英國心理學會的認證，無論在計分與結果的解釋上，都應具有專業的水準。

五、運用測驗評量中心所得資訊做決定

決定過程的產出通常係依據測驗評量中心的目標而定，但並不能僅利用一種評量結果就希望得到最好的決定，需廣泛地評量受試者的各種行為，例如能力、態度、人格等。以下分別從評量中心做決定的角度以及使用其它資訊的方法再詳加說明做決定時特別重要的議題。

(一) 評量中心做決定

在評量中心內，透過工具蒐集所有的資料後，受試者接著必須被判斷是否成功？在完成評量前，所有評量需要獨立進行，評量者不可隨便下結論。基於工作分析或生涯決定所得到的判斷標準，需考量的因素除了重要性、經常出現的頻率、還有受試者未來的可發展性與培育性。另外做決定時也可使用某些標準進行加權計算，或藉由某些計分或計算規則，基於工作分析中的發現，結合可接受的最低評分標準與合用性。

有時候測驗評量中心的不同評量者，經常使用設計者相同的評量方法及活動，去應用在每一個受試者身上，在這種情況下，對於清楚確定評量

標準使其在應用過程中有一致性是相當重要的。而且幫助評量者維持與更新標準，從某一家測驗評量中心適用或轉移到另外一家測驗評量中心應該是相當容易的事，但仍須有品質管控的團隊經常去檢查測驗評量中心的標準。評量中心經常運用評量者所組成的團隊對多數的受試者進行評量，焦點即在於比較不同受試者間的表現，同時評量標準的一致性應該具體考量受試者未來的表現，也就是強調其預測效度。

測驗評量中心的管理階層經常需要依賴較資深評量者的評估報告做決策，使用工作分析資料，評量團隊較能從評量活動中得到一致性的共識 (Edurala & Carla, 2012)。評量者個別的分數評定有時常利用統計數字的結合取代決定的一致性，或者取代整體評量的過程，研究也建議此為一種相當有效的方法，統計的決定系統至少會被強而有力的內容效度及效標關聯效度所支持。假如前述的統計系統可以自動產生，對最後的選擇做決定，受試者也可在資料受到保護的情況下得到分數的合理解釋。

(二) 使用其它資訊

測驗評量中心經常使用蒐集資料所得的證據做決定，通常也有相當好的理由如此做，例如避免在相同的工作場所使用不同的標準、對於受試者正向或負向評量的確認、避免偏見和固著的觀念、工作設計的限制使其無法展現出某些行為。然而，做決定時或在評分的過程中，須適當地納入一些內在的資訊，內在資訊整合為最後的結果時，需考量這些資訊是否存在於所有的應試者、評量中心運用時對於能力評量的面向是否經過詳細計劃、確保效度的資料是否完整蒐集。

清楚整合內在資料的架構，須釐清屬於評量過程的一部分或是獨立存在，這將依據評量中心的目的而定。譬如說，評量的結果是為了應用在升遷上，主要的焦點是在於評量應試者發展他們工作技能的潛能上，評量活動須考量工作表現和受試者現在的情況，若是內部的資訊考量應用於評量中心以外的場合，則需要另行召開會議研商資料的整合、應用與解釋的方式。

六、倫理、專業和合法的考量

測驗評量中心因掌握了許多受試者的個人資料，因此須考量倫理、專業和合法的議題，以下分三部分加以說明：

(一) 確定倫理、專業和合法的議題都已完整顧及

評量中心進行各項評量活動時，在設計與執行時，倫理、專業性和合法性的議題應充分確認。倫理和專業議題下節將進一步闡明，而合法性指的是對受試者的平等性，以及資料受到完整的保護，不同的僱用行為會與測驗評量中心的目的（例如升遷或解僱）以及受試者的資料型態（內部或外部資料）有關。

(二) 受試者的資料議題

受試者參加評量中心前應先提供給他熟悉參加評量的目的、過程，以讓受試者決定是否參加評量，假如參加測驗評量中心的評量係屬僱用決定的重要部分，則需要告知受試者他們的權利。理想上，測驗評量中與受試者應該在測驗進行前三週進行，這些溝通包括：

1. 廣泛地了解測驗評量中心的測驗類型與模擬測驗進行的方式。
2. 提供有關評量人員的專業訓練、在評量過程中他們的角色以及如何做決定。
3. 對於評量的結果，測驗評量中心應安全地保存一段期間。
4. 確認提供受試者何時以及何種評量的回饋。
5. 確認誰有權力以及在何種情況下可以獲得測驗評量中心的報告。
6. 有關態度測驗的評量資訊應提供網路上的練習。
7. 假如評量的結果儲存在一個合法的電腦設備中，受試者應該可以要求給予合理的結果回饋，測驗評量中心也可給予一個適合的建議。
8. 對於回饋的形式而言，對內部受試者而言，面對面提供較為合適；但對於外部受試者而言，提供書面形式的回饋可能較為實際。
9. 在最短的時間內應該提供回饋，理想標準是在一個月內。
10. 若回饋可以做到告知受試者為什麼未被選取的原因則更為理想。

(三) 使用工具和資料的議題

對於測驗評量中心的工具更新和維護是相當重要的，所有的工具應該被妥善的保存和管制，只能對於被授權及受過相當訓練的評量者才可使用。而測驗資料的生命週期應視工作性質的發展與改變，或有特殊事件的介入而定，一般對於測驗資料的解釋效力大約在 12 個月到 24 個月內。另外對於資料的保護，在行政管理上，需注意誰可以獲取以及何時需予以銷毀，若測驗評量中心需要利用原始測驗資料做其它目的的運用，則需要得到受試者的同意。

七、監控評量結果

來自於測驗評量中心的評量結果應該加以監控，監控的原則應該就評量的進行次數、匯集成冊的資料，定期加以檢視。計劃加以監控時，檢視測驗評量中心所進行各項結果的檢視與確認有效性，應該從最初的階段開始，監控的議題應該包括剛開始的檢視、內容涵蓋面的適當性、公平和區辨度、資料蒐集和統計評估。

(一) 剛開始的檢視

測驗評量中心剛開始的檢視是否符合測驗的目的，評量的問題是否與工作情境相類似，假如他們遵守組織的標準，而且評量者也熟悉並遵循觀察、記錄、分類與評估的程序，一個獨立性的品質管控檢查就可以順利完成。

(二) 內容涵蓋面的適當性

涵蓋內容的適當性，首先指的是測驗評量中心對於每一種能力評量面向的基本設計是否周全。在實際運作上，目的就是將想要評量的能力予以揭露並評定層次；其次，紀錄的檢查將能夠揭露不同能力的差別。

(三) 公平性和區辨性

測驗評量中心要進行預測就必須提供明確的目標與真實的資訊，意即測驗評量中心能夠做出區別是因為基於能力的判斷而非既有的立場或觀

點。不同的表現或分數究竟與特定的種族、性別或其它群體的因素有何關係，假如測驗不能反映出潛能與表現的差異，可能就違反規定也不公平。測驗評量中心不同評量活動的表現可以依據受試者背景因素加以追蹤分析，對大的群體樣本或受試者而言，統計分析應該持續進行。無論多少人通過測驗評量中心的測驗，對於評量過程中的品質檢視是必要的，而且團體中的差異需經常檢驗，看看是否有其偏差存在，偏差的因素包括：

1. 過度強調因為性別或種族差異造成的測量誤差，譬如說，競爭性活動過度的採用，例如分派角色的團體討論，常常因為文化常會影響受試者的表現，也較不能為社會文化所接受。
2. 評量活動所偵測出的各項特質的混合，希望能夠描繪出受試者的不同能力，但有時只能測量出主要能力，忽略了某些特質。
3. 訓練過少使得公平評量的涵蓋面不足。
4. 對於身心障礙受試者未給予適當的處置。

假如受試者表現出內外混合的特質，對於評量者而言較重要的是內在的特質和表出現有知識面所代表的潛力。

（四）資料蒐集和統計評估

測驗評量中心有效的科學評量來自於與組織目標清楚的連接，這對於實證性證據幫助提升評量效度相當有助益。對於評量有效性應該應用下列的最低標準：

1. 執行評量的程序應該確認資料蒐集的正確性與真實性。
2. 評量在方法上應該盡可能嚴格和科學化，包括內容的質性分析、統計分析、受試者或評量者的態度調查。除此之外，進行實證性的效度研究，將測量結果與工作表現考驗其相關程度。

八、測驗評量中心完整的政策說明

測驗評量中心的人力資源策略很可能是要強調組織全面的有效性，也進而澄清組織的政策。總結言之，一個具有實效的測驗評量中心，完整的

建構應包括各項評量計畫的目標確認，界定參加各項評量的受試者或參與者特性，進行評量的專業人員或觀察者都須經過專業的訓練，熟悉各項評量方法覺察應試者之間的差異，為用人機關選擇最適合的人才。至於在評量過程中蒐集的各種資料應有良好的保存與利用方法，更能將這些資料透過量化與質化的整合做最好的解釋，提供給受試者回饋，同時藉由各項監控程序確保測驗品質。

小結

歐美先進國家設立測驗評量中心(機構)，主要目的是在於確保評量的公平性、精確性與科學性，並提供給用人機關資訊做決定；除此之外，對測驗評量結果的解釋與改進，對整體評量過程的監控以及研發各項評量方法更是機構主要工作任務。對於辦理考試機關而言，辦理考試的各項流程除重視形式公平性外，更需注重其內在公平性，而內在公平性指的就是評量的「信度」與「效度」。以目前考選部內部單位的設計，除了有專責的題庫建置單位(主要負責測驗題)，也有所謂依考試性質分化的三個考試業務單位高普考試司、特種考試司、特種考試司(主要負責試務及申論題臨時命題的試政工作)，對於評量方法的研發改進，題庫全面性的建置、測驗結果的解釋與應用仍力有未逮，爰有設置「測驗評量專責機構」之議，但定位與任務有待研議。

從先進國家測驗評量中心建置經驗觀之，他們對執行計畫前的規劃、建置過程中的發展、執行業務的過程以及執行業務後的檢討都有清楚的重點與任務界定，各項角色設計例如「評量者」、「管理者」、「角色扮演者」與「設計者」都有其工作內容，以及執行測驗評量過程中對於協助用人機關做決定的貢獻。考選部目前正研議建立測驗評量專責機構，惟思考面向應不僅止於補充建置題庫人力不足以及擴大題庫建置類科，而應思考此機構的建立對我國考選工作辦理測驗評量品質，究竟有何關鍵性的幫助。爰此，擬於第六章結語與建議部分提出對於試題型態、工作分析、測驗解釋、機構各項角色設計以及如何利用所得資料與統計數字提出筆者的看法。

肆、測驗的編譯—測驗國際化的挑戰

受到企業跨國化的影響，以及各國對於多元文化主義(multiculturalism)的體認，即便在單一國家範疇下，亦逐漸意識到提供多語言版本之測驗工具的必要性。因此，測驗國際化及跨文化的議題，在近幾屆國際測驗年會即持續地受到重視（ITC，2010），本次年會亦有針對此一議題之測驗工作坊(workshop)。另國家考試每年辦理 19 至 24 種次之考試，列考之科目數每年超過 2,500 科，使用之試題量相當龐大。因此，有論者提及可引進其他國家之試題以為因應。然而，翻譯其他國家開發之測驗試題，就如同跨國測驗一樣，在測驗理論上仍有許多面向需要考慮，與此一議題相關者即為測驗的改編及翻譯。

一、測驗翻譯(translation)或編譯(adaptation)

一般偏好採取直接翻譯一個已存在的測量工具，而非發展一個新的工具，其主要原因有：

- (一) 節省成本及時間：既有的測驗已具備測驗信、效度相關資訊，且較為人熟知，可以直接採用。
- (二) 發展新測驗工具所需之測驗專家可能不易覓得。
- (三) 可以對於不同文化及語言族群間就測驗結果的差異進行比較。
- (四) 藉由提供受試者以其熟悉的語言版本接受測驗，可以提高測驗的公平性。

然而，直接引進已開發的測驗工具的問題是，因為無法確保測驗所評量的概念一定存在於另一個標的人口的文化中，可能把一個在特定文化存在的測驗結果誤用於另一個文化中，此在心理測驗、人格測驗中最常存在。因此，將一測驗工具運用於不同語言、文化之受測族群，均必須要在新的語言文化下重新定義此一測驗工具的信度與效度。

根據國際測驗委員會的分類，測驗編譯的方法主要有二種：順向(forward translation)或逆向(backward translation)。其中順向翻譯

(forward translation)，是將測驗由原始語言翻譯為標的語言，然後由雙語者比較二個語言版本，此一過程通常是由一位雙語專家進行，其通常不需對於測驗目的及測驗試題清楚瞭解，此種方式較為便宜、簡便，是最常採用之方式。至於逆向翻譯(backward translation)，則將測驗由原始語言翻譯為標的語言，然後再翻譯回原始語言，此一翻譯程序可能重複數回。待翻譯程序完成後，再將最後的翻譯版本與原始版本進行比較。此種翻譯方式通常採取小組進行，小組中包括數名翻譯人員，還包括根據標的語言進行測驗試題改寫、重組的測驗專家，以確保編譯完成的測驗工具適用於跨語言/文化的受測團體。以上二種方式均可採用，惟若必要採行翻譯其他語言存在的測驗工具，以逆向翻譯為較佳，其過程不僅僅進行測驗的語言翻譯(translation)，尚涉及在語言及文化框架內的調適。

測驗編譯是一需要長期、專家技術及高度投資的複雜過程，涉及去除某些不適合的試題，而代之以其他符合新受測者的語言或文化，因此，在測驗編譯的過程中，必須注意到一測驗工具可否達到在另一個文化相關層面的相等性(equivalence)。Lonner & Berry (1986) 提出 4 個層面供思考相等性議題，包括：1. 功能相等 (functional equivalence)：指功能或角色在不同文化仍然存在；2. 概念相等 (conceptual equivalence)：不同文化關於行為或概念在意涵上有相同的意義；3. 語言相等 (linguistic equivalence)：指涉實際的翻譯過程。4. 度量相等 (metric equivalence)：又稱心理距離等值、數量等值，指一項測量工具上獲得的分數，可在不同文化間進行比較，如百分制是國人較為熟悉的度量 (引自 Chang, 2001)。Iliescu (2012) 則在本次測驗年會工作坊中列舉一般性向或人格測驗之內容所可能產生相等性問題的實例：

「測驗所包含的面向是否只適用於特定國家或文化區域，如：貨幣、度量衡。」

11. If a tool rents for 80 cents an hour, how much would it cost to rent it from 8:30 a.m. to 11:00 a.m.?

- A. \$2.40 B. \$2.80 C. \$3.60 D. \$2.00 E. \$5.00

12. Three \$5 bills are equal to how many dimes?

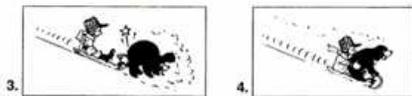
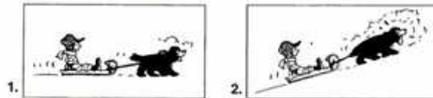
- A. 50 B. 30 C. 150 D. 15 E. 350

13. In a broad-jump competition, John jumped 14ft., 5 in. and Bill jumped 11ft., 9 in. How much farther did John jump than Bill?

- A. 3 ft., 4 in. B. 2 ft., 14 in. C. 2 ft., 8 in.
D. 3 ft., 8 in. E. 26 ft., 2 in.

• Culture / civilization (2)

7.



- A. 1 2 4 3
B. 2 3 1 4
C. 1 4 2 3
D. 3 1 4 2
E. 1 2 3 4

14.



- A. 1 4 3 2
B. 4 3 2 1
C. 1 3 4 2
D. 1 2 4 3
E. 1 3 2 4

例題中的”dime”，因為大部分國家的貨幣並沒有「分」的概念，所以此類試題必需要依據受測者之國家進行調整；另像「度量衡」（公制或英制）。其他如運動項目（棒球、美式足球及英式足球），歷史事件、社會文化名人、地區性地理位置等亦是，如：上圖圖片中的雪上活動，在未下雪的區域即無法使用，可能也沒有適當的概念用以取代該試題所指涉的評量標的。

二、國際測驗委員會（ITC）測驗編譯之指導原則

國際測驗委員會為建立跨語言及文化測驗進行測驗成績相等性

(equivalence) 的指導原則，自 1992 年起，協同歐洲心理測驗協會、歐洲測驗出版機構、國際應用心理學協會、國際心理科學會聯盟、國際教育成就評量協會、國際語言測驗協會等測驗組織，展開撰擬測驗編譯的指導原則。該指導原則主要分成四個面向，以下摘要 2010 年改編版之指導內容：

(一)基本條件

1. 改編測驗之前，必須取得測驗版權所有人的授權。
2. 符合擬採用該改編測驗之所在地或國家關於測驗相關法令規定以及專業的實務規範。
3. 採取適當的測驗編譯設計。
4. 以有利於標的團體之觀點，評估測驗擬評量的相關內容及架構。
5. 評估擬採用該改編測驗之標的團體，因文化及語言差異所可能產生的相關影響。

(二)測驗的發展

1. 必須充分考量將要使用該翻譯測驗工具之標的人口的語言及文化差異。
2. 確認測驗的方向、評分原則、試題以及測驗指導手冊等所採用的語言，不論在語言及文化上，均適用於擬使用該編譯測驗之標的團體，並提供相關證據。
3. 對擬採用該編譯測驗工具之標的團體，提供其選擇所熟悉的測驗技術、試題形態、測驗規範及測驗管理程序的資訊。
4. 對擬採用該編譯測驗工具之標的團體，提供其所熟悉之測驗試題內容及刺激材料的相關資訊。
5. 在語言、心理及文化面上，採用提高不同語言測驗版本間編譯過程正確性及相等性的系統性判斷資料。
6. 必須運用適切的統計技術，以建立不同語言版本測驗工具之相等性；確認測驗工具可能不適用於某一個或更多標的團體的問題內容或面向。

7. 必須提供所有採用該編譯測驗工具之標的團體測驗效度的評估資訊。
8. 應提供關於擬採用該測驗之標的團體有關等化議題的統計資訊。
9. 不同測驗團體所使用之測驗版本間不相等的議題，不應作為不同團體上之比較，惟或可用於個別團體在進行測驗成績報告時增加其內容效度。

(三)測驗的管理

1. 測驗管理者必須預估可能遭遇的問題，並準備適當的材料或指示，用以採取修正前述問題的適當行動。
2. 應注意可能會限制測驗分數效度的測驗刺激材料、行政程序及回應模式等因素。
3. 影響測驗工具管理的環境面向，應盡量在不同標的團體間維持相同。
4. 測驗管理指導應同時具有原始及標的語言版本，以使測驗受標的團體間非預期之變異因素影響最小化。
5. 測驗的操作手冊應依新的文化架構，敘明所有必須審視之行政管理面向。
6. 測驗管理者應是不引人注意，且管理者與受測者之互動應降至最低，在管理手冊中明列管理者應遵循之規範。

(四)成績解釋

1. 當一個測驗工具被運用於其他的標的團體時，應提供其改變的文書記錄以及測驗相等化的資料。
2. 標的團體樣本在測驗成績產生差異，研究者應以其他實證資料證實其差異。
3. 不同標的團體間的測驗結果比較，只有在成績報告的結構上無差異的情形下才可以進行。
4. 測驗發展者應提供關於可能影響測驗標的團體表現的特定社會或

文化生態背景，並提出在解釋測驗結果時如何處理此類影響的程序建議。

三、歐盟機構人事考選制度經驗分享

目前歐洲聯盟（European Union）計有 27 個會員國家，其下有 10 個機構、3 個金融組織、2 個諮詢委員會、3 個跨機構組織及 6 個獨立機構，其中 9 個機構（如表 4-1）人員進用原則由歐洲人事遴選辦公室（European Personnel Selection Office；EPSO）負責辦理其公開考選。由於歐盟會員使用的國家語言計有 23 種。因此，在從事考選測驗時，語言及文化差異即成為辦理考選機構的重要課題。在本次國際測驗委員會年會中，歐洲人事遴選辦公室亦派員參加會議及發表報告，其對於測驗編譯此一議題亦投入相當的關注。以下為該辦公室在本次年會所發表關於其考選制度的簡介。

（一）考選組織

歐盟機構的人員，原則應透過公開競爭的考選程序進用，進用的人員包括永久職及契約職，歐洲人事遴選辦公室每年考選之人員以契約職為主。契約職通常分為二類：一為長期契約，即初期契約為 5 年，期滿後可最長再延為 5 年；二為短期契約，契約期間為 3 個月至 3 年，通常是作為正式職員之代理人、因旺季或特定工作技能需求進用人力。所有的人事考選活動，係由歐洲人事遴選辦公室及獨立的遴選委員會二個單位共同辦理，其職掌分別為：（European Personnel Selection Office，2009）

1. 歐洲人事遴選辦公室的職掌：

- (1) 製作及公告甄選須知。
- (2) 依據用人機關的指派，選任遴選委員會之委員成員。
- (3) 審核是否受理申請者之報名申請。
- (4) 審核申請者是否符合須知所定之資格及各種規定（如：於報名期限內提交報名、只有一個報名帳戶、符合基本資格要求）。
- (5) 提供遴選委員會行政支援。

- (6)負責有關測驗之所有行政事務。
- (7)組織初試及複試各階段之測驗。
- (8)將合格候用名單遞送各用人機關。
- (9)代表機關管理候用名單。

2. 遴選委員會的職掌：

- (1) 審查應試者是否符合工作之特定條件，如經歷審查。
- (2) 依據甄選的型態，制訂筆試、口語及實作測驗之試題。
- (3) 在不同的測驗中，評估應試者的表現。
- (4) 在甄選程序結束時，提交候用名單給歐洲人事遴選辦公室。

遴選委員會可以請求專家擔任諮詢角色之主考委員，以協助其任務，但這些專家不列為遴選委員會成員，遴選委員會仍自行決定候選人各項測驗成績。遴選委員會名單在口試前會由歐盟人事遴選辦公室公布於網站上，參與甄選的應試者依法不得與委員有直接或間接之接觸，違反者會被逕行判定為不適格。所有與委員會的聯繫，均必須透過人事遴選辦公室進行。

表 4-1 歐洲人事遴選辦公室考選人力之機構及其職員數

機構	至 2012 的職員人數
歐洲議會 (European Parliament)	6,655
歐洲理事會 (或稱歐盟各國部長理事會，the Council of European Union)	3,153
歐盟執委會 (European Commission)	25,478
歐洲法院 (Court of Justice)	1,952
歐洲審計院 (European Court of Auditors)	887
歐洲經濟暨社會委員會 (European Economic and Social Committee)	724
歐洲監察使 (European Ombudsman)	531
歐洲資料隱私保護監督官 (European Data Protection Supervisor)	66
歐洲對外行動部 (European External Action Service, EEAS)	43
總計	41,159

資料來源：Nussbaum，2012。

在以往的考試內容中，有關歐盟機構和歐盟政策的專業知識十分重要，而且某些職位的錄用還需加試專業知識。例如報考經濟職務者加試經濟學知識，報考律師者加試法律知識。從 2010 年開始，歐洲人事遴選辦公室開始引進新的甄選方式，新考試體系將重點置於測試應考者的學習能力、適應能力、口頭表達及寫作能力、溝通能力和團隊合作的能力。考試內容從傳統的智商(IQ)測驗轉為情緒商數(EQ)、能力和潛能。人事遴選辦公室認為，專業知識並非不重要，但學歷證書已經足夠證明應考者已經掌握基本的知識結構，不需要在考試中進行重複測試。因此在新考試體系中，專業知識部分僅占總分的 20%左右。

(二) 考選制度簡介

1. 應考資格

(1) 基本應考資格：

- A. 必須是歐盟會員國的公民。
- B. 須擁有完整的公民權。
- C. 必須完成任何有關於軍事服務之義務。
- D. 必須符合該職務所需要之特質。

2. 個別應考資格：

- A. 學歷：行政職類，若職務為 AD5，至少具備 3 年制專科以上教育程度；AD7 則至少具備 4 年制學士以上教育程度；助理職類，則至少具備相當我國高中教育程度。
- B. 專業經驗：視出缺職務需要。
- C. 語言的知識：通常並不需要提出證明文件，特定職務會在須知中註明需要提出之語言能力證明文件或學位。
- D. 主要語言：即歐盟會員國所使用之 23 種官方語言，部分視職缺會指定第一語言。
- E. 第二語言：必須除前揭第一語言以外之語言，且為英、法、德語

之一。

一般不會有年齡限制，不過，必須注意須知中註明各機關的退休年齡限制。

2. 報名方式：均採網路線上報名，完成報名後，再依規定下載報名表，並將證明文件（如學歷、工作經驗證明、經歷表等），以掛號寄出。

3. 考選方式

考選方式分為兩個階段（考選內容及時程如表 4-2），第一階段是初試（Admission Tests），重點在測試應考者的語文、數字、抽象推理能力以及情境應變能力等一般性向及能力，初試之試題除情境應變能力測驗部分僅有英、法、德語三種語文外，其餘所採用語文為第一語言。考試採電腦測驗（Computer-based tests）方式、題型均為選擇題，考試的地點及時間通常在 27 個會員國有多個測驗中心及指定日期，應考者可於獲得應試通知時自行選擇地點及時間。只有通過第一階段，才能夠進入第二階段考試，惟初試之成績不列入最終測驗結果。

一般依原公告的需用名額錄取約 2.5 倍進入複試階段（評鑑中心），複試通常集中於歐盟辦公中心比利時布魯塞爾舉行，為期 1 天，每天最多 36 人分成 6 組進行測驗，以英、法、德語三種語言之一進行，考試之內容主要分為二部分：一是專業領域：通常以個案研究進行，占總成績 25%；二是基本核心職能：即表 4-3 所列 8 種核心職能，一般而言，領導能力只有在特定 AD7 的職務才會列入，占總成績 75%。

表 4-2 歐盟機構人力考選內容及時程

職務	行政職類(AD) 如：公共行政、人力資源、經濟、法律、財政、審計及統計	語言類 如：翻譯或通譯	助理類(AST) 如：行政助理、人力資源、財政、資訊科技、傳播、秘書	特殊職務類 如：法律語言專家、護士、核子觀察員、科學研究人員	
所需期程(月)	公告甄選須知及自我評估				
1	線上報名 春季	線上報名 夏季	線上報名 冬季	線上報名 不定期	
2	電腦化測驗：內容包括語文推理、數字推理、抽象推理能力及情境判斷測驗	電腦化測驗：內容包括認知推理能力及專業能力測驗	電腦化測驗：內容包括語文推理、數字推理、抽象推理能力、模擬判斷測驗及專業技能測驗	依所定資格之簡歷資料 (CV)	
3					初試
4					初試
5	評鑑中心： 1. 專業領域個案研究(採書寫) 2. 團體討論 3. 口頭簡報 4. 結構化面試	評鑑中心： 1. 實用語言測驗 2. 團體討論 3. 口頭簡報 4. 結構化面試	評鑑中心： 1. 專業技能測驗(採實作) 2. 結構化面試	評鑑中心： 1. 一般性個案研究 2. 結構化面試 3. 團體討論 4. 語文推理、數字推理、抽象推理能力測驗	
6					複試
7					複試
8					複試
9	候用名單(合格者列入資料庫，不同之機構可檢視合格者之資料，然後提供職務；通常機關在進用之前，會再邀請該人員進行面試。名單有效期間原則為1年)。				

資料來源：European Personnel Selection Office，2010。

表 4-3 評鑑中心所評量的測驗項目及欲評量之核心職能

測驗方式 核心職能	個案研究	團體討論	口頭簡報	結構化面試
分析及問題解決	◎		◎	
溝通	◎		◎	
服務品質及結果	◎		◎	
學習與發展		◎		◎
權衡輕重緩急及組織力	◎	◎		
順應力			◎	◎
與他人共事		◎		◎
領導		◎		◎

資料來源：European Personnel Selection Office，2012。

每年大約會有 4 萬多人報名甄選，經過第一階段的電腦測驗，錄取約 1 千人進入第二階段評鑑中心之測驗，最後約只有近 300 人通過被列入候用名冊。遴選委員會依據應考者在評鑑中心所得成績及其專業證明內容(如相關專業工作經驗、證照)，按其報考領域及測驗成績分類排序及格之候用名單(通常最大分為 4 類，同類之及格名單則依字母排序)。候用人員通常約 80-90% 經機關面試後進用(如預算有問題則儲備資格可以延長 1-2 年)，試用 9 個月後，如不合格則予以解聘。

依據本次國際測驗年會歐洲人事遴選辦公室發表之報告，由於測驗採多種語言進行，所有測驗試題均需進行測驗編譯 (Test Adaptation) 程序及進行試題校準 (Item Calibration)，並採用試題反應理論 (Item Response Theory, IRT) 中之二元計分 (Dichotomous) 及多面向 Rasch 計分模式進行計分。

參與考選之應考者對於測驗結果如有異議，可於收到通知 20 個工作天

內向遴選委員會的主席提出異議；或在3個月內向歐盟執委會提出；另對於考選階段，亦得以管理不善、違法行政、不公平、歧視、權力濫用、不回覆、拒絕提供資訊、不合理的延遲等理由，向歐洲監察使申訴，要求調查。每年監察使所受理的申訴案件中約12%與歐洲人事遴選辦公室相關。

歐盟機構人力考選的運作，是國際測驗委員會近幾年所關注之測驗跨語言及跨文化議題的縮影，在本次年會「The Assessment of Critical Thinking: Cross-cultural and Validity Issues」此一場研討會中，來自美國、香港及葡萄牙等3國學者所發表的4篇研究報告中，即提出了語言對於評量思辯能力（Critical Thinking）此類之認知能力，確有運用及評量的文化差異存在，而造成測驗結果的不確定性。而依據歐洲人事遴選辦公室與會人員的報告，在該辦公室辦理測驗的過程中，多語言及多文化也成為其評量上的障礙。

伍、測驗資訊科技運用的進展

本次年會的開幕演講主題「評量之演進：模擬及非娛樂用途的遊戲(The Evolution of Assessment: Simulations and Serious Games)」，主講人 Mike Fetzner 先生提到，自 1980 年代中期，由於電腦的運用，使得測驗開始有新的發展方向，除了早期以電腦取代紙筆為作答工具，運用電腦登錄及管理成績，到電腦適性化測驗，乃至於近年運用資訊科技中的多媒體虛擬動畫 (multimedia simulations)、3D 技術，增加評量內容與實際工作的真實性。新的電腦科技亦提供測驗在評量工具及在評量內容更豐富化的可能性，使得新的試題形態及互動機制變得更為可行。

本次會議中有許多議題與電腦化測驗相關，電腦化適性測驗 (Computer adaptive tests, CAT) 仍是討論最多的議題。此外，因應 21 世紀技能 (21st century skills) 概念的提出，如何利用資訊通信科技 (Information and Communication Technology, ICT) 將測驗的形態由傳統的選擇式試題轉移至綜合形態，包括實作評量 (performance assessment)，以評量 21 世紀社會所需的認知技能，如問題解決、思辯 (critical thinking) 及推理能力，在本次會議中亦介紹幾個採用 ICT 的評量應用實例。相關議題說明如下：

一、電腦化測驗的挑戰

近年來，由於電腦測驗可以節省紙筆測驗印製及運送試卷、試題等紙本之成本，且可以提供較快速的成績計算及進行測驗結果的統計及分析等便利性，在電腦效能大幅提升與網際網路蓬勃發展的推導下，不論是採用電腦化測驗 (CBT) 或電腦化適性測驗 (CAT)，電腦測驗已有逐漸取代紙筆測驗的趨勢。

本次年會第 26 場研討會「Paper-pencil and computerized testing: Practical matters」中，即有 4 篇針對電腦化測驗議題提出研究，其主題包括如下：

- (一) 比較相同內容的紙筆測驗與電腦測驗的結果，以確認此二種評量工具可提供一致評量效果的程度；電腦管理的評分與紙筆測驗評分是否相等。
- (二) 受試者對於電腦測驗工具的接受度。
- (三) 電腦化適性測驗 (Computer adaptive tests) 中，試題參數的不同，是否會造成測驗結果的差異。

其中，關於議題 (一)，在過去二、三十年紙筆測驗轉移至電腦測驗的發展過程中，已有不少研究針對電腦測驗與紙筆測驗的可比較性 (comparability)，及測驗形式、介面 (test media) 對於受測者的影響進行探討。但至今研究結果仍相當分歧，有的研究發現紙筆測驗受測者表現較佳，也有研究的結果顯示，電腦和紙筆測驗成績沒有差異或差異很小。(Schroeders, 2009；林玲英、藍珮君, 2011) 因此，雖然從紙筆測驗轉進至電腦測驗在國際測驗機構已蔚為主流，然而，關於測驗形式、介面對於測驗效度是否有影響的議題，仍然持續被討論，也被認為是必要的。

此外，自電腦測驗發展以來，即不斷有相同之議題被提出討論，其主要的議題有：

- (一) 成本效益：採取電腦測驗的起始成本通常高出紙筆測驗非常多，包括電腦軟體、硬體設備、網路、標準化測驗環境設置及技術支援與維持的成本。(Kozma, 2009；Kikis-Papadakis & Kollias, 2009) 且相較於紙本測驗，每一次的電腦化測驗，都需要投入更大量的前置規劃流程。(Kyllonen, 2009) 雖然部分的研究指出，電腦測驗第一階段的成本問題，可以經由長期運用得到舒緩，但是，相對於紙本測驗，電腦測驗無疑地還是被認為需付出較高成本。(Farcot & Latour, 2009)
- (二) 測驗公平性：因為與擬評量之內容無關的呈現方式之差異，可能造成對於受測者表現的影響，如電腦螢幕尺寸、螢幕呈現效果、使用的鍵盤種類等都可能產生差異的影響。(Bridgeman, 2009) 若干研

究則提出因貧富或城鄉差距造成之資訊熟悉度落差，形成對測驗公平的影響。(Buško, 2009; Kikis-Papadakis & Kollias, 2009)而 2006 年在愛爾蘭、韓國及丹麥 3 國舉行的經濟合作暨發展組織國際學生能力評量計畫 (OECD PISA) 的電腦測驗結果，則引起了電腦測驗造成性別差異影響的討論 (Martin, 2009)。

(三) 測驗建構無關的變異：受測者可能受到與測驗目的無關的資訊素養 (如電腦打字速度) 干擾而影響其在測驗上的表現。(Bridgeman, 2009)

二、電腦測驗的創新--21 世紀技能與評量方式的改變

雖然 21 世紀技能概念的提出，主要在針對於教育體制內之教育系統、課程內容及教學的變革，但是，因其關注的知識技能無法使用傳統的測驗加以評量，由其衍生對於評量 (測驗) 方法的再思考，以及運用資訊通信科技發展新的評量方式，因此，在本次年會中除由荷蘭測驗機構 Cito² 進行專題演講外，在年會第 27 場研討會「The assessment and teaching of 21st century skills」中，亦由澳洲、芬蘭、荷蘭等 3 個參與前述 21 世紀技能評量及教師計劃國家之學者提出相關報告。以下摘要說明其概念及其運用電腦測驗實例。

(一) 21 世紀技能概念簡介

美國在 2002 年成立一個產官學界合作的組織 Partnership for 21st Century Skills，目的在於創建一個能將 21 世紀的公民所應具備的知識和技能融入教育系統的成功學習模式。在該組織所提出一份先期的研究報告《Learning for the 21st Century》中指出，經濟、科技、資訊、人口統計、和政治上的力量已使人們的工作和生活產生了重大變化，這些改變

² Cito (Central Institute for Test Development) 是由荷蘭學者參照美國測驗服務社 (ETS) 於 1968 年成立的測驗組織，其在 1987 年依據荷蘭教育福利法成爲該國的公立教育測驗機構。嗣於 1999 年經修法民營化，成爲一提供教育測驗及企業人力測驗工具發展、訓練及諮詢等服務的測驗公司，總公司位於荷蘭，目前在德國、美國及土耳其均設有分公司。(請參考：http://www.cito.com/about_cito/history_of_cito.aspx)

以及它們改變的速率會持續加快，而學校必須和業界、社區以及家庭一樣，針對這些改變加以調適，才能成功和繁榮（Partnership for the 21st Century Skills, 2002）。目前該組織正在和美國民營事業、聯邦政府和許多州政府合作，大力推動這方面的教育革新。

其後，在 2009 年由思科(Cisco)、英特爾(Intel)和微軟(Microsoft)三家科技公司，夥同 Partnership for the 21st Century Skills 和其他教育相關機構，提出一個跨國的計畫 Transforming Education: Assessing and Teaching 21st Century Skill³。在這個計畫中，主張目前的教育體系需要改變，理由有二：一、由於資訊通信科技的發展，全球經濟結構從今天看來已和 20 世紀大不相同。一些全球經濟領先的國家，現在是奠基於製造和提供更多資訊產品及服務，而非仰賴材料的製造；即使是製造業也是依賴資訊通信技術的創新使用。同樣地，由於資訊通信科技運用的無所不在，人們獲取、運用及創造資訊與知識的方式也異於以往。二、21 世紀的公民和從業人員需要新的技能：在當代的工商業界，人們在跨越領域組成的團隊中工作，他們使用各種社會、數位和物質上的資源，並且利用學科知識但不受其所限，來解決複雜且不具良好結構的問題。他們創建新理念、產品和服務，並且和同事、客戶或更廣大的對象來分享。

在此一計畫下，歸結定義出 21 世紀的技能架構為：1. 思考的方式：包括（1）創意與創造力；（2）思辯能力、問題解決與決策能力；（3）學習如何學習，後設認知（metacognition⁴）。2. 工作的方式：包括（1）溝通能力；（2）合作、團隊能力。3. 工作的工具：包括（1）資訊素養（information literacy），指知道何時需要資訊，並能有效找尋、評估和利用資訊的能力；（2）資通科技素養（ICT literacy），指能利用資訊科技蒐集、管理、整合、評估、溝通及創造多元的資訊，並且有意義地運用。4. 生活能力：包

³詳細資料請參考：<http://atc21s.org/>

⁴後設認知（metacognition）：簡言之，就是對自己的認知過程（包括：記憶、感知、計算、聯想等各項）的思考。（參考：<http://zh.wikipedia.org/zh-tw/%E5%BE%8C%E8%A8%AD%E8%AA%8D%E7%9F%A5>）

括(1)公民行為(citizenship),不論是區域性的或全球性的;(2)生活與職涯能力;(3)個人與社會責任,跨文化認知能力。(Assessing & Teaching of 21st Century Skill:2010)

(二) 運用資訊通信科技 (ICT) 的評量工具實例

Kyllonen (2009) 曾提出,論及電腦化測驗的應用,除了便利性外,另一個無法為人所爭辯的論據,即在於資訊科技可以提供豐富的刺激來源,如動態影像、聲音或圖畫,以及提供受測者與評量資料進行互動(Interconnected),而得以評量紙筆測驗無法評量之技能。此一論點在討論 21 世紀技能評量中則得到更多的實現,以下是在 21 世紀技能白皮書中,所列出符合應用資訊通信科技評量綜合技能的測驗實例 (Assessing & Teaching of 21st Century Skill:2010):

1. 美國醫師執照資格考試第 3 階段臨床能力綜合考試 (United States Medical Licensing Examination; USMLE Step 3) 之 Primum® 臨床個案模擬測驗

此一測驗是運用電腦虛擬及劇本評量方式所發展出的電腦化測驗 (Computer-based Case Simulations; CCS)⁵,在此電腦測驗中,每一個臨床病人之個案是由 Primum® 這個電腦軟體控制,經由電腦的四個模擬介面提供:(1)病人資料:關於模擬病人之病史、主訴及主要生命徵象。(2)診斷介面:此一介面中的診療單中可輸入給予病人之病理檢查、施予藥物、進行處方或建議等診療行為。(3)虛擬時間介面:透過虛擬時間使需要時間發展的病理檢查結果或治療反應,得以進行。(4)病人醫療地點介面:將醫療地點改變至開刀房、加護病房或回家休養。(電腦測驗介面如圖 5-1) 此一測驗乃利用多媒體科技,提供動態及互動式的模擬狀態,個案病症之狀態會隨著受測者所施予的治療或檢驗等作為,以及虛擬時間的變化,產生

⁵美國醫師執照資格考試個案模擬電腦化測驗模擬網站:<http://www.usmle.org/practice-materials/index.html>

如真實病人的臨床變化，用以評估受測者包括診斷、治療及監測的臨床技能，內容含蓋 12 個個案模擬，總測驗時間為 4 小時。

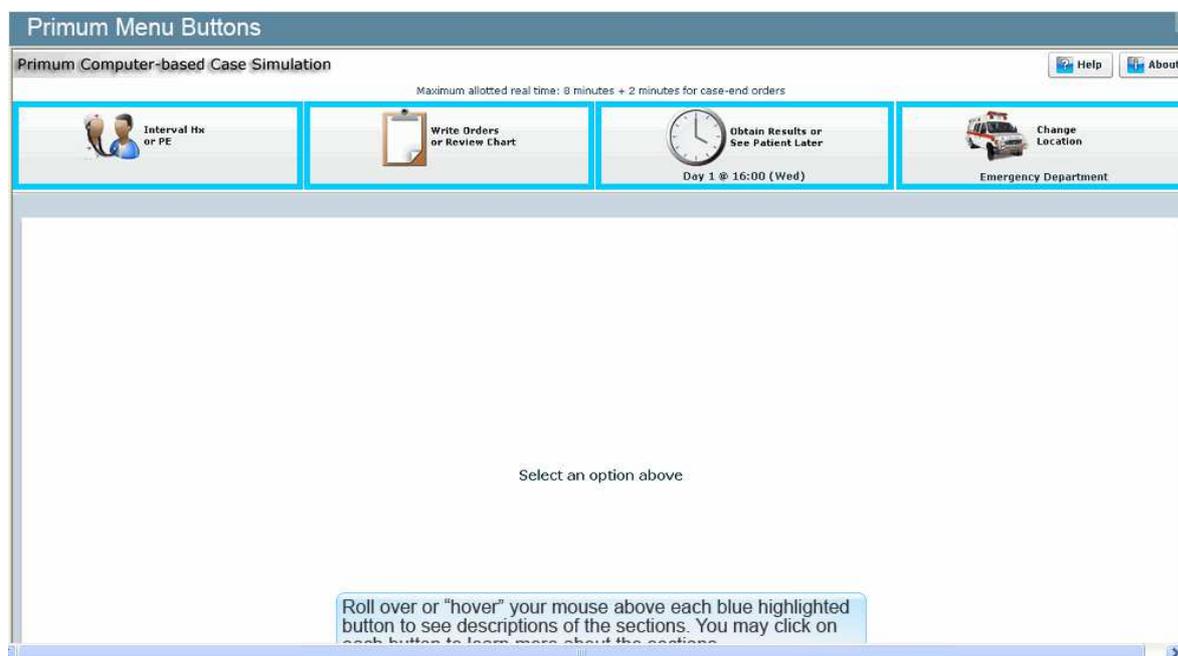


圖 5-1 Primum®臨床個案模擬測驗電腦測驗畫面

資料來源：美國醫師執照資格考試個案模擬電腦化測驗模擬網頁

(<http://www.usmle.org/practice-materials/index.html>)。

2. 世界數學測試 (World Class Tests)⁶

此一測驗是由英國教育及技能部於 2001 年發展，用於評量 8 至 14 歲學生在數學及科學的問題解決能力，其利用電腦模擬使測驗的內容更生活化、具互動性，目前已被部分國家採用作為評核數學資優能力的評量工具。香港科技大學於 2008 年 2 月成立世界數學測試亞洲中心，目前在包括中國、英國、美國、澳洲、香港、紐西蘭、沙烏地阿拉伯、斯洛伐克、阿拉伯聯合大公國、日本、韓國及新加坡等國家和地區設有考試點。(世界數學測試亞洲中心網頁，2012) 圖 5-2 為該測驗的試題樣本，受測者需於電腦上完成解題。

⁶ 詳細資料請參考：<http://www.worldclassarena.hk/tc/about/aboutWCT.htm>

旋轉三角形

若點擊紅色三角形上任何一隻角，
三角形會以該角為中心轉動90°



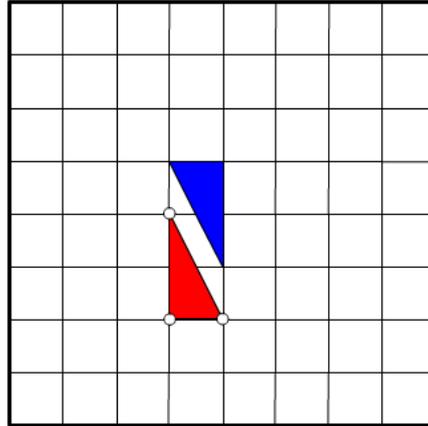
步驟數目



最佳成績

試利用最少的步驟，
把紅色三角形移動至
藍色三角形的位置上。

你應可以少於12個
步驟完成任務。



重新開始

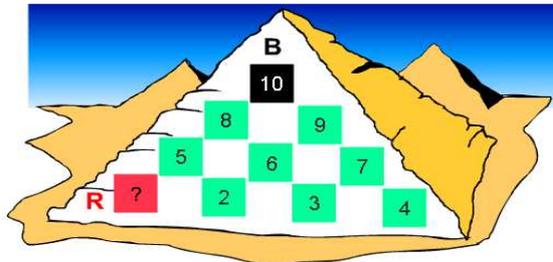


後退一個步驟

金字塔

第一頁 第二頁 第三頁

試在紅色格內輸入一個少於10的整數，
其他格內的數字會隨之而變更。
試重覆以上的指示，並觀察其變化。



1. 當 $R = 8$ 時， B 是多少？

$B =$

2. 當 $B = 68$ ， R 是多少？

$R =$

圖 5-2：世界數學測驗試題樣本

資料來源：世界數學測試亞洲中心網頁 (<http://www.worldclassarena.hk/tc/about/ExampleQuestions.htm>)。

3. 哈佛大學教育研究所虛擬表現評量計畫(Virtual Performance

Assessment Project ; VPA)⁷

此一計畫與美國國家科學教育標準(National Science Education Standard)相銜接，目的在於評量學生是否具備未來職場與公民生活全球經濟發展所需的核心理念，包括創造力、通訊科技、資訊處理及問題解決能力。

計畫之測驗內容係利用 3D 電腦技術模擬實際問題，提供受測者執行提問、資訊蒐集的電腦作答介面，最後指出問題解決方案。以該計畫網頁所提供之模擬測驗為例，受測者在預設的環境問題中，透過角色扮演進行有關問題的實驗設計、資料蒐集，最後提出因果判斷及解決方案。



圖 5-3：虛擬表現評量計畫測驗畫面

資料來源：哈佛大學教育研究所虛擬表現評量計畫網頁

(<http://vpa.gse.harvard.edu/home>)。

4. 盧森堡大學電腦問題解決模擬情境式智力測驗 (The Genetics Lab - computer-based problem solving scenario to assess

⁷ 詳細資料請參考：<http://vpa.gse.harvard.edu/>

intelligence)

此一問題解決模擬情境式智力電腦測驗，係由盧森堡大學 Genetics Lab (GL)與 Henri Tudor 公共研究中心 (Public Research Centre Henri Tudor) 共同開發。測驗之內容與上述哈佛大學的虛擬表現評量計畫 VPA 相似，係運用微世界 (microworlds) 的概念，提供受測者在一個含有問題解決過程及結果的互動與動態模擬劇本 (scenarios) 下進行相關問題解決能力測驗，其測驗的試題架構及檢測訊息內容如圖 5-4 及圖 5-5。

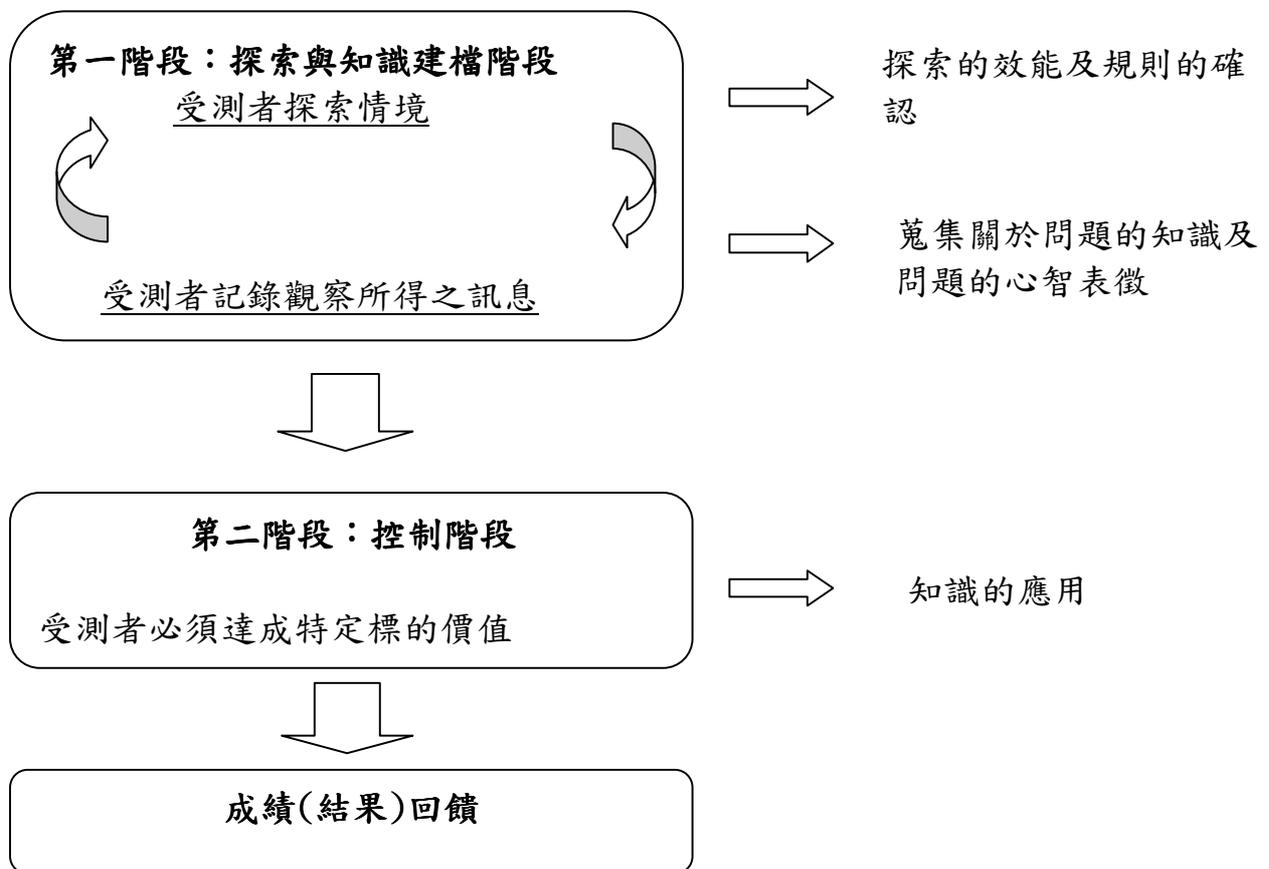
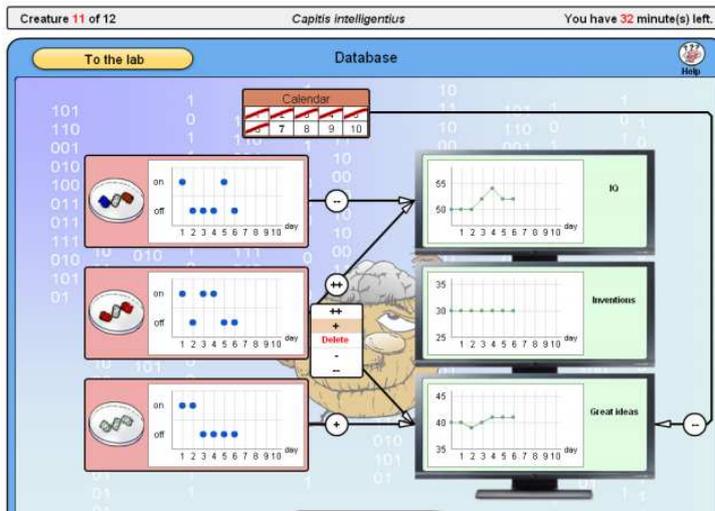
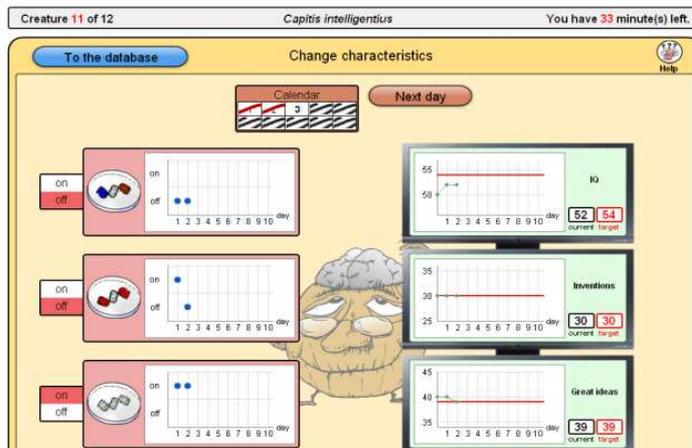


圖 5-4 Genetics Lab 電腦問題解決模擬情境式智力測驗試題架構
資料來源：Somleitner, P., Burnner, M., Keller, U., Martin, R., & Latour, T. (2011)。

Phase 1: Knowledge Documentation



Phase 2: Knowledge Application



Feedback after each item



圖 5-5 Genetics Lab 電腦問題解決模擬情境式智力測驗畫面
資料來源：Somleitner, P., Burnner, M., Keller, U., Martin, R., & Latour, T.

(2011)。

將資訊通信科技 (ICT) 運用於電腦測驗上，應得以評量傳統紙筆測驗無法評量的新概念，如複雜問題解決能力 (complex problem solving) 即是其中之一，惟學者 Kozma (2009) 指出，除了與一般電腦化測驗面臨相同的挑戰外，如將資訊通信科技運用於評量，在測驗理論上另面臨其他的挑戰如下：

1. 必須確認資訊通信科技試題內容所評量的主體知識，與紙筆測驗所評量的結果具相等性。
2. 必須釐清那些技能只能藉由資訊通信科技評量。
3. 必須將主體知識的基本概念與 21 世紀技能結合於評量中，同時，在評量結果釐清那些是主體知識，那些是 21 世紀技能。
4. 發展如何針對此類評量的測驗過程、策略及結果進行評分的新方法或模式。
5. 必須事先建立判斷此類能力連續表現之品質的預測能力。

陸、結語與建議

本次參加 2012 年國際測驗委員會年會，收穫良多，以下僅就報告所提之各項研究發展趨勢，與我國國家考試相關的部分，提出綜合結語及建議。

一、電腦化測驗使用多媒體優勢，檢驗特殊效度

採用電腦作為介面進行評量，多數研究均指出電腦化測驗因有各項多媒體工具的優勢，可以表現原有紙筆測驗試題無法達成的功能，進而提升其效度，若僅是將試題從紙本呈現轉為電腦螢幕介面呈現，就有些可惜了。尤其目前電腦化測驗多應用於醫事人員類科，多數醫事人員也經常使用電腦判斷病人的狀況，但目前的試題呈現除文字外僅有彩色圖片加以輔助，若可使用其影音動態的多媒體優勢，對於檢驗應考人對於某些測驗內容的精熟與判斷力，有助於特殊效度的確認。惟此種進步須在作答時間、試題曝光程度、命題者配合以及本部提供製作試題的協助下才能完成此類試題。

二、新增科技功能，保持系統彈性

一套電腦化測驗系統的建置，所費不貲，其中又包括派題與作答系統、題庫系統彼此的配合。測驗試題呈現的形式有日新月異的發展，也有不同的類科屬性，目前電腦化測驗系統多屬委外建置，以本部目前的人力狀況，並不容易將完整的需求陳述清楚，即使是委外廠商也經常在摸索所中進步。考選部除了不斷蒐集國外電腦化測驗系統建置的進步情況，也需預留新的科技功能應用於電腦化測驗中的彈性，即使需求變動擴增，也無須重新建置，對於減少經費的支出將有實質的幫助。

三、編寫發展電腦化測驗的參考手冊

先進國家發展電腦化測驗雖然依據工作階段不同，分為測驗發展者、測驗出版者及測驗使用者，提供不同的參考規範，但三者之間彼此

的聯繫關係非常緊密。考選部發展電腦化測驗系統之軟體、硬體則由資訊管理處委託廠商建置，試題則由題庫管理處遴聘學者專家命題、審查，行政人員則負責相關行政程序及藉由系統完成校對工作，職責雖然明確，但彼此聯繫性卻不高。國外的經驗提供了我們很好的參考，電腦化測驗的系統建置，包括網路的傳遞安全、軟體的設計、設備的完善性、試題的型態等等，都須要各項完整的文件或注意事項提醒，且國外電腦科技應用於電腦化測驗不斷的進步，除持續蒐集各項改進重點外，將這些蒐集的資料編寫成考手冊且持續增補，對確保電腦化測驗軟、硬體的設計更新與維護，將有極大的助益。

四、透過比對系統處理試題曝光議題

歐美國家的電腦化測驗多屬適性化測驗，爰此，試題多屬不公布狀態，惟使用多次以後，即須檢討其試題曝光問題。我國國家考試考畢後試題隨即公布，應考人基於知的權利與民主化的氛圍，目前很難改變試題公布政策，然而試題公布使考選部面對另外一種試題曝光的問題，即考古題是否能重複使用。以往考古題的重覆使用已發生了數次應考人對於考選部舉辦考試公信力的質疑，現在除提醒委員命題注意避免與三年內考畢試題雷同外，也建構電腦比對系統，有其實質避免試題雷同之成效，但也可能造成委員未來選擇冷僻評量內容命題的機率，此點須加以注意。

目前採紙筆測驗的國家考試已有完善的比對系統處理已曝光(考畢)試題，惟電腦化測驗尚無完善的比對系統，其雷同比率比對值也尚未建立。理想的作法應該是委員命題完畢將試題傳送給考選部後，就應進行雷同值的比對，若部分試題雷同度過高，就應退回給委員重新修改試題，以提高試題的品質及可用率，除了與歷屆考試試題比對外，委員所命擬試題彼此間的比對也必須進行。雖然應考人對於使用歷屆試題相當歡迎，尤其使用某一通過標準的專技人員考試測驗題，也比較不會向考選部抗議，但為了舉辦考試的公平性以及維持測驗的評量效能，電腦化測驗的試題仍須加以

比對。

五、利用電腦化測驗蒐集的各項作答樣態與統計數字，提升未來試題品質，也可提供應考人回饋

現行國家考試電腦化測驗的優點在於即測即評，應考人可以在作答完畢後選擇是否在第一時間得知分數與及格狀態，惟我們應可對它有更高的期待。無論是電腦化測驗（CBT）或是電腦化適性測驗（CAT）與傳統紙筆測驗最大的不同點在於可利用電腦軟體的設計蒐集應考人的作答資料，且即時加以分析試題的難易度、鑑別度與信度，利用試題反應理論（IRT）還可得知其猜測度，分析應考人對於試題的偏好、作答差異，對於提升未來命題品質有極重要的參考效果。

且目前專技人員電腦化測驗多有命題大綱，在整份測驗中，依據命題範圍也可切割成各項分測驗，給予應考人一個整體的分數雖然代表及格與否，但對整份測驗的試題內容缺少整體性的判讀，若能歸納統計應考人各分測驗的表現情況，對命題委員或抽題難易度的分配將更具參考價值，若能繪成分測剖面圖提供應考人回饋，更可提升考選部對於評量分數的專業處理形象。

六、釐清測驗評量專責機構的人員角色

考選部研議成立測驗評量專業機構時，除討論如何設計它的機關屬性可以充分發揮其功能外，還要思考清楚的組織目標、承諾以及如何與考選部內部的原有單位協調合作。尤其在工作任務的設計上，應配合各種人員的配置，如評量方法研究者、評量活動設計與執行者、行政管理者以及測驗資料分析與報告撰寫者的角色，依據組織希望達成目標與範圍之大小分別遴聘，前述幾種角色有些也可同時扮演兩種功能，但其工作成果的導向更應思考如何落實於國家考試試務與試政工作之改進執行。

七、在試題與評量程序中，加入模擬測驗的概念

模擬測驗在國外測驗評量專業機構中使用的頻率相當高，有些為用人機關設計的測驗中甚至規定至少需使用兩種模擬測驗。模擬測驗的概念即在於模擬真實工作中可能產生的情境，藉以評量其專業知識、臨場反應與處理方法的適當性，通常此種測驗較具有預測效度。目前的國家考試應試科目的設計大多趨向分科專業知識成就測驗，應試科目的設計有時其區別效度並不明顯，評量受試者的能力單一面向的性質較高；即使有口試，往往限於時間與配分比例過低的關係，對整體測量結果的改變影響並不大。較接近模擬測驗性質的國家考試評量方法，是從民國 100 年開始的警察情境測驗，這是一個很好的開始與突破，惟目前仍僅限於筆試科目的一科。模擬測驗的概念可思考突破應試科目設計，成為一項「評量程序」，在人力資源不足的情況下，似可研究與口試的結合，雖然有人會批評口試在未結構化的情況下，會影響評分信度，惟「模擬測驗」若導入結構化的評量與計分，反而可增進其評量的整體效能與有效性。

八、加強核心職能、工作分析、應試科目與評量方法的連結

工作分析是國外測驗評量機構重要核心業務之一，設計測驗方法與評量活動時，均須對用人機關的工作內容與所需能力加以分析，以確保選才的過程正確無誤。考選部自兩年前開始進行各用人機關核心職能的評估工作，希望從用人機關的立場確定所需人才的能力屬性。核心職能評估可視為工作分析的一部分，評估者主要係由該職系類科之中央主關機關負責執行，亦無可厚非，惟國外的工作分析通常與評量方法相結合，才能彰顯考用合一的目標。核心職能的評估既已全面展開，但評估後的結果如何與應試科目與評量方法互相連結，應預先做準備，所有評量階段與評量活動的設計都應以核心職能評估作基礎，完成用人機關的職能評估項目後，更應思考的面向是如何應用在國家考試各項甄選程序與評量重點的改變。

九、強化測驗結果解釋的能力

長久以來，考選部日復一日、年復一年的舉辦考試，一項考試的放榜，又是另外一項考試報名的開始，部內同仁也很少對各項考試的評量結果，進行完整的分析，以瞭解評量所呈現的優點與缺點。測驗解釋的能力對於辦理考試機關相當重要，各項評量方法、評量活動、試題刺激能否符合測驗的目標，試題是否可充分發揮評量功能，評量的誤差，都應仔細檢驗，以作為下一次辦理考試的參考。在未成立測驗評量專責機構前，可選擇特殊需要之類科（例如錄取不足額或錄取率偏高或偏低），進行測驗結果完整的解釋，以作為下一次命題、題庫建置與閱卷的改進。

十、善用統計資料擴大決定之層次與精確性

現代科技與統計軟體的進步，對於評量結果的決定可發揮極大的功能，例如決定及格與否。惟進行評量過程中仍有許多的統計資料可供其它「行政決定」所用，例如命題委員與閱卷委員的遴聘，藉由應考人作答資料所得之統計數字，可瞭解委員的命題與閱卷品質，尤其對於閱卷委員的閱卷寬嚴程度與分數偏好可得到較詳細的資訊，平行兩閱時計算其閱卷相關係數進而確認其評分者信度，對於未來委員的遴聘更有實質的幫助，這些都屬於行政決定的一部分。如果閱卷委員其寬嚴標準飄忽不定，又不依據參考答案評分，對國家考試的公平性影響甚鉅；命題內容的設計與評量層次的設定亦是如此，善用統計資料，可擴大辦理國家考試各項試務或試政工作之決定層次，以及提升評量之精確性。

十一、視考試性質，評估將具評鑑效度之考畢試題重新列入題庫運用之可行性

我國國家考試以筆試測驗作為主要評量工具，測驗試題之內容亦偏向學科成就測驗，每年列考之考試科目超過 2,500 科以上，此於先進國家人力考選及測驗機構中均鮮少採用。另為配合試題疑義處理程序，國家考試之試題於考畢後均對外公布，此又與一般測驗機構原則不公布

試題，以精簡試題題庫開發成本之思維及作法大相逕庭。

目前國家考試為避免試題雷同爭議，原則並不重新運用考畢試題，除使少數科目之試題設計落入注重學科知能細節或是否偏離核心知能之爭議外，亦增加國家考試試題的開發成本。如同前述，翻譯採用國外相關測驗試題，在適用上必須考量其在語言及文化上的限制，其過程並不見得比自行開發試題簡單省事。在本次測驗年會中，已有測驗機構基於防範窺探考畢試題等行為的安全成本日益增加，發表針對將公布之考畢試題以固定比例（如：20%）重新列入題庫，對於測驗成績之影響進行長期研究。目前經 6 年累積研究資料顯示，初步顯示其及格率並未有所改變。未來國家考試似可參照選擇採全程到考一定比例錄取之專技人員考試，試行將考畢鑑別度良好之試題，以較低比例重新列入考題，再觀察其對於考試效度之影響，以評估重新用運考畢試題之可能性。

十二、驗證電腦測驗應用的介面影響效度

由於電腦測驗具備許多優點，如：減少大量印刷題本及運送成本、可即時評分、立即進行統計分析，國家考試自 93 年開始推動航海人員及部分醫事人員類科考試採行電腦測驗，雖然宥於公布試題對題庫設計及試題校準之限制等因素，尚無法採行電腦適性測驗，近年來，亦逐步擴大電腦測驗適用的考試類科範圍及適用的試題題型。

目前以電腦環境施測的國際測驗，其最常應用的型態，仍以選擇題（單選或複選）、是非題等題型為主，相關研究就電腦測驗的施測形式對於應考人成績究竟有無介面效應，尚無法輕易下結論，國際測驗委員會在 2005 年出版的「電腦及網路傳送測驗國際指導手冊」中，在就電腦測驗在測驗方法論應注意事項中即敘明：應確認測驗的信度、效度不會受測驗傳送媒介影響；並注意電腦測驗的採行不需要與測驗無關的電腦技能或知識，以及注意測驗介面的改變不應影響受測者的能力表現。因此，在擴大電腦測驗應用範圍的同時，建議可參照國家華語測驗推動工

作委員會⁸，進行應考人在兩種施測介面—紙筆與電腦間成績表現的比較，確認應考人成績表現不會受到測驗介面的影響，以確保測驗之效度及公平性。

⁸ 詳細請參考國家華語測驗推動工作委員會網站 <http://www.sc-top.org.tw/chinese/publication.php>

參考文獻

- 林玲英、藍珮君。(2011)。華語文能力測驗電腦與紙筆測驗試題難度比較研究，2011 ACTFL 年會發表論文，美國丹佛。取自
<http://www.sc-top.org.tw/download/research/Comparability%20of%20Paper-and-pencil%20and%20Computerized%20Versions%20of%20a%20Chinese%20Proficiency%20Test.pdf>
- 楊亨利、應鳴雄。(2005)。線上測驗是否可能具備與紙筆測驗相同評分效力？*師大學報：教育科學類*，50 (2)，85-107。
- Bridgeman, B. (2009). Experiences from Large-Scale Computer-Based Testing in the USA. *In Scheuerman F. & Björnsson, J. (Eds.) The Transition to Computer-Base Assessment: New Approaches to Skills Assessment and Implication for Large-scale Testing.* (pp. 39-44) European Communities.
- Buško, V. (2009). Shifting from Paper-and-Pencil to Computer-based Testing: Requisites, Challenge and Consequences for Testing Outcomes. *In Scheuerman F. & Björnsson, J. (Eds.) The Transition to Computer-Base Assessment: New Approaches to Skills Assessment and Implication for Large-scale Testing.* (pp. 117-126) European Communities.
- British Psychological Society (2012). *Guidelines for the Development Computer-Based Assessment.*
- Carlson, F. & Geisinger, F. (2012). Test Reviewing at Buros Center for Testing. *International of Testing V12, N2, 122-135.*
- Chang, C. Y., (2001). *Cross-Culture Assessment: A Call for Test Adaptation.* Association for Assessment in Counseling. AACE Newsnotes Article. Available form

- <http://aac.ncat.edu/newsnotes/y99sum1.html>
- ETS (2009) . *Constructed-Response Test Questions: Why We Use Them; How We Score Them*. R&D Connections. N.11. Available form http://www.ets.org/Media/Research/pdf/RD_Connections11.pdf
- European Personnel Selection Office. (2010) *EPSO Development Program: Roadmap for Implementation*. Available form http://europa.eu/epso/doc/edp_11_2010.pdf
- European Personnel Selection Office (2012) . *Information and Notices. Official Journal of the European Union, 55*, Available form <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2012:055:FULL:EN:PDF>
- Edurola C. Casslar & Carla Quesada. (2012) . *A neural network predictive model to estimate future transfer of knowledge to the work environment*. 2012 ITC Commission essay.
- Enrique MERINO-Tejedor & Manuel Fernandez-Rios. (2012). *New Issues In the assessment of work design*. 2012 ITC Commission essay.
- Esther Care & Patrick Griffin. (2012). *Delivery of tasks online-creating tasks to indicate the constructs*. 2012 ITC Commission essay.
- Elosua, P. & Iliescu, D. (2012). Test in Europe: Where We Are and Where We Should Go. *International of Testing VI2, N2*, 157-175.
- Farcot, M. & Latour, T. (2009). *Transitioning to Computer-based Assessment: A Question of Cost*. In Scheuerman F. & Björnsson, J. (Eds.) *The Transition to Computer-Base Assessment : New Approaches to Skills Assessment and Implication for Large-scale Testing*. (pp. 108-116) European Communities.
- Geisinger, F. (2012). *Worldwide Test at Beginning of the Twenty-First*

- Century. *International of Testing V12, N2*, 103–107.
- Hagemeister, C, Kersting, M. & Stemmler (2012). Testing Review in Gemery. *International of Testing V12, N2*, 185–194.
- International Test Commission (2005) . *International Guidelines on Computer-Based and Internet Delivered Testing*.
- International Test Commission (2012) . *Design, Implementation and Evaluation of Assessment and Development Centres*.
- International Test Commission (2010) . *International Test Commission Guidelines for Translating and Adapting Tests*. Available form <http://www.intestcom.org/upload/sitefiles/40.pdf>
- International Test Commission (2012) . *ITC Guidelines for Quality Control in Scoring, Test Analysis, and Reporting of Test Scores*. Available form http://www.intestcom.org/upload/sitefiles/qcguidelines_2012-07-13.pdf.
- Iliescu, D. (2012). *Test adaptation: The strife for equivalence*. 2012 ITC Commission workshop essay.
- Janel F. Carlson & Kural F. Geisinger. (2012) . Test Review at Buros Center for Testing. *International of Testing V12, N2*, 122–135.
- Kikis-Papadakis, K. & Kollias, A., (2009). Reflections on Paper-and Pencil Tests to eAssessment: Narrow and Broadband Paths to 21st century Challenges. *In Scheuerman F. & Biörnsson, J. (Eds.) The Transition to Computer-Base Assessment : New Approaches to Skills Assessment and Implication for Large-scale Testing*. (pp. 99–103) European Communities.
- Kozma, R. (2009). Transforming Education: Assessing and Teaching 21st Century Skills Assessment Call to Action. *In Scheuerman F.*

- & Biörnsson, J. (Eds.) *The Transition to Computer-Base Assessment : New Approaches to Skills Assessment and Implication for Large-scale Testing.* (pp.13-23) Luxembourg: Office for Official Publications of the European Communities.
- Kyllonen, P.C. (2009). New Constructs, Methods, & Direction for Computer-Based Assessment. *In Scheuerman F. & Biörnsson, J. (Eds.) The Transition to Computer-Base Assessment : New Approaches to Skills Assessment and Implication for Large-scale Testing.* (pp.151-156) Luxembourg: Office for Official Publications of the European Communities.
- Lindley, A. & Bartram, D. (2012). Use of the EFPA Test Review Model by the UK and Issues Relating to the Internationalization of Test Standards. *International of Testing VI2, N2*, 108-121.
- Martin, R. (2009). Utilising the Potential of Computer-delivered Surverys in Assessing Scientific Literacy. *In Scheuerman F. & Biörnsson, J. (Eds.) The Transition to Computer-Base Assessment : New Approaches to Skills Assessment and Implication for Large-scale Testing.* (pp.172-177) European Communities.
- Nussbaum, M. (2012) . *Implementing a New Selection Model across 27 European Countries.* 2012 ITC Commission essay.
- Partnership for the 21st Century Skills. (2002). *Learning for the 21st Century : A Report and Mile Guide for 21st Century Skills.* Available form http://www.p21.org/storage/documents/P21_Report.pdf
- Patricia A. Lindley & Dave Bartram (2012) . Use of the EFPA Test Review Model by UK and Issue Relating to the Internationalization of Test Standards. *International of Testing VI2, N2*, 108-121.

- Paula Elosua & Dragos Iliescu (2012) .Test in Europe:Where We are and Where We Should Go. *International of Testing V12, N2*. 157-175.
- Scheuerman F. & Biörnsson, J. (2012) *The Transition to Computer-Base Assessment : New Approaches to Skills Assessment and Implication for Large-scale Testing*. Luxembourg: Office for Official Publications of the European Communities. Available form [http://publications.jrc.ec.europa.eu/repository/bitstream/11111111/8713/1/reqno_jrc49408_final_report_new\(1\)%5B1%5D.pdf](http://publications.jrc.ec.europa.eu/repository/bitstream/11111111/8713/1/reqno_jrc49408_final_report_new(1)%5B1%5D.pdf)
- Schroeders, U. (2009) .Testing for Equivalence of Test Data across Media. In Scheuerman F. & Biörnsson, J. (Eds.) *The Transition to Computer-Base Assessment : New Approaches to Skills Assessment and Implication for Large-scale Testing*. (pp.164-170) Luxembourg: Office for Official Publications of the European Communities.
- Sonnleitner, P. , Burnner, M., Keller, U., Martin, R., & Latour, T. (2011) *The Genetics Lab: A new computer-based complex problem solving scenario to assess intelligence*. EAPA-Congress presentation. Available from <http://www.ecpall.lu.lv/files/Philipp%20Sonnleitner%20et%20a1.pdf>
- Sonnleitner, P., Brunner, M., Greiff, S., Funke, J., Keller, U., Martin, R., Hazotte, C., Mayer, H., & Latour, T. (2012). The Genetics Lab: Acceptance and psychometric characteristics of a computer-based microworld ssuming complex problem solving. *Psychological Test and Assessment Modeling*, 54, 54 - 72.

The 8th Conference of the International Test Commission



Date	July 3-5, 2012 July 2, 2012 (<i>Pre-Conference Workshops</i>)
Venue	University of Amsterdam, The Netherlands
Theme	Modern Advances in Assessment: Testing and Digital Technology, Policies, and Guidelines
Sub-themes	<ul style="list-style-type: none">■ Testing and Digital Technology■ International Testing and Testing in Heterogeneous Populations■ Developments in Theory of Testing■ Professionalization, Guidelines, and Training in Testing■ Testing, Policy and Ethics Issues

THAT IS THE QUESTION:	Raadzaal	Bestuurskamer	Clauszaal	Emmazaal
08.00 - 17.00	Registration			
09.00 - 10.30	Workshop 3 <i>Testing measurement and structural equivalence across culture: Procedures, issues and complexities</i> Barbara Byrne	Workshop 2 <i>Evaluating skills: Fundamental concepts and skills for psychologist and researchers</i> Kurt Geisinger	Workshop 1 <i>Analyzing non-cognitive data with item response theory: What can we learn from modern test theory?</i> Rob Meijer & Iris Egberink	Workshop 4 <i>Modern advances in dynamic testing: Practical solutions to identified concerns</i> Aletta Odendaal & Marié de Beer
10.30 - 11.00	Coffee break			
11.00 - 12.30	Workshop 3 Continued	Workshop 2 Continued	Workshop 1 Continued	Workshop 4 Continued
12.30 - 13.30	Lunch			
13.30 - 15.00	Workshop 6 <i>Item response theory: Concepts, models, and applications</i> Ron Hambleton	Workshop 7 <i>Measurement matters: Applying psychological measurement principles in modern clinical testing and assessment</i> Gary Canivez	Workshop 8 <i>Practical R for everyday data analysis</i> Richard Morey	Workshop 5 <i>Test adaptation: The strife for equivalence</i> Dragos Iliescu
15.00 - 15.30	Tea break			
15.30 - 17.00	Workshop 6 Continued	Workshop 7 Continued	Workshop 8 Continued	Workshop 5 Continued

Tuesday, July 3, 2012

Time	Grote Zaal	Mauritszaal	Leeszaal	Bestuurskamer
08.00 - 17.00	Registration			
08.45 - 09.30	Opening ceremony (sponsored by NIP/GMAC/SHL) (Grote Zaal)			
09.30 - 10.30	State-of-the-Art Speech (Grote Zaal) Mike Fetzer: <i>The Evolution of Assessment: Simulations and Serious Games</i> Session chair: John Hattie			
10.30 - 11.00	Coffee/Tea break			
11.00 - 12.30	Invited symposium 1 <i>The Assessment of Critical Thinking: Cross-cultural and Validity Issues</i> Chair: Diane Halpern 2nd discussant: An Verburgh Heather Butler Kelly Y. L. Ku Vivian Miu-Chi Lun Amanda H. Franco	Special session 1 (11.00-12.15) Panel discussion : <i>The ITC Guidelines: Past, Present, Future</i> Chair: Dragos Iliescu Ron Hambleton Dave Bartram Iain Coyne Avi Alalouf David Foster Alina von Davier	Symposium 1 <i>Linguistic Diversity and Testing</i> Chair: Paula Elosua Discussant: Kurt Geisinger Guillermo Solano-Flores Edurado Ubieta-Muñuzuri, Kadriye Ercikan Maria Martiniello	Symposium 2 <i>IRT and computer based testing: some empirical applications</i> Chair: Rob Meijer Discussant: Wilco Emons Danny Brouwer Iris Egberink Annette Maij-de Meij, Jorge Tendeiro
12.30 - 13.45	Lunch			
13.45 - 15.15	Keynote 1 (13.45-14.45) <i>The exciting future of educational testing</i> Marten Roorda	Invited symposium 2 <i>Reflections on the International Journal of Testing: History, Influential Articles, and Getting Your Research Published</i>	Symposium 4 <i>Assessment of Linguistic Minority Students in Canada and the USA</i> Chair: Debra Sandilands Discussant: Kurt Geisinger	Symposium 5 <i>Testing Resources Help Promote Test Development and Use in Emerging Countries</i> Chair and discussant: Thomas Oakland

	Session chair: Ron Hambleton	Chair: Stephen G. Sireci Moderator and Discussant: John Hattie Stephen G. Sireci Rob Meijer April Zenisky	Debra Sandilands Kadriye Ercikan Guillermo Solano-Flores John W. Young	Hazel Wheldon Paul McKeown Dragos Iliescu Ilke Inceoglu
15.15 - 15.45	Coffee/Tea break			
15.45 - 17.15	Keynote 2 (15.45-16.45) <i>The challenge of setting standards for testing and test use in a rapidly changing online assessment environment</i> Dave Bartram Session chair: Sverre Nielsen	Invited symposium 3 <i>Dealing with Response Distortions in Applied Questionnaires</i> Chair: Olexander Chernyshenko Discussant: Pat Kyllonen Jonas P. Bertling Anna Brown Stephen Stark Olexander Chernyshenko	Symposium 7 <i>Reducing Bias and the Achievement Gap of Minorities in Selection Procedures in the Low Countries</i> Chairs: Johnny Fontaine & Eva Deros Discussant: Fons van de Vijver Veronique Verhees Britt De Soete Symen Brouwers Johnny Fontaine	Symposium 8 <i>Challenges of test adaptations in special contexts: The role of the ITC Guidelines</i> Chairs: Jose-Luis Padilla & Stephen G. Sireci Discussant: Kurt Geisinger Joseph A. Rios Itziar Alonso-Arbiol Anita Obrycka Isabel Benitez
17.00 - 18.00	Transport by canal boat from conference venue to welcome reception			
18.00 - 20.00	Welcome reception (sponsored by NIP/GMAC/SHL)			

Due to limited space, only the first authors of each presentation are mentioned; the full author lists are published in conjunction with the respective abstracts.

Tuesday, July 3, 2012 (continued)

Raadzaal	Emmazaal	Clauszaal	Ronde		Time
Registration					08.00 - 17.00
Opening ceremony (sponsored by NIP/GMAC/SHL) (Grote Zaal)					08.45 - 09.30
State-of-the-Art Speech (Grote Zaal) Mike Fetzter: <i>The Evolution of Assessment: Simulations and Serious Games</i> Session chair: John Hattie					09.30 - 10.30
Coffee/Tea break					10.30 - 11.00
Symposium 3 <i>Examinee motivation and secondary data analysis of large-scale assessment</i> Chair: Christina van Barneveld Discussant: Don Klinger Gul Shahzad Sarwar Christina van Barneveld Carlos Zerpa Christina van Barneveld	Oral paper session 1 <i>Test development</i> Chair: José Muñiz Olga Mitina Djurre Holtrop Marië de Beer Susan Davis-Becker Felicity Hill-Miers	Oral paper session 2 <i>Tests for challenging concepts</i> Chair: April Zenisky Joseph Kui Foon Chow Iris A.M. Smits Sara Ibérico Nogueira Sophie Brasseur Kostas Mylonas	Poster session 1 (10.45-12.45) Hayri Eren Suna Yoon Soo Park Jung Yeon Park Mohammad Asgari AbdelAziz Sartawi Liang-Cheng Zhang Ludwig Kreuzpointner Fraide A. Ganotice Meneses Alba Jia-Jia Syu Majid Akbari Kazuhiro Yasunaga Eleonora Nosenko Almut Thomas Hassan Gholami José Silva Ekaterina Orel Nicole de Jong Vytautas Laitusis Pierre Vrignaud		11.00 - 12.30
Lunch					12.30 - 13.45
Symposium 6 <i>Psychometric Properties of the Multiple Mini-Interview in Medical School Admissions</i> Chair: Don A. Klinger Discussant: Christina van Barneveld	Oral paper session 3 <i>Psychometric issues in test construction</i> Chair: José Muñiz Brad Ching-Chao Wu Peter M. Krueger Nancy Anderson	Oral paper session 4 <i>Different aspects of</i> Chair: Paula Elosua Lale Khorrandel Hanwook Yoo Lawrence M. Rudner Christine Hohensinn	Poster session 2 (13.30-15.30) <i>2a: Sponsored scholar session *</i> Sobia Aftab Ella Gift Banda Hudson F. Golino Ronnell B. King Khrystyna Rakhubovska Betty-Jean Usher-Tate <i>2b</i> Friederike Hohenstein Yang Wu Masoomah Khosravi Ramezan P.		13.45 - 15.15

Don A. Klinger King Luu Stefanie Sebok Don A. Klinger	Jacob Arie Laros Rudolf Debelak	Maria Elena Oliveri	Zornitsa Kalibatseva Jieting Zhang	Dowlati Claisy Marinho-Araujo Cynthia Bisnot Josu Mujika	
Coffee/Tea break					15.15 - 15.45
Symposium 9 <i>Assessment of Children's Temperament: A Summary of International Work</i> Chair: Thomas Oakland Discussant: Solange Wechsler Ricardo Primi Jan Ciecich Dragos Iliescu Thomas Oakland	Oral paper session 5 <i>The use of tests in work-related settings</i> Chair: Iain Coyne Oliver Zierke Rainer Kurz Thomas B. Sitser Elly Broos Ilke Inceoglu	Oral paper session 6 <i>Assessment of student performance</i> Chair: Aletta Odendaal Yesim Capa-Aydin Saw Lan Ong Lingbiao Gao Flaviu A. Hodis Richard P. Phelps	Poster session 3 (15.30-17.15) Jianping Xu Felipe Valentini Jacob Arie Laros Denise Balem Yates Elena C. Dieti Gary L. Canivez Christina M. Ravert Thierry Lecerf Sotta Kieng Krunoslav Matesic Anders Sjöberg		Demetrios Katsi Aristotle Ivan Rabelo Silvia Galli Hudson F. Golim Priska Hagmann Arx Ursula Kastner-Koller Jan van Bebber Tatjana Turilova Mišcenko
Transport by canal boat from conference venue to welcome reception					17.00 - 18.00
Welcome reception					18.00 - 20.00

* Sponsored by: *APA - Boom test uitgevers - Cubiks - GITP Research - Measured Progress - Meurs HRM - National Board of Medical Examiners - NOA - Pearson Assessment and Information - Prevue HR Systems - Profiles International - Schuhfried GmbH - TestGroup Consulting - Zeeuws Bureau voor Toegepaste Psychologie.*

Due to limited space, only the first author of each presentation is mentioned; the full author lists are published in conjunction with the respective abstracts.

Wednesday, July 4, 2012

Time	Grote Zaal	Mauritszaal	Leeszaal	Bestuurskamer
08.00 -	Registration			

17.00				
08.45 - 10.30		Invited symposium 4 <i>ITC Standards for Testing Individuals with Limited Proficiency in the Language of the Assessment</i> Chair: Alina Von Davier Discussant: Thomas Eckes Fons van de Vijver Stephen G. Sireci Alina Von Davier Thomas Oakland Avi Allalouf		Symposium 11 <i>Cooperating to enhance test security in international testing programs</i> Chair: John Fremer Discussant: Eugene Burke Aranka Krechtling Cor Sluijter David Foster John Fremer
09.00 - 10.30	Keynote 3 (09.00-10.00) <i>Current challenges of international student assessment</i> Eckhard Klieme Session chair: Aletta Odendaal		Symposium 10 <i>Developments in the Netherlands 1 (COTAN)</i> Chair and Discussant: Jan Henk Kamphuis Bas Hemker Wilco Emons Remko van de Berg Klaas Sijtsma	
10.30 - 11.00	Coffee/Tea break			
11.00 - 12.30	Invited symposium 5 <i>Network Psychometrics</i> Chair: Denny Borsboom Discussant: John Hattie Denny Borsboom Angélique O. J. Cramer Sacha Epskamp Verena D. Schmittmann	Special session 2 (11.00-12.15) <i>Test User Qualification in Europe: the work of the EFPA Board of Assessment</i> Chair: Dave Bartram Ana Hernandez Patricia Lindley Sverre Nielsen Eva Bergvall Lars Michaelsen	Symposium 13 (11.00-12.15) <i>Developments in the Netherlands 2 (COTAN)</i> Chair and Discussant: Arne Evers Wilma Resing Petra Hurks Tatjana van Strien	Symposium 14 <i>Cross-cultural Assessment of Psychopathology</i> Chairs: Frederick Leong & Zornitsa Kalibatseva Discussant: Thomas Oakland Zornitsa Kalibatseva, Frederick Leong Andrei Ion Eduardo Fonseca-Pedrero Barbara M. Byrne
12.30 - 13.45	Lunch			

13.45 - 15.15	<p>Keynote 4 (13.45-14.45)</p> <p><i>Recent Developments in International Testing</i></p> <p>Fons van de Vijver</p> <p>Session chair: Paula Elosua</p>	<p>Invited symposium 6</p> <p><i>The Combined Emic and Etic Method in Personality Assessment: Cross-cultural Relevance of the CPAI-2</i></p> <p>Chair: Fanny Cheung</p> <p>Discussant: Rob Roe</p> <p>Fred Leong Sarah Wan Marise Born Dragos Iliescu</p>	<p>Symposium 16</p> <p><i>Test Progress in Iberian Latin American Countries: trends and opportunities for international collaboration</i></p> <p>Chair: Solange Wechsler</p> <p>Discussant: Thomas Oakland</p> <p>Solange Wechsler Leandro Almeida Maria Pères Solis Margarida Pocinho</p>	<p>Symposium 17</p> <p><i>Cross-cultural comparability of noncognitive assessments – Construct and Method Issues</i></p> <p>Chairs: Jonas P. Bertling & Patrick C. Kyllonen</p> <p>Discussant: Johnny Fontaine</p> <p>Patrick C. Kyllonen Jia He Dave Bartram Jonas P. Bertling</p>
15.15 - 15.45	Coffee/Tea break			
15.45 - 17.15	<p>Invited symposium 7</p> <p><i>Advances in Assessment Theory and Practices</i></p> <p>Chair: April Zenisky</p> <p>Discussant: Ron Hambleton</p> <p>Paul DeBoeck Wim J. van der Linden Bruno Zumbo Ronald Hambleton</p>	<p>Special session 3 (15.45-17.00)</p> <p>Panel discussion : <i>ISO 10667: A Standard for Assessment in Work and Organizational Settings</i></p> <p>Chair: Dave Bartram</p> <p>Anders Sjöberg Marise Born Wayne Camara Sverre Nielsen</p>	<p>Symposium 19</p> <p><i>Short scales for psychological research - Applicability, benefits, and potential limitations</i></p> <p>Chairs: Christoph J. Kemper & Beatrice Rammstedt</p> <p>Discussant: Johannes Lutz</p> <p>Markus Zenger Stefan Schipolowski Matthias Ziegler Hilde Tobi</p>	<p>Symposium 20</p> <p><i>New developments of observed-score equating</i></p> <p>Chairs: Marie Wiberg & Alina von Davier</p> <p>Discussant: Kurt Geisinger</p> <p>Marie Wiberg Alina von Davier Anton Beguin Per-Erik Lyrén</p>
17.15 - 18.45	ITC-Meeting (Grote Zaal)			
19.30 - xxx	Conference dinner			

Due to limited space, only the first author of each presentation is mentioned; the full author lists are published in conjunction with the respective abstracts.

Wednesday, July 4, 2012 (continued)

Raadzaal	Emmazaal	Clauszaal	Rotonde		Time				
Registration					08.00 - 17.00				
Symposium 12 <i>International Perspectives on Test Reviewing</i> Chair: Kurt Geisinger Discussant: Dave Bartram Patricia Lindley Janet F. Carlson José Muñiz Carmen Hagemeister Paula Elosua	Oral paper session 7 <i>Differential Item Functioning</i> Chair: Jan Bogg Nedim Toker Tanya Yao Stephen Anwyll Ada Woo Katja Päßler David Magis	Oral paper session 8 <i>Test bias across cultures</i> Chair: Cheryl Foxcroft S. Kanageswari d/o Suppiah Shanmugam Van Nguyen Magdalena S. Halim Saima Ambreen Serkan Arikan Efi Fitriana	Poster session 4 (8.45-10.45)		08.45 - 10.30				
			Eunice Alencar Sara Ibérico Nogueira Margarida Pocinho Emils Kalis Leonor Almeida Maiana Nunes Mercedes Liporace	Lucas Guimarães Ahmad Shahidov Sergio Zavarize Micheline R. Bastianello Cristian Zanon Alessandra Coscarelli Katharina Lochner Paola Bully (1)					
	Coffee/Tea break					10.30 - 11.00			
	Symposium 15 <i>Current issues in gifted children identification and assessment</i> Chair: Jacques Grégoire Discussant: Pending Todd Lubart Sophie Brasseur Catherine Cuche Jacques Grégoire	Oral paper session 9 <i>Computer Adaptive Testing & Testlet research</i> Chair: April Zenisky Anna Brown Lihshing Leigh Wang Kyung T. Han Luc T. Le Muirne C. S. Paap	Oral paper session 10 <i>New developments in work related testing</i> Chair: Iain Coyne Harald Ackerschott Aylin Dincer Eduardo C. Cascallar Enrique Merino-Tejedor Markus Nussbaum	Poster session 5 (10.45-12.45)		11.00 - 12.30			
				Tanja Krone Thomas Proctor Rudolf Debelak Brad Ching-Chao Wu Tzu-Yun Chin William Durán Josu Mujika Hadassah Weitzman B. Carlos Nunes (1) Carlos Nunes (2)			Pei-Chun Shih Matteo Ciancaleoni Po-Hsi Chen David Arribas-Aguila Ji Eun Lee Matthias Stadler Iryna Arshava Eduardo Doval Andress P. Kooij Lisabet M. Hortensiu Paola Bully (2)		
		Lunch					12.30 - 13.45		
		Symposium 18	Oral paper session 11	Oral paper session 12			Poster session 6 (13.30-15.30)		13.45 -

<i>So Short! That Valid?</i>	<i>The use of tests in educational settings</i>	<i>Item design and test performance</i>	Richu Wang Jelena Kolesnikova Krunoslav Matešić, j Nelson Hauck-Filho Tove K. Marberger Jorge Costa Joice Dickel Segabin Roselaine Silva (1) Roselaine Silva (2) Eduardo Fonseca- Pedrero Dina Horowitz	Pia Deimann Rebecca Gokiert V. Prado-Gasco Juliane Borsa Jan Ciecich (1) Jan Ciecich (2) Zlatka Cugmas Eliezer Gums Yung S. Lee Damascen Susanna S. Yeung	15.15
Chair: Rab MacIver Discussant: Helen Baron Matthias Burisch Rab MacIver Rainer Kurz Tom Hopton	Chair: Aletta Odendaal Eileen Talento-Miller Janet Clinton Debra Masters Saad Chahine Hanna Eklöf	Chair: Sverre Nielsen Leanne Ketterlin-Geller Mariel F. Musso Claire Stevenson Nicolas Becker Maria Araceli Ruiz-Primo			
Coffee/Tea break					15.15 – 15.45
Symposium 21 <i>The Hierarchical Structure of Personality and Performance</i> Chair: Rainer Kurz Discussant: Jörg Prieler Rainer Kurz Dimitri van der Linden Paula Cruise Rab MacIver	Oral paper session 13 (15.45–17.30) <i>Faking and test security</i> Chair: Jacques Grégoire Katharina Lochner Cristina Anguiano-Carrasco Mireia Ruiz-Pàmies Chad W. Buckendahl Steve Addicott Dev Ramnane	Oral paper session 14 <i>Test development around the world</i> Chair: Jan Bogg Andrew A. Mogaji Cheryl Foxcroft Elena Brenlla Paulo Roberto Teixeira Linda Platas	Poster session 7 (15.30–17.30) Nicole Hirschmann Vivian Lago Lucas Guimarães Elisa M. P. Yoshida Kai Welzen Jean Natividade (1) Jean Natividade (2) Pablo Santamaría Ioana R. Cocia		15.45 – 17.15
			Zheng Zhang Flaviu A. Hodis Maria Angela Louren- çoni Kostas Mylonas Katja Schlegel Fraide A. Ganotice, Pedro Portes Maria Anna Donati		
					17.15 – 18.45
Conference dinner					19.30 –

Due to limited space, only the first author of each presentation is mentioned; the full author lists are published in conjunction with the respective abstracts.

Thursday, July 5, 2012

Time	Grote Zaal	Mauritszaal	Leeszaal	Bestuurskamer
08.00 - 14.00	Registration			
08.15 - 10.00				Symposium 23 <i>Monitoring the quality of performance assessment raters</i>
08.30 - 10.00	Keynote 5 (08.30-9.30) <i>Engineering the Design of Assessment Task Models and Templates</i> Ric Luecht Session chair: Wim van der Linden	Invited symposium 8 <i>Attitudes of Psychologists on Tests and Testing: The Results of an International Survey</i> Chair: Arne Evers Discussant: José Muñiz Solange Wechsler Saul Fine Carina McCormick Arne Evers	Symposium 22 <i>Personality assessment across and within cultures: Addressing some methodological issues</i> Chair: Ilke Inceoglu Discussant: Fons van de Vijver Anne Herrmann Mathijs Affourtit Dave Bartram Marise Ph. Born	Chairs: Avi Allalouf & Alvaro Arce-Ferrer Discussant: George Engelhard Jr. Alvaro J. Arce-Ferrer Avi Allalouf Jo-Anne Baird Iasonas Lamprianou George Engelhard, Jr.
10.00 - 10.30	Coffee/Tea break			
10.30 - 12.00	Invited symposium 9 <i>Psychometric modelling and the management of response biases in personality assessment</i> Chair: Anna Brown Discussant: Dave Bartram Luis Eduardo Garrido Natalia Kieruj Matthias Ziegler	Special session 4 (10.30-11.45) <i>Consequences of the results of the European survey on test attitudes</i> Chair: José Muñiz Giulia Balboni Alessandra Coscarelli Carmen Hagemeister Andrei Ion	Symposium 25 <i>Is the end in sight for traditional psychometric testing?</i> Chairs: Janwillem Bast & Nadine Schuchart Discussant: Mark Schittekatte Angélique O. J. Cramer Dylan Molenaar	Symposium 26 <i>Paper-Pencil and Computerized Testing: Practical matters</i> Chair: Paula Elosua Discussant: David Foster David Arribas-Aguila

	Anna Brown	Lars Michaelsen	Paul Oosterveld H. Steven Scholte	Dragos Iliescu Annemarie M. F. Hiemstra Sarah L. Hagge
12.00 - 13.15	Lunch			
13.15 - 14.15	State-of-the-Art Speech (Grote Zaal) <i>Test Validation for 21st Century Tests</i> Stephen G. Sireci Session chair: José Muñiz			
14.15 - 14.30	Closing ceremony (Grote Zaal)			

Due to limited space, only the first author of each presentation is mentioned; the full author lists are published in conjunction with the respective abstracts.

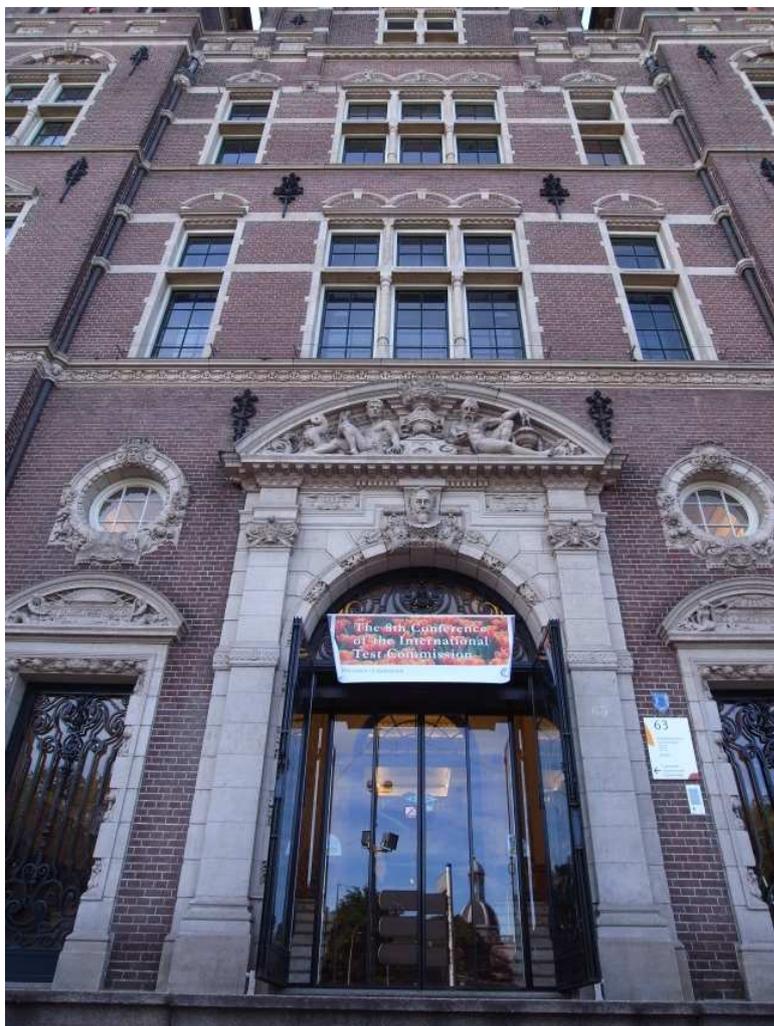
Thursday, July 5, 2012 (continued)

Raadzaal	Emmazaal	Clauszaal	Rotonde		Time
Registration					08.00 - 14.00
Symposium 24 (08.30-10.00) <i>Advances in Testing and Measurement: Job Performance, Change and Innovation</i> Chairs: Neil Anderson & Kristina Potocnik Discussant: Neil Anderson Jesús Salgado (canceled) Gouras Athanasios, Kristina Potocnik Rob Roe	Oral paper session 15 <i>Standard setting and calibration</i> Chair: Paula Elosua Stefan Bondorowicz Çiğdem Güzel Klaus D. Kubinger Abdullah Ferdous Rosemary Reshetar Gad S. Lim	Oral paper session 16 <i>New methods of testing</i> Chair: Sverre Nielsen Phillipp Sonnleitner Jeroen Kuntze Alec W. Serlie Mark Schittekatte Dustin Wahlstrom Yehuda Esformes	Poster session 8 (8.15-10.15)		08.15 - 10.00
			Rainer Kurz Paula Cruise Erdogan Gultekin Helle Pullmann Louisa Tate Maria Veltmann (1) Maria Veltmann (2) Haniza Yon (1) Margreet Kerkmeer Chenxiao Tang	Haniza Yon (2) Morteza Taheri Janneke Oostrom Elke Ve V. Prado-Ga Cornelia Moll Andreea Butuces Dmitry Leontie Heidi N. Keiser Maiana M	
Coffee/Tea break					10.00 - 10.30
Symposium 27 <i>The assessment and Teaching of 21st Century Skills</i> Chair: Patrick Griffin Discussant: Martina Roth Patrick Griffin Arto Ahonen Diederik Schonau	Oral paper session 17 <i>The consequences of feedback, coaching and (re)testing for test scores</i> Chair: Jacques Grégoire Maïke Wehrmaker Anely Ramirez Nicholas Marosszeky	Oral paper session 18 <i>Test fairness</i> Chair: Cheryl Foxcroft Steve Dept Magda Chia Alvaro J. Arce-Ferrer Guillermo	Poster session 9 (10.15-12.15)		10.30 - 12.00
			Björn Johansson MingJie Zhou Helle Pullmann Joaquin Caso Niebla José Miguel Juliana Pacico Enrique Merino- Tejedor Enrique Vila Abad	Bruno Damásio Bruno Damásio Sofia Sjöberg Malgoza Rascevs Scott R. (1)	

Esther Care	Yongwei Yang Dorothea Klinck	Solano-Flores Jelte M. Wicherts		Scott R. (2) Olga Mi Michael Christo Haniza	
Lunch					12.00 - 13.15
State-of-the-Art Speech (Grote Zaal) <i>Test Validation for 21st Century Tests</i> Stephen G. Sireci Session chair: José Muñiz					13.15 - 14.15
Closing ceremony (Grote Zaal)					14.15 - 14.30

Due to limited space, only the first author of each presentation is mentioned; the full author lists are published in conjunction with the respective abstracts.

附錄二、第八屆國際測驗年會照片



2012 年 ITC 會議會場



會議報到



年會開幕式（阿姆斯特丹市副市長 Lodewijk Asscher 先生代表市長出席致辭）



開幕表演



開幕演講



工作坊(Work shop)5 : Test Adaptation: The strife for equivalence



會議研討(一)



會議研討 (二)



會議研討(三)



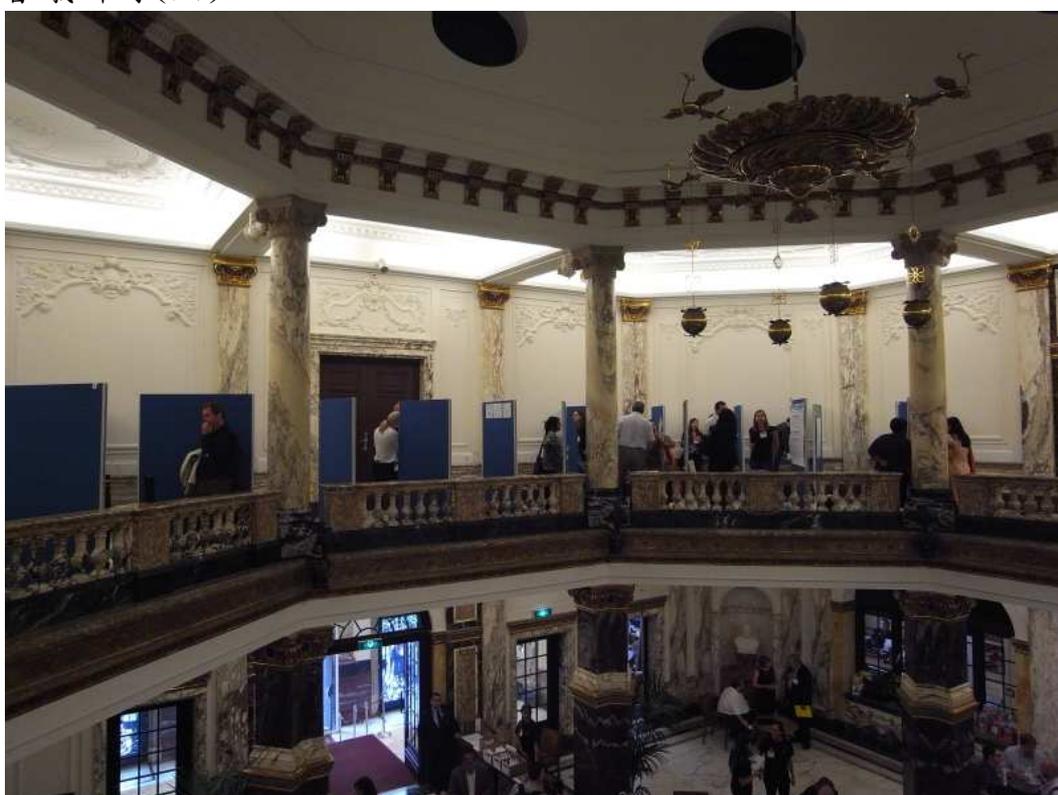
會議研討(四)



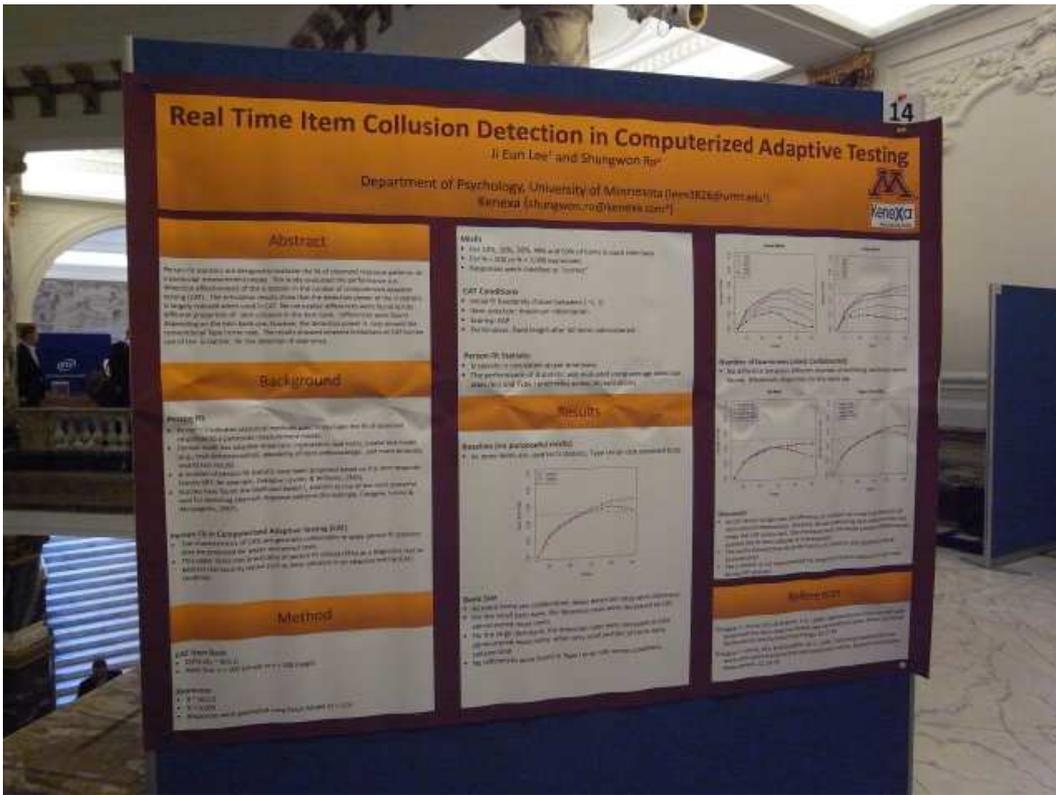
會議研討(五)



會議研討(六)



海報展覽(一)



海報展覽(二)



贊助商展覽(一)



贊助商展覽(二)



茶敘



參加迎賓晚會--與國內學者及新加坡國防部與會人員合影



與北京師範大學心理學系學者合影



黃專門委員立賢與張專門委員麗雪於會場合影



閉幕式