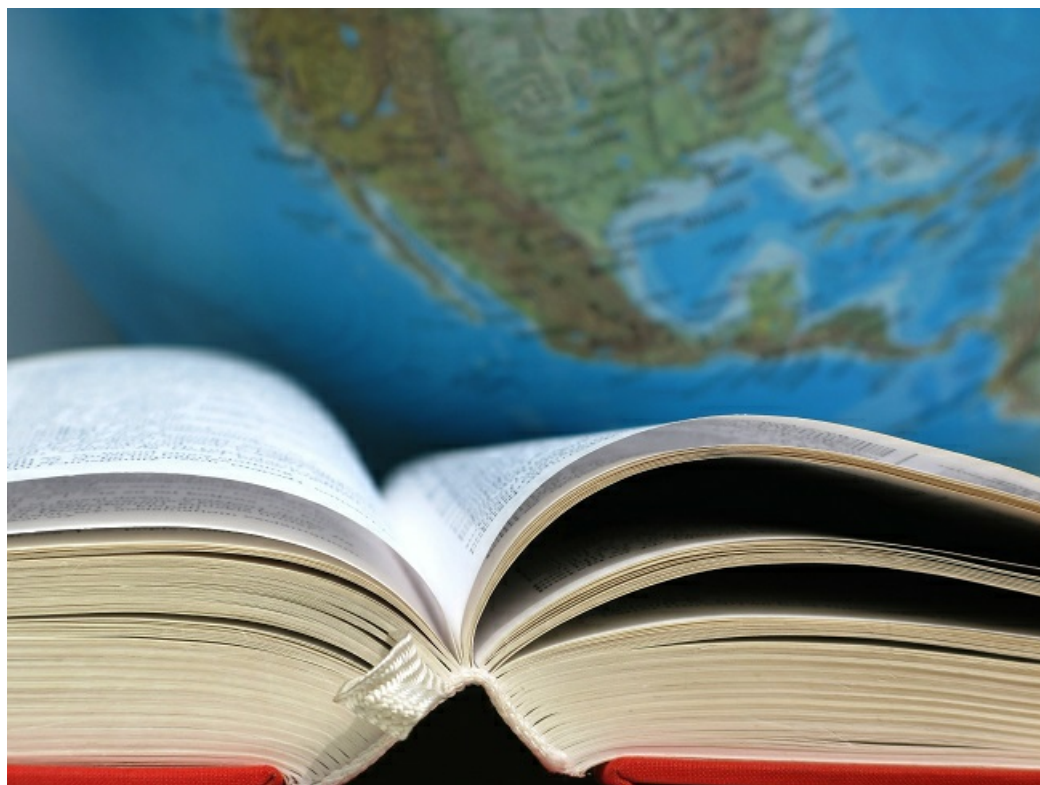


21世紀評量的發展趨勢

曾建銘

國家教育研究院測驗及評量研究中心，測驗及評量組



壹、議題重要性

全球化競爭的來臨，在20世紀末，澳洲、聯合國教科文組織、歐盟、美國等國家或國際組織，先後皆提出對於人民或學生應習得能力的研究，如1992年澳洲提出的七項關鍵能力、1996年聯合國教科文組織提出終身學習四大支柱（已於2003年修訂為五大支柱）、2005年歐盟提出終身學習八項關鍵能力、2003年美國也提出21世紀能力(張鈿富等，2009)。由於傳統教育最重視的3R能力，即「讀（reading）、寫（writing）、算（arithmetic）」，在工作應用上已日漸不足，因此各國紛紛提出4C及ICT方面的能力，所謂4C即「批判性思考與問題解決（critical thinking and problem solving）、有效溝通（effective communication）、團隊共創（collaboration and building）、創造與創新（creativity and innovation）」（黃子璿，2010）。

隨著科技的日新月異，全球環境也跟著改變，OECD結合來自全球60個研究機

構，超過250位專家學者共同提出了「21世紀需要被評量與教導的技能」，包括一、思考的訓練方法：包括創意、批判思考、問題解決、決策、學習的能力；二、工作的方法：溝通與合作的能力；三、操作工具的方法：運用資訊科技與網路素養；四、生活的技能：公民素養、生涯發展、個人與社會責任（ATCS，2014）。蘋果、思科、微軟、戴爾電腦等公司，與美國教育部，共同創立了21世紀關鍵能力聯盟（P21, Partnership for 21st century skills），認為21世紀關鍵能力為：集體合作的創造力與創新力、有效的社交溝通與合作能力、網路資訊的整理與分析、媒體素養、跨文化的合作技能。而根據美國國家科學院定義21世紀學習所須具備的能力，可簡單區分為三大類：認知技能（Cognitive skills）：如批判思考、分析推理（analytic reasoning）等；人際關係技能（Interpersonal skills）：如團隊合作和溝通能力等；個人內在技能（Intrapersonal skills）：即自我表達。包含自我覺察的反省能力及誠實耿直等，此與擁有好職業收入和健康生活形態息息相關。為了了解是否達成21世紀關鍵能力，21世紀評量有何因應？是本文擬探討的。

貳、各國主要具體作法與評量趨勢

在過去的20年中，評量發揮了美國教育政策核心的作用，就如同過去幾十年它也在其他國家一樣。例如大規模、總結性評量，被視為有力影響學校和教室所發生的事情，正因為如此，定期進行評量研究，以了解學生的長處和短處。此外，2001年「沒有孩子落後的法案(No Child Left Behind Act)」中提及，測驗已更頻繁，也越來越有影響力，並專注於核心內容領域。從大規模總結性評量的結果，以及其他成就測驗，經常被用來決定學生是否可以進到下一個年級，以及判斷學校和教育工作者的品質。

21世紀的學習標準強調的是，學生如何能夠運用將所習得的學習內容知識，運用到有用的工作上時，能夠展現的能力。而不是傳統的讓學生用同樣速度和順序學習，獨立完成作業，接著考試檢驗成效。若按傳統教學，學生和老師幾乎沒有多少時間可深入主題與合作探究議題，以達成21世紀的關鍵能力。評量學生的能力和知識，以及讓學生、老師知道，他們是否朝向21世紀的學習目標前進，是非常重要的。

過去的學習標準和評量，都只著重背誦大型考試中可能會考的內容。這些考試因將決定學生未來的學習和就業之路，造成學生家長與老師莫大的壓力，而且常常被誤用來評估學校與教師的教學品質。太過強調總結性測驗—評量學習成果，使得大

家忽視形成性測驗的重要性。相較於總結性測驗，形成性測驗對學生與老師而言更有價值，因為形成性測驗可即時提供回饋，老師可迅速調整教導方式，滿足學生當下的需求，以免形成迷思概念或產生學習落後的情況。

21世紀的評量應能針對學生某一科目的學習進度或關鍵能力的學習成果提供有用的回饋，並評量更大範圍的能力與才能，全面性地反應學習者的學習成果。因此，21世紀的評量需要更好的總結性與形成性測驗，來評量所有的學習成果，包括內容知識、基本技能、高階思考能力、深度理解與應用知識，以及21世紀關鍵能力的運用。另也需發展許多不同的即時形成性測驗，來評估學生的內容知識了解程度、理解力、基本與高階思考能力和21世紀關鍵能力的表現。如美國大學與準備度測驗，學生需運用研究報告、預算編列和其他文件，來解決一個複雜的問題，例如如何管理因人口成長而產生的交通阻塞。

有些評量工作需要大量的人力來進行，而科技化的測驗可將這些評量工作自動化，並提供一套新方法來評估學生的技能表現，特別是根據現實情況模擬設計的問題。

以下提供一些21世紀評量的趨向：

一、21世紀技能的評量與教學(Assessment & Teaching of 21th century skill, ACT21S)

ATC21S是一個國際性的研究工作，由Cisco, Intel and Microsoft等著名公司贊助，參與國家包含澳大利亞，芬蘭，新加坡和美國，另哥斯達黎加和荷蘭也加入該計畫，以幫助了解語言和文化對21世紀技能評量的差異影響。目的在增強學生使用正確的技能，使其能在21世紀的工作場所成功。

為了改變課堂水準，政策制定者需要有準確了解學生群體技能的訊息。透過評量蒐集該數據是一個重要關鍵的部分。ATC21S正在開發的方法來評量技能，來形成21世紀課程的基礎，其重點放在溝通和合作、問題解決、公民權和數字的流暢性。ATC21S的目的是提供21世紀課程用於教育系統的建議，以支持提高勞動力。我們的目標是創建一個具有教學和學習資源的新評量架構，幫助學生培養21世紀技能。轉換這些技能到課堂中，將塑造未來幾年國家社區的經濟和社會發展。ATC21S的成功取決於各國政府廣泛政府間組織（IGO）、學術界和工業界等公共和私營部門的合作。國際ATC21S計畫包括五個階段，將建立世界各地學生21世紀技能之評量和教學的基礎(ACTS, 2014)。

二、結合認知診斷與內容標準提升學生學習

CBAL (Cognitively Based Assessment of, for, as Learning) 以認知診斷為基礎的評量不只是一種測量學生的形成性 (formative) 與總結性 (summative) 評量模式，並且是一種幫助學生學習與教師教學的教育評量行動理論模式。此模式架構的構成要素包括1、特定領域素養 (competency) 模式 (包含一系列的學習)、2、跨年度與整合績效責任的總結性評量、3、教室任務、延伸活動、教師指導及解釋性素材的形成性評量、4、幫助教師了解、實施CBAL的形成性與總結性評量及幫助教師發展知識與技能的專業支持。配合構成要素的假設行動機制有11個，以達到預期的中介與最後效果，詳請可以參閱Bennett (2010)。

三、模擬、實地演練以及追蹤學生解題的想法

從早期電腦讀卡的紙筆測驗到1980—1990年代以表現為基礎的測驗，然後又回到以大規模選擇的紙筆測驗，但都無法符合學校教育與幫助學生真正的學習。但由於認知科學與電腦技術的日新月異及具有功能強大的計算技術的結合，產生了新一代的學生測驗—建立利用電腦來對選擇題與建構開放題計分，同時又可提供嚴謹與可信的學生學習證據，因此顯著地改進教師教學與增進學生學習。開發模擬真實世界情境的試題讓學生進行解題，美國NAEP自2003年春天即開始在豐富科技環境下之解題計畫 (Problem Solving in Technology-Rich Environments project – TRE)，此計畫說明科技測驗的特性，此計畫可以提供學生更複雜多步驟的問題，還包括多媒體及動畫等。學生亦可藉此採用不同的方法與多種解題法，藉由電腦記錄學生解題活動與過程。評量者可以透過這些資訊來計分，以了解學生真正學習多少，優缺點如何？而非單一的對與錯。

另可同時致力於測驗實質內容之提升，如多媒體試題、聽力測驗、互動式測驗、情境式測驗等，以發揮紙筆測驗所不能及之處，並加速推動申論題、繪圖題等電腦作答介面，整合後續人工電腦閱卷作業，除提供應考人較熟悉且便利之作答介面外，並可有效提升評分信度，開創更多資訊科技所帶來之加值服務。

四、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：

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），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。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。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>

五、哈佛大學教育研究所虛擬表現評量計畫（Virtual Performance Assessment Project；VPA）

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能力，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，受測者透過虛擬代理人（avatar）在預設的環境問題中、觀察、進行資料蒐集，以解決科學問題。

六、21世紀關鍵能力的未來學校

當21世紀學生所應具備的關鍵能力，從傳統的专业學科領域知識轉變成4C能力時，不管是老師的教學方式或者是學校的環境，皆須依照教學目標的轉變而調整，才能夠有效的培養出具有國際競爭力的學生。有鑑於此，美國史丹佛國際研究中心

(SRI International)提出了一個幫助學校、地區和社群的發表創新性意見的工具：school 2.0。希望透過school 2.0，協助學校、地區與社群發展一個共同的教育願景，運用科技來協助願景的未來與探索，創造符合現今教學目標的教學生態環境，以培育具有全球競爭力的人才。School 2.0提供一個學習生態系統（Learning Ecosystem），清楚描繪出一個能夠培養21世紀能力的校園藍圖，這個系統中包含四大元素，分別是：一、學校基礎科技建設（infrastructure），二、政策規劃與實施（planning and implementation），三、學校相關的人員參與（stakeholder），四、學習環境營造（Learning Environment）。

在政策規劃與實行中，則必須不斷地透過各種數據資料的確認分析做為績效導向的管理、持續領導學校進行專業性發展及如何多方（如社區）合作幫助學生成功學習、在學校環境、設備建置、學生評量上規劃各種專業技術支持、以及透過實行嚴謹的學習標準與評估系統來幫助學生與學校持續不斷的進步。在學校相關人員中，則包含了校長、行政人員、老師、技術人員、學生、家長、以及社區人員。由此可以了解，在這個持續改變的世代中，為了培養學生擁有生存之必備能力，必須要由家庭、學校以及社區的共同合作，才能夠成功（張瓊方，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、必須釐清哪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，在評量結果釐清哪些是主體知識，哪些是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。

參、我國現況概述

目前我國除了，升學考試分發的大學學科能力測驗、大學指定考試、四技二專聯合甄試入學考試、國中會考外，還有臺灣學生學習成就評量、補救教學篩選考試與不定時參加的國際大型測驗，其餘則是學校的平時考、定期考試等。除國際大型測驗外，其餘大都屬於成就測驗，而測驗內容則大部分屬於知識記憶、理解與運用等層次的課程內容。

對於線上合作學習解題、虛擬表現評量與具即時性回饋的形成性評量，目前可能皆止於研究階段，對於像國外利用資訊科技大規模開發，進行實際電腦測驗評量結果分析，尚有一段距離。

肆、對我國的啟示或建議

根據以上國際對於21世紀所提出的關鍵能力與評量趨勢，提供以下建議：

一、根據課程目標建立一系列科目的各級標準是首要工作。因應21世紀關鍵能力—集體合作的創造力與創新力、有效的社交溝通與合作能力、網路資訊的整理與分析、媒體素養、跨文化的合作技能。參考國際對於21世紀所提出的關鍵能力結合課程目標來建立一系列科目的各年級標準是亟需進行的工作。

二、結合認知診斷與內容標準，以提升學生學習與教師教學的測驗，是可行與發展趨勢。利用所訂出的標準於教學中進行形成性評量，提供回饋給學生與教師，以讓教師因應學生學習情況隨時調整教學方式與進度。

三、建立高階互助合作解題的電腦總結性評量，來評估學生的內容知識了解程度、理解力、基本與高階思考能力和21世紀關鍵能力的表現，也是必要的。即我們必須從主要測驗離散知識，轉移到衡量學生的批判性思維能力、研究問題、蒐集信息，並同時採用技術做出明智的、理性的決策。

四、透過測驗資料庫的建立，監控學生的學習成長歷程，利用這些大數據資料分析幫助政策決定與課程改進，是政府可以思考建立的方向。

由於評量是推動其他教育支援系統的動力，許多全國性和全球性的計畫，都在著手設計一套屬於21世紀的評量，新的評量方式將在總結性與形成性測驗中取得平衡，並能夠真正反映當代所需要的深度了解和能力表現。21世紀的評量希望能帶給學生全人教育，使孩子們在認知、情意、社交或道德上，都能得到充分的發展，教導出健康、安全、有熱忱、享有充分資源，能夠正面迎接挑戰的學生。

參考文獻

(駐洛杉磯文化組吳迪珣／藍先茜摘譯) <http://yumi26.pixnet.net/blog/post/37037237>

美國國家科學院定義21世紀學習所須具備的能力，2012年7月10日，教育週刊，

取自：[http://blogs.edweek.org/edweek/inside-school-](http://blogs.edweek.org/edweek/inside-school-research/2012/7/study-21st-century-learning-demands-mix-of-abilities/)

[research/2012/7/study-21st-century-learning-demands-mix-of-abilities/](http://blogs.edweek.org/edweek/inside-school-research/2012/7/study-21st-century-learning-demands-mix-of-abilities/)

張鈿富等（2009）。歐美澳公民關鍵能力發展之研究。台北：國立教育資料館。

張瓊方（2011）。從21世紀關鍵能力淺談未來學校。數位典藏與學習電子報，10（1）。

教育部（2013）。教育部電子報576期，取自http://epaper.edu.tw/windows.aspx?windows_sn=13354

黃子璿（2010）。從3R到4C：淺談21世紀能力的發展與趨勢。數位典藏與學習電子報，9（11）。

ACTS(2014).*Assessment & Teaching of 21th century skill*, ACT21S. Retrieved from <http://atc21s.org/index.php>

Bennett, R. E. (2010). Cognitively Based Assessment of, for, and as Learning (CBAL): A Preliminary Theory of Action for Summative and Formative Assessment. *Measurement: Interdisciplinary Research and Perspectives*, 8:2-3, 70-91.

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.

PARCC(2013).*Assessment Administration Guidance*. Retrieved March 10, 2014
from <http://www.parcconline.org/>