

創造力的測量與共識衡鑑

邱皓政

【作者簡介】

邱皓政，美國南加州大學University of Southern California哲學博士，主修心理計量與統計學。研究興趣為統計方法與應用技術、心理測驗學、創造力研究、組織創新、社會與人格心理學等，專長議題為結構方程模式與多變量統計方法。目前為輔仁大學心理學系副教授、台灣統計方法學學會理事長、中國測驗學會常務理事、國立政治大學創新與創造力研究中心研究員，並任教於國立交通大學、國立台灣師範大學等校。歷經美國加州大學洛杉磯分校（UCLA）神經醫學研究中心統計分析師、教育部訓育委員會助理研究員，台灣心理學會秘書長。

摘要

本文的目的在回顧心理與教育理論與實務上所使用的創造力測量方法與相關議題，並介紹當代普遍使用的共識衡鑑策略。傳統以來，創造力的研究多延續智力的觀點，以一種「能夠創造出新奇有價值的產品的能力」來定義創造力的內涵，因此著重於發散性思考的測量。另一個主流則是以個別差異的觀點，藉由高創造者與一般個體的鑑別比較，找出能夠反應創造力高低不同的特質因素加以測量。此外，有關行為經驗的測量，或是歷史人物的研究，也能夠反應創造力的具體成果，或是以他人評定的方法，來針對創造性的個人或產品進行評鑑。本文除了針對不同的測量典範進行介紹，也特別針對創造力測量的相關議題進行討論。

關鍵詞彙：創造力、測量、心理測驗、共識衡鑑

Measurement and Consensus Assessment of Creativity

Hawjeng Chiou

Department of Psychology

Fu-Jen Catholic University

ABSTRACT

This present article is to review the assessment methods and measurement issues of creativity in psychological research as well as educational practices. Traditionally, creativity that defined as ‘the ability to create novel and value products’ is assessed as cognitive efforts and/or intellectual abilities. Measures of divergent thinking, for example, is the best known creative constructs for creativity. Another main stream of study of creativity is based on the individual differences, by focusing on the assessment of personality traits and comparisons among high creative people. In addition, creativity could be inferred by past experiences, behavioral observation, or eminence persons. This article also introduce a common used strategy, consensus assessment technique (CAT), for evaluating creative products. Finally, the significances as well practical issues about CAT along with the recent development on psychometric methods were discussed.

Key Words: creativity, measurement, psychological testing, consensual assessment

壹、前言

創造力 (creativity) 是什麼？是一種潛在的心智能力 (mental abilities)，還是能夠被客觀評定的一種社會產物 (social products)？這些問題在學術界已經爭執了將近一甲子，雖然學者對於這個問題逐漸獲得共識，但是仍然顯得意見紛歧，各有偏重。Mayer (1999) 指出，雖然花費了50年的時間，學者們最後多以「能夠創造出新奇有價值的產品的能力」來界定創造力，但這種觀點顯然太過於狹隘，沒有能夠對於創造力的實踐有更具體的幫助，因此呼籲學者繼續以更廣泛、寬闊的觀點，更多元的研究與測量方法來探討創造力是什麼。尤其是到了2005年的今天，研究者紛紛對於人類的正向特質感到高度興趣，作為正向心理學 (positive psychology) 當中不可或缺的一環的創造力研究，其測量議題也受到學者的高度重視 (Kerr & Gagliardi, 2003)，儘管如此，從定義到測量，創造力的各層面仍充滿有待釐清的問題 (Runco, 2004)，尤其是當我們希望能夠落實研究成果，對於實務領域產生具體的作用時，創造力的「捕捉」，也就是創造力的評量與衡鑑，成為定義與內涵問題以外，更值得探究的議題。

一、捕捉創造力：測量與衡鑑

測量與衡鑑可以說是非常相似，但是層次不同的兩種概念。所謂測量 (measurement) 是利用一套客觀程序與符號系統 (測量尺度)，來描述被測對象在某一個被測特質的個別差異，其關鍵在於標準化與數量化。而衡鑑 (assessment) 則是指利用多重管道所得到的多種資訊，來評估某個對象在某一特質的狀態與意義。多重管道可能是多位評定者或多種工具，而各種資訊則包括了主觀或客觀程序所得到的質化或量化資料，例如問卷、測驗、訪談、日記、檔案資料，其關鍵在於多元性與整合性。如果測量與衡鑑涉及了好壞對錯的價值判斷或評估，又可稱為評量 (evaluation)。

在科學研究當中，測量扮演著重要的角色。因為如果沒有精確的測量，那麼所有的研究數據即失去客觀分析的基礎。尤其為了追求真相，科學的測量較一般生活上的各種度量有更高的要求，因此具備客觀性與標準化的測驗工具能夠用來評估個體或產品的創造力，成為研究者主要的依據。但是如果是實務工作者，對於創造力的瞭解是為了能夠具體解決問題，藉以評估個體或產品的各個面向，得到一個全面的瞭解與評估。因此宜採衡鑑的觀點來看待創

造力。本文在介紹創造力的評估方式時，所著重的是測量方法，但也希望能夠涵蓋衡鑑的概念，強調多重測量方法的優點，以彌補測量程序的缺點與不足之處。

二、測量的原則

測量的一個基本原則，是要有具體的對象與內容。當我們想要瞭解一個人的創造力高不高時，測量的對象可能是這個「人」本身，也可能他的「作品」，此時對象即有不同。另一方面，測量的內容、方法與程序，則需視創造力的客觀定義，也就是說，研究者必須先能清楚的說明到底所要測量的對象為何，才可能具體的提出測量的方法。例如一個人身高之所以能夠被測量，是因為我們定義了身高是指在地表脫鞋不戴帽，從腳底到頭頂的垂直距離；一個人的年齡是指出生到現在為止的年紀，如此定義就沒有實歲與虛歲問題。

在心理學領域，由於所測量的概念多為潛在的構念（constructs），此時抽象構念是否能夠被具體的測量，端視研究者如何對於該構念進行操作性的定義，由操作型定義，即可決定測量與衡鑑的對象與程序。研究者除了能夠從理論邏輯層面來界定所欲測量的對象的性質與內涵之外，還需言明如何能夠觀察到該構念的個別差異的方法，此種定義稱為操作型定義（operational definition）。由此可知，創造力的操作型定義，決定了創造力的評量與衡鑑，而創造力測量的最大問題，也就在於每一位研究者的定義不同，導致測量的形式與結果也就不同（Kerr & Gagliardi, 2003）。

測量的進行必須倚賴評量工具，一個評量工具的好壞最重要的指標是信度與效度的評估。其中信度估計（estimation of reliability）是針對整個測驗的穩定性與可靠度的檢驗。從測量誤差的觀點來看，信度指的是一個量表能夠偵測到真實分數的程度，反應了測量誤差的大小。一個測驗如果有越高的信度，則有較強的偵測真實分數的能力，反之則無法從測量分數當中獲得穩定可靠的測量數據作為描述分析之用。所以信度係數可以說是一個測驗最基本的一個優劣指標。最常用的信度估計方法為Cronbach（1951）所發展的 α 係數，目前在心理學研究中，幾乎所有的測量工具都須報告Cronbach's α 值，以說明該量表的可信賴程度。此外，用以估算一個量表在兩個時點之間的穩定情形的再測信度（test-retest reliability）也經常被使用，又稱為穩定係數。

效度所反應的測量分數的真實內涵與意義，雖然信度係數提供了測量分數是否能夠測到真實分數，但是接下來的問題卻是，這個真實分數到底是什麼？對於一個測量工具進行效度

的評估，目的就是在回答這個問題。有趣的是，效度的評定只是一種參考性的價值，不像信度係數可以對於測量工具的穩定性進行絕對性的評定。效度的好壞評定程序並沒有一套共同的標準，在測量工具評估中可以視為一個充分條件，若各種效度指標係數十分理想，表示測量工具的效度良好，對於測量工具有相當的加分效果，但是效度係數偏低，不一定表示量表品質不佳。

貳、創造力定義與評量方法

一、創造力被視為能力

Creativity一詞多被翻譯成創造「力」，即反應了人們對於創造力是一種能力的直觀看法。事實上，多數的學者對於創造力的定義，也與創造的知覺與認知歷程有關。以影響創造力研究最深遠的心理學家Guilford為例，他在1950年的演說中即指出，創造力是指那些具有傑出創造表現的人們所具有的能力（“Creativity refers to the abilities that are most characteristic of creative people”）。而韋伯字典中對於creativity的定義則為“the ability to bring something new into existence”。目前最通用的創造力測驗，也多與認知思考能力的測量有關（例如Torrance的創造思考測驗）。這些現象顯示創造力與人們的心智能力有密切的關聯，因此許多研究都圍繞在創造力與聰明才智、批判力、發散性思考能力的關係的討論上。

二、創造力被視為個人特質

另一種對於創造力的界定，在於高低創造者的個別差異的區分，也就是尋找高創造者的特殊特質。累積了多年的研究之後，這個取向的研究已經累積了相當豐富的成果，早在1981年，Barron與Harrington（1981）在心理學年度評論的一篇回顧性論文，已經完整的整理了創造者所呈現的一般性特質，例如偏好複雜、自主、自信、容忍自己與其他的人的差異。此外，從不同的領域來看，高創造的科學家具有下列特殊的人格特質：不喜歡與人來往、喜歡與「事」有關的工作，而不喜歡與「人」有關的工作、較高的自我控制與自我指導、思想與行動較獨立而不從眾、獨立判斷、易受複雜的情境與人際關係所影響、高度的自我強度、講求精密與正確、較能容忍模糊情境等。高創造性藝術家則有獨立、敏感而不受限制、喜好幻想、比較自我中心、較強的認知能力、較直觀等等。高創造的建築師則傾向於獨立性、較強

的個性、決斷力、認知複雜性、開放自我的知覺與感覺、敏感的理解力、廣泛的興趣、較女性化、拒絕接受信條、較高的自我知覺、較高的自我期望等特性。

除了Barron與Harrington（1981）的整理，Basadur與Hausdorf（1996）認為個體所具有的價值觀扮演重要的角色，當一個人看重新觀念的重要性之時，往往就能有較好的表現。Barron（1969）則將重點放在個人自我的強度（ego strength），包括個體自己內在存在矛盾與衝突的接受容忍度。Raudsepp（1981）認為個人情緒及態度方面的障礙會影響創造性的發展。例如：逃避問題的挑戰、缺乏自信、害怕被批評、不斷與他人競爭、早期生活經驗受挫、缺乏自我知覺、缺乏正向情感、企求熟悉的事物、過分順從他人意見、害怕獨處、過分沈溺於緬懷過去和憧憬未來、情緒混亂等。Sternberg（1995）認為高創造力的人具有面對障礙時的堅持、願意冒合理的風險、願意成長、忍受模糊的狀態、接受新經驗、對自己有信心等特質。

從上面的各項特質中，可以看到與創造力有關的個人特質包含範圍甚廣，包括人格特質、動機、認知偏好、態度、能力或經驗因素等等。因此促成了許多創造力的特質測量的工具的發展。

三、作品取向的創造力觀點

近年來大多數學者多從創作的產品（Product）觀點對創造力進行研究與定義，原因有二，第一，就目前的心理學理論與研究而言，以過程取向的定義來進行實質的測量有其難度，尤其是創造思考的進程與內容難以清楚觀察測量。第二，思考過程究竟是否具有創造力，最終仍需倚賴所產出之作品或其它結果而得知，因為在人面之面上仍存在許多未知之影響因素所干擾（Amabile, 1983）。

Amabile在其創造力評量研究中強調，「產品或可觀察的反應才是創造力最終之證明」（Amabile, 1983）。一件創作物是否具有創造性乃受控於觀察者及評價大眾，根據他們對創造力一詞之知識、甚或是想像，形成了是否具有創造力之評判效標。此一標準之最大特性在於，創造力之具備與否主要依據於評判者之主觀效標。其優點在於它克服了客觀評定指標定義不容易之困難，但容易因為社會環境中之主流價值限制（Barbara、Camea, 2003）與態度上之喜好與否、生理吸引力之影響所致，而在評斷上產生不同差異。本文將針對創造作品的共識性評鑑方法（CAT技術）進行介紹。

參、創造力評量的分類

「測量」是創造力研究的重要環節，Guildford於1967年提出「智力結構理論」(Structure of Intellect)係由思考內容(content)、運作方式(operation)、產出(product)等三向度所構成。每一向度下又各有不同的結構單位，彼此交互建構成180種智慧組合。其中又以「發散性思考」(divergent thinking)與創造力最為相關，Guildford並據此理論架構編制了數份相關之測驗。從此，關於創造力之評量開始受到世人廣泛的注意與重視，並開始有若干其它相關之測驗陸續問世。近年來關於創造力研究正方興未艾，許多學者對「創造力」之觀點也漸趨於分歧與多元化，加上測驗理論與統計方法不斷地進步與創新，「創造力」之評量方法益發多元！Hocevar與Bachaler(1989)整理了所有有關創造力的評量方法，區分為八大類，Kerr與Gagliardi(2003)則以歷程、特質為主，區分為兩大類，本文以Hocevar與Bachaler(1989)的分類系統為基礎，簡化成創造歷程、創造者特質、行為與經驗測量、他人評量技術四大類，分別描述如下。

一、創造歷程的測量

(一) 發散性思考測驗 (tests of divergent thinking)

能力導向的創造力測量可以說是共識最高的一種創造力測量方式。多數的學者對於創造力的定義，多與創造的認知歷程有關此，其中最重要的理論觀點源於Guildford之智力結構中的發散性思考(divergent thinking)，具有流暢力(flucency)、變通力(flexibility)、原創力(originality)、精進力(elaboration)、定義力(redefinition)等不同成分。基本上，發散性思考反映了創造力的認知屬性的部分面向強弱，並不能完全代表創造能力，因此這一類的測驗僅可被稱為「發散性思考測驗」，而不宜作為創造能力的代表(Davis, 1989; Zarnegar, Hocevar, & Michael, 1988)。

發散性思考的測驗中，以「拓弄思創造思考測驗」(Torrance, 1974; TTCT)使用最為普遍，並被吳靜吉教授於70年代引進國內。根據Guilford的智力結構理論，人類認知能力的運作可以區分成不同的素材(material)，因此TTCT區分為語文與圖形兩類。語文創造思考測驗中另有「產品改良」、「不尋常用途」等分測驗；非語文創造思考測驗中另有「完成圖畫」、「平行線」等分測驗。

其它的一些測驗在型式、內容、操作和計分上，皆類似於拓弄思創造思考測驗。例如Wallach和Kogan所於1965年發展的測驗，包括了舉例（Instances）、替代用途（Alternate）、相似性（Similarities）、形狀意義（Pattern Meanings）與線條意義（Line Meanings）等五個分測驗。每一個分測驗均可獲得一個獨特性分數（Uniqueness score）和數量分數（Number score）。此外，尚有一些測驗試圖測量個體創造歷程，Ghiselin、Rompel和Taylor（1964）所發展的創造歷程檢核表（Creative Process Checklist）即為一例，該測驗是設計去衡鑑科學家在發明的過程中，注意力和情緒的狀態。

二、創造者特質的測量

（一）態度與興趣量表（attitude & interest inventories）

除了能力取向的測量方法，在態度與情意向度上的個別差異，也能夠反應創造力的強弱，因此促成了創造力的態度與興趣測驗的發展。過去的研究發現，具有創造力的人會展現出有利於創造的態度與興趣傾向，因而可藉由個人對創造的興趣與態度來鑑別其創造力之高低（Barron & Harrington, 1981; Runco, 2004）。此外，幽默、玩興或玩性（playfulness）亦是與創造力有密切關係的概念，也可以說是創造者的一種情意與興趣表現。過去的研究發現孩童在觀賞完歡樂的影片後會有較佳的創意表現（Isen, Daubman, & Nowicki, 1987）。

Davis與Rimm（1982）所發展的興趣測驗（Group Inventory for Finding Interests; GIFFI）就是一個典型的興趣測驗，測量的題目包括「我有豐富的幽默感」、「我喜歡嘗試冒險的行為」。Khatena & Torrance所編之「創造知覺問卷」（Creative Perception Inventory），其中包含了「你是哪一種人」（What kind of person are you），共有50題強迫選擇題，用以評量與創造力相關之心理特質，例如接受權威、自信、好奇、想像力等等；以及「我自己」（Something about myself）以自陳式方法評量創造性人格特質與創造行為經驗，例如主動性、自我強度、獨特性等等（Khatena & Torrance, 1976）。

普林斯頓大學教授Raudsepp（1981）在從事二十餘年的創造研究的經驗中，根據其對企業界或一般具有創造才華的個體所作的觀察，編製了創造商數量表（Creativity Quotient Inventories, CQI），用以測量十個不同創造性行為層面的創造性傾向：包括左右腦傾向，有關於人類創造思考的能力與右腦功能有密切的關係。價值取向、工作態度、問題解決行為、青少年經驗、興趣、人際關係、人格屬性、自我知覺與負向自我意象，並由丁興祥、陳明

穗、蔡啟通、邱皓政、李宗沅（民80）等人引進國內。

（二）人格測驗（personality inventories）

心理學的觀點認為，個人所獨有之人格特質對其行為之影響扮演關鍵性之角色。故有許多學者認為在創造力的行為表現上，人格特質因素同樣發揮了重要之影響。故以評量與創造力相關之人格特質來推斷其所具有之創造能力高低。過去的研究發現，高創造的科學家不喜歡與人來往、喜歡與「事」有關的工作，而不喜歡與「人」有關的工作、較高的自我控制與自我指導、思想與行動較獨立而不從眾、獨立判斷、易受複雜的情境與人際關係所影響、高度的自我強度、講求精密與正確、較能容忍模糊情境等。高創造性藝術家則有獨立、敏感而不受限制、喜好幻想、比較自我中心、較強的認知能力、較直觀等等。高創造的建築師則傾向於獨立性、較強的個性、決斷力、認知複雜性、開放自我的知覺與感覺、敏感的理解力、廣泛的興趣、較女性化、拒絕接受信條、較高的自我知覺、較高的自我期望等特性。因此，人格測驗便被應用於創造力領域上來衡量創造性人物與非創造性人物兩者在人格特質上之分別。

Goughs於1957年所發展的加州心理測驗（California Psychological Inventory）；Cattell和Eber所發展的16PF（Sixteen Personality Factor Questionnaire）；Gough和Heilbrun於1965年所發展的形容詞檢核表（Adjective Check List）以及Heist和Yonge於1968年所編製的綜合性人格量表（Omnibus Personality Inventory），都可以測量人們的性格特徵。此外，尚有專門設計來測量創造者人格特質的人格測驗，例如 Davis和 Subkoviak於1975年編製的「你如何思考？」測驗（"How Do You Think?" test）；Rimm 1976年所發展的「發現創造性天才的團體量表」（Group Inventory for Finding Creativity Talent）以及Torrance和Khatena兩人所共同發展的「你是那一種人」測驗（"What Kind of Person Are You?" test），也有關於人格特徵的測量。

三、行為與經驗測量

（一）傳記式問卷（biographical inventories）

俗語說「鑒往知來」，因此許多研究者認為「過去行為是未來行為最好的預測指標」，創造力行為表現也是如此。個人之生活經驗與周遭環境所接觸的人、事、物、情境等皆可能有助於甚或有害於創造行為之發展，因而發展出「傳記式問卷」，藉由個人過去所遭遇的經

歷、對某問題所做出之行為反應、家庭與學校之環境狀況，來據以評量創造之潛能。

創造力的傳記量表的原理，主要是取用具有創造力高者的代表樣本，與一般人相較而有高傾向的行為經驗與歷史。例如創造行為研究機構（Institute for Behavioral in Creativity）所發展的阿法傳記量表（Alpha Biographical Inventory），便是發展自對於科學家與工程師的測量，其內容包括數百個有關童年生活、興趣、嗜好等之項目。

Schaefer於1970年所發展的「傳記量表：創造性」（The Biographical Inventory：Creativity）則包括了165個有關家族史、教育史、休閒活動、身體特徵和才藝能力的項目，所涵蓋之範圍包含了科學、文學、藝術等方面之創造人才評定；Taylor於1963年則發展一種50個項目的傳記量表，編製樣本係來自一所電子工場中的員工，比較管理人員對於受試者創造力與生產力的評量結果而獲得量表的內容。

（二）傑出人士（eminence）

有些學者認為研究傑出人士所具有之特質可以提供評量創造力之指標（丁興祥，1991, 1992, 1993; Simonton, 1976）。針對傑出人士的著名者為人格評量與研究學院（Institute of Personality Assessment & Research, IPAR）在1970年代所進行之一連串調查研究，他們的研究對象包括建築師、作家、數學家等，然後從這些特殊表現的個人加以分析，探討他們的人格特質或家庭背景因素各種因素，做為創造力高低的依據。Simonton（1976）曾以結構方程模式（structure equation model）之統計方法，針對音樂家與科學家的歷史與傳記資料中，找出個人因素（如年齡）與社會環境（如政治環境穩定性）對創造力產生之影響。Csikszentmihalyi（1996）也曾針對不平凡的傑出人物進行關於創造力之深度訪談研究，其研究結果有助於說明創造性人物與非創造性人物於行為表現、思考方面之分別，並瞭解於哪些情境狀況下會有助於或有礙於創造力之展現。

（三）自陳式創造活動與成就（self report creative activities & achievements）

評定創造力最直接的方法即是依照個人所陳述之活動經驗與成就故事來加以評斷，例如競賽的獲勝經驗、專利權的獲得、個人作品公開展覽等等，這些具體創造表現來表示個人創造力之高低。此種方式的優點在於其表面效度甚高，可以具體瞭解個體在創造力的哪些面向有客觀的表現，但是缺點是標準化程度低，個體間的比較有其困難，而且創意活動或表現的內容與水準的定義亦有其主觀性。因此在創造力的衡鑑上，多作為補充性的測量，或是作為

個案研究的素材。

四、他人評量

（一）教師、同儕或主管評定（ratings by teachers, peers, supervisors）

前面幾種測量方法都是由受測者本人進行自陳式的填答，另一類型的評定方法則是由其他人來針對受測者的個人屬性或狀態進行評定，以描述被測者的創造力。其中最常使用的是教師評定法（teacher nominations），亦即由學生的老師就學生的能力或個人屬性進行評定。評定的內容多為與創造力相關之指標，例如與發散性思考有關的思考流暢性、變化性、發明能力、原創性或精進力（Yamamoto, 1963）；或是與創造表現有關的人格特質，如統整性、獨立性、彈性、合作意願等，請教師以形容詞檢核表的方式來評定學生的特性（Harrington, Block, & Block, 1987）。也可以直接對於學生是否具有創造力進行整體評估（Rimm & Davis, 1980）。類似於教師的評量，在企業領域則是以上司評定（supervisor rating），對員工進行一連串之評量與審核，以甄選出具有創意之員工（Keller & Hollandm, 1978）。

除了老師的評量，研究者也經常使用同儕評量法（peer nominations），就某些具體的向度或屬性來評定（Torrance, 1974; Foster, 1971）。雖然同儕評量多應用於兒童或學生的創造力評定，在企業界也應用同儕評定法來進行員工能力的評估，所評估的指標非常廣泛，包括語言表述能力、接受困難挑戰程度、堅持力、流暢力、思考演繹能力、與人相處能力、專業能力等等，不一而足（Specher, 1964; Taylor, Smith, & Ghiselin, 1963）。

（二）客觀產品評鑑（judgments of products）

另一種創造性評量方式，是由其他評定者直接對於個體的創造性產品加以評鑑，如果評定者不只一位，則是一種共識衡鑑的應用。Ghiselin（1963）認為個體是否具有創造性，可以由產品的「內在品質」（intrinsic quality）的客觀分析來認定。Simonton（1980）曾經發展一種可靠、客觀的量化技巧。他使用兩部音樂曲目的字典，找出樂曲發表的日期以及作曲者的生日等資料，在所蒐集的479位作者的15,618首樂曲中，他利用樂曲中的前六個音符來決定該曲的原創性，每一個音符皆與下一個音符配對比較，區分成五組配對的音符，經由電腦的分析，以音符和間距的罕見性為基礎，計算每一樂曲的轉化（transition）情形，並與全部的音符配對相比較。由此一過程所計算出來的罕見性分數（rarity scores）便代表該曲的原創性。

產品評鑑的評量方式，在操作型定義上非常的明確，操作過程簡易且具有合於邏輯的客觀性，但是其數學模式卻不易直接衍用於其它的領域，而且此一技巧無法清楚的區分創造性與「怪異性」(bizarre)之間的差別，又重新面臨了創造性定義的問題 (Amabile, 1983)。事實上，對於創造性的評鑑，僅單獨使用此一技術可能會有所不足，個人創造性所涉及的現象，可能不一定會反應在實際的作品上，例如發自人格本身的創造性氣質、幽默感等等日常性的創造性行為表現即無法客觀的予以分析。

肆、創造力測量的相關議題

一、效度的問題

創造力評量的最大爭議是在效度的問題之上。基本上，一個測量要有效與否與三個層面有關，第一，被測量的構念的特性，第二，測量分數的獲取過程與意義，第三，測量分數與其他分數的關係。

效度 (validity) 即測量的正確性，指測驗或其他測量工具確能測得其所欲測量的特質或功能之程度。測量的效度愈高，表示測量的結果愈能顯現其所欲測量對象的真正特徵。若以內容來看，效度的評估有三種不同的模式，從測量的內容與範圍來著手的內容效度 (content validity)，從外在標準契合程度的評估模式所發展出的效標關聯效度 (criterion-related validity)，以及強調概念意涵釐清的建構效度 (construct validity)。

(一) 構念效度問題

基本上，建構效度 (construct validity) 係指測量工具能測得一個抽象概念或特質的程度。建構效度的檢驗，必須建立在特定的理論基礎之上，透過理論的澄清，引導出各項關於潛在特質或行為表現的基本假設，並以實徵的方法，查核測量結果是否符合理論假設的內涵。最常提及的建構效度評量技術，係由Campbell和Fiske於1959年所提出的多元特質多重方法矩陣法 (multitrait-multimethod matrix)，據以考驗聚斂效度 (convergent validity) 及區辨效度 (discriminant validity)。MTMM設計的原理是以多種不同方法 (如自評法、同儕評量法) 測量多種不同的特質。

1. 理論與定義選擇問題

對於創造力的測量，最大的挑戰就在於是否該測量工具能夠反映出一個人的創造力高低。創造力測量所面對的第一個問題，是什麼是創造力？創造力的內涵或要素為何？創造力這個概念是不是與其他構念不同？這些問題回答之後，研究者才能據以發展測量工具，並進而評估是否測到他。在定義與說明的過程中，所牽涉到的關鍵問題是研究者所採取的理論觀點。

如果按照Guilford的觀點，他認為創造力是一種廣泛性、擴散性思考的認知能力，這種思考能力表現在思考的流暢性、變化性、獨創性與精密性四種成分上，這些思考特性潛力較佳者，可以表現出創意的行為，得到與眾不同的答案或作品。Guilford對於創造力的看法，源自於他的SOI智力模式（structure-of-intellect; SOI），因此他的創造力定義是一種以認知能力為基礎，在測量上即需以一般能力測驗的方法來評定。

顯然的，一旦研究者採用了不同的理論觀點，所獲得創造力的評量結果也就有所不同，對於一個個體或產品的評價也就不同。充分反映了創造力構念的模糊性，導致構念效度的檢驗受到理論選擇主觀性的威脅。很不幸的，許多創造力的實徵研究，在構念效度（Construct validity）上，無法獲得理想的結果（Bastos, 1974; Goolsby & Helwig, 1975; Holland, 1968; Jordan, 1975; Kazelskis, 1972）。

2. 構念區辨問題

如果理論與定義層次的問題獲得了解決，進一步的在測量上如何證明創造力與其他心理構念間的關係，是研究者的一個挑戰。傳統上的建議，是利用MTMM設計來計算不同構念間的聚斂性與區辨性，但是創造性測驗之間彼此有關，並不能解決效度上的問題（Amabile, 1983）。尤其是利用外在評定者的評定程序時，區辨力並不理想，顯示評分者評定分數受到月暈效果等方法效應的影響甚大。例如Holland（1959）的研究發現，不同評定者對於十幾種不同的能力，其中包括發散性思考能力，評定分數具有高度的相關。例如原創力與語言能力及寫作能力間的相關達.72與.84區辨力甚低。Wallen與Stevenson（1960）的研究也發現教師去評定學童的創造力與IQ相關達.57，與其他學習成就測驗間的相關達.66至.72之間。其他一些對於藝術創造力的測量研究，發現創造力評定與其他相關技能的構念的區辨問題更加困難（Brittain & Beittel, 1964; Csikszentmihalyi & Getzels, 1970; Rossman & Gollob,

1975)。

以評定者進行創造力評分的高低區辨力現象，顯示評定者偏誤不僅造成信度不足，也帶來效度上的問題。因此，研究者發現若能改善評分程序，採用趨於標準化的策略，或是加強評分者的溝通與訓練，可以提高評量的品質，提高構念區辨能力（Milgram, 1983; Richards, Kinny, Benet, Merzel, 1988; Runco, 1986; Runco & Albert, 1985; Sternberg, 1985）。

除了評定程序技術上的改善，近年來則建議採取統計控制的方法，以驗證性因素分析（confirmatory factor analysis）來排除高度測量誤差的影響，並藉以區辨構念間的關係（Noar, 2003; Ding & Hersherberger, 2002; Hocevar & Bachelor, 1989），同時可以檢驗創造力測量是否具有高階的潛在構念或測量恆等性（measurement invariance）問題（Marsh & Hocevar, 1985; Hocevar & El-Zahhar, 1984）。顯示當代的結構方程模式技術的發展，可以協助研究者進行創造力的測量。

（二）構念效度的衡鑑策略

1. 因素分析策略

為了檢驗測驗的構念效度，當代最廣為使用的策略是因素分析法，透過測量分數之間的共變關係的分析，將測量分數的結構化關聯，視為共同因素，也就是潛在構念存在證據。在實際研究上，因素分析有探索性（exploratory）與驗證性（confirmatory）兩種不同的操作策略。如果在進行因素的萃取之前，研究者並未對於資料的因素結構有任何預期與立場，而藉由統計數據來研判因素的結構。此種因素分析策略帶有濃厚的嘗試錯誤的意味，因此稱為探索性因素分析（exploratory factor analysis; EFA）。然而，如果研究者在研究之初既已提出某種特定的結構關係的假設，例如Guilford的創造力理論觀點，此時研究者可以利用觀察資料與理論觀點的適配程度來確認理論模式的合適性，此種因素分析稱為驗證性因素分析（confirmatory factor analysis; CFA），具有理論檢驗與確認的功能。

顯然的，CFA在偏重理論檢驗的創造力測量當中較佔優勢，因為創造力的內涵為何必須有一理論作為基礎，進而才能進行測量工作，因此較宜採用CFA的觀點來進行構念效度的檢驗。但是值得注意的是，CFA策略僅是一種統計的方法，僅能用來檢驗觀察資料與理論觀點的符合程度，統計指標並不能反映「構念」是否被真實

的測量，因此CFA或EFA程序都僅是作為一種構念效度的部分證據，單獨使用CFA或EFA都不足以說明是否測到了想要測量的構念。

2. 實驗策略

除了使用統計程序來檢驗測量分數的意義，另一種更能反應測量結果是否能夠反應所測特質的方法是利用實驗設計或實際的觀察紀錄，來確認是否測量得分較高者，真正具有創造的能力。例如我們可以將一群學生隨機分配兩組當中，其中一組實施有助於創造力提升的訓練課程，另一組則為對照組，實驗完成之後，實驗組學生在創造力測驗的得分上應顯著高於對照組學生。

雖然實驗方法可以直接檢驗是否創造力能夠被具體測量，優點是可以得到強而有力的因果論證，但實驗方法最大的困難是耗費成本與可行性不高。不論是把創造力視為認知能力或人格特質，都不容易改變與訓練，因此難以在實驗情境中來提升，藉以觀察到測量分數的差異。

3. 神經心理策略

除了實驗方法之外，由於科技的發展，使得心理學家得以利用先進的儀器設備來探討人類大腦與神經機制的運作狀況。例如利用大腦斷層掃描（computerized axial tomography, CAT）或是腦波測量（electroencephalograph, EEG），可以偵測人類進行思考或行為時，大腦運作的狀況。Hudspeth（1985）就曾經利用CAT與EEG，瞭解不同創造得分的受試者的腦部活動，得到顯著的證據。

如同實驗方法的限制，神經心理測量的策略雖然可以協助我們瞭解測量是否真實的反應創造力的內容，但是也具有相當高的可行性門檻，除了設備設施的有無與操作能力，更大的挑戰在於創造力究竟與腦或神經活動的關聯為何，如果研究者不能明確的指出神經生理基礎與創造表現的關係，那麼CAT、EEG的測量數據徒留更多的問號而已。

總而言之，構念效度可以說是創造力測量最為棘手但是卻是最為關鍵的議題。從實體論的觀點來看，如果創造力是真實存在的一種能力，創造力的個別差異應可以反映在具體的客觀行為表現上，而得以計量方法來測得其內涵（Borsboom, Mellenbergh, & van Heerden, 2004）。然而上述這些構念效度的方法，對於創造力的評量皆有其侷限，除了在高等計量方法（例如結構方程模式、項目反應理論）力

求提升以外，其他效度證據的補足與其他測量方式的提出，皆有助於此一議題的解答。

（三）內容效度

如同構念效度的問題，內容效度牽涉到測量範疇的定訂，也是創造力測量的另一大困難之處。一般的創造性測驗，僅在評鑑有限範圍之內的創造性表現，對於概念複雜的創造性現象而言，測驗所測得的分數是否能夠代表個體的創造性，頗受質疑。此外，在行為經驗的評量方面，由於創造力的表現在不同知識領域有很大的不同，因此內容範疇的變動甚大。

（四）效標關聯效度

效標關聯效度又稱實證效度（empirical validity）或統計效度（statistical validity），係以測驗分數和特定效標（validity criterion）之間的相關係數，表示測量工具有效性之高低。

效標關聯效度最關鍵的問題在於效標（criterion）的選用。做為測量分數有效性與意義度的參照標準，效標必須能夠作為反應測量分數內涵與特質的獨立測量，同時也需為社會大眾或一般研究者接受能夠具體反應某項特定內涵的指標。因此研究者除了透過理論文獻的證據作為選用效標的基礎，且能提出具有說服力的主張，來支持其效標關聯效度檢驗的評估。

基本上，效標混淆（criterion contamination）是創造力測量最明顯的一個問題，創造力測量的目的在瞭解受試者的創造能力，同時創造表現的認定本身即是創造力測量本身的一個目的，兩者間非常可能具有重疊的現象。創造力的測量的內容多以創造有關的作業或行為經驗有關，但是效標的測量分數也可能包含著某些行為作業或經驗內容，造成效標混淆的問題。

很明顯的，在創造力測量中，效標的選擇是非常不容易的一件事，因為創造行為並不是非常容易觀察的現象，而且在不同領域，創造表現也有明顯的差異，要能得到客觀的創造行為表現，是非常困難的一件事，因此所得到的效度係數一般均不理想（Kerr & Gagliardi, 2003）。根據Hocevar與Bachlor（1989）的整理發現，以各種效標來檢驗創造力分數的研究，相關多在.30以下，顯示效標與創造力測量間的共同變異部分不及10%，Hocevar與Bachlor（1989）認為是一種邏輯效度（nomological validity）不足的現象，也就是測量分數並無法反應理論上或邏輯上應該具有的關聯。

邏輯效度偏低的一個原因，是創造力測驗本身會由於測量誤差的存在而降低效度的估計，也就是誤差扭曲問題（attenuation due to measurement errors）此時如果效標測量也具有誤

差時，效度係數低估的情形會更加明顯。除此之外，如果測量時，樣本的異質性不足，導致樣本變異量過低時，對於相關係數的估計也會有一低估的傾向，這些問題都可能造成測量上的低估問題。

二、測量信度問題

在測驗領域，信度表示測量的可靠性（trustworthiness），也就是測量結果的一致性（consistency）或穩定性（stability）。任何一種測量，或多或少會有誤差，誤差由機率因素所支配，也就是一種隨機誤差（random error）。誤差愈小，信度愈高；誤差愈大，信度愈低。因此，信度亦可視為測驗結果受機率影響的程度。

以能力測驗形式為主的創造力測驗，例如發散性思考，由於多採開放式問題，因此計分程序多需倚賴主觀評定程序，因此測量誤差較大，測驗的信度偏低。另一方面，如果創造力測驗的內容涉及不同知識領域的內容，也會因為提目的異質性導致題目間的一致性偏低，信度係數偏低。

在各種創造力的測量工具中，以人格與情意測驗的信度較高，因為測量的內容是與創造力有關的人格特質，穩定性高，題目間的同質性較能夠維持一致，內部一致性係數多可達.70以上；但是在行為經驗的評量，由於題目異質性較高，內部一致性通常較低。

雖然自陳式的人格與情意測驗具有較為理想的信度，但是測量過程容易受到社會讚許性的影響，使得測量分數的信度所受到的威脅不是隨機誤差的影響，而為一種系統性的誤差。解決之道是在測驗發展的程中，落實項目分析的工作，儘量避免使用容易引發社會讚許性的題目，提高題目的表面效度，維持題目的難度適中，減少高低難度落差太大的創造力作業。

創造力評量的評分者間信度問題，特別發生在當測驗需要由他人進行評定時。尤其是共識衡鑑法，對於創造性產品的評估，是由不同評分者的平均分數作為創造力的指標，因此影響評分者評定的因素，皆可能影響信度。我們將在下一節有關共識衡鑑部分詳細討論。基本上，如果是需要評分者進行評分的測驗，則應強化評分者的訓練，藉以提高分數的一致性，必要時可以採用類化理論（generalizability theory），來探討是否有其他的因素會影響評分者間評定的一致性。

三、常模問題

創造力測驗與智力測驗的一個明顯差異，是創造力測量多沒有標準答案，而其他能力測量皆有正確答案可供對照。因此，創造力得分高低通常需藉由常模的對照來得到創造力高低的分數。以TTCT為例，受試者所完成的答案，需對照常模得到獨創力的得分。某一個受試者完成10個答案，每一個答案是否具有獨創性必須參酌常模樣本的稀有度，才能得到具體的分數，也就是說，創造力的得分高低是相對分數而非絕對分數，如果沒有評分常模，即無法得到獨創力分數。

另外，常模的應用也會隨著各種不同的情境而有變化，例如不同性別、種族、文化脈絡、學門、年齡，而會有變化，如果要區分成不同類型的常模，使得常模的建立難度與成本更高。如此一來，立即凸顯出創造力測驗的常模問題。如果創造力測驗缺乏常模，根本無法得到正確的分數，測驗所得到的數據也毫無意義可言。即使完成了常模的建立，也必須時時進行更新、維護。對於測驗編制者來說，是一個非常沈重的負擔。在我國，心理與教育學術領域規模並不大，在測驗的編制上以經顯得捉衿見肘，在測驗的維護上，更顯得困難。這也是國內欠缺本土型創造力測驗的主要原因。

四、測驗實施問題

創造力評量的另一個問題是實施上的問題。如果從認知測量的觀點來看，創造力的作業是一種費心的認知作業，因此對時測驗時間的限制與標準化就成為得分高低的影響因素。另外，題目的難度如何安排，也會影響受試者的得分。

如果從情意或人格測量的角度來看，創造力測量很容易受到作答者的反應心向或反應風格影響其作答行為。此外，施測者本身的暗示、主觀期待、經驗的多寡、指導語的設計，都可能影響受試者的表現，使得創造力測驗的實施程序非常容易影響測量的結果，此時必須加強施測者的訓練，也必須透過標準化作業程序來降低人為的影響與其他不預知的干擾。

肆、創造力的共識衡鑑

創造力研究領域的一位重要學者MacKinnon，在1975年提出了一個迥異於傳統的「創作者本位」的創造力研究觀點，認為「社會評價」對於創造力具有決定性的影響，他認為如果一個作品無法獲得他人正面評價與肯定，創造力是無法在一個社會中立足與發展，也就是說，在創造力的背後，存在一個特殊的社會歷程。二十年後，另一位重要學者呼應了他的想

法，Sternberg（1995）指出：「創造力之評量取決於評定者之喜惡，可以有某組裁判認為其有創意，也可能找到另一組裁判評定者的看法完全迥異，可見得評量者的因素對於是否具有創造力的評定扮演舉足輕重之角色。」，這句話明確的指出，創造力的共識衡鑑（consensual assessment）對於創造作品的認知與評價具有重要的意涵，而且創造作品的評定，與評定者的組成有關，不同「組」的裁判，可能得到不同的評定結果，顯示評定者的特質與團體歷程，是創造力評定的重要變數。

從學術的觀點來看，MacKinnon與Sternberg的看法不約而同的指出，創意產品或觀點除了要有相當的新奇、獨特、與創見之外，也需具有相當的社會性與適切性，尤其對於創造性作品的評價，創意本身被他人接受的程度，以及審美上與技術上的優劣，更可能是評估者關切的主要焦點。顯見創造力的研究，可以從評定者的角度來切入，將研究視野擴及人際互動與社會層面，更豐富創造力研究領域的知識內涵。

一、共識衡鑑技術（CAT）

對於創造物的評量，最重要的評量技術是透過評定者（judgers）對於某一個被評定物（創造作品）評定是否是新奇且適當的作品的共識決定，此一策略稱為共識衡鑑技術（consensus assessment techniques; CAT）（Amabile, 1996）。在CAT技術下，一件作品或個人創造能力的有無，取決於評定者的主觀看法。因此，此一評定者的特徵及評定者的內在歷程，可以說是CAT技術最關鍵的核心議題。

Amabile（1996）指出，適當的共識衡鑑程序應具有下列特質：第一，評定者對於被評作品具有相當程度的領域背景知識或經驗，例如對於一部電影的評定者，應對於電影有相當的瞭解與接觸經驗；第二，評定者必須獨立進行他/她的評定，而不受他人的左右或影響；第三，從研究的觀點，CAT應不僅要求評定者對於「創造性」進行評定，而需同時在其他相關的指標上進行評定（例如新奇性、適當性、技術品質等等），如此可以更進一步的瞭解評定者對於該作品的看法，以及瞭解各指標間的關係；第四，評定者應被要求針對每一件作品在各指標的上的評定其感受，而不是就各產品在某一特質上進行相對比較，如此才能凸顯各作品在評定者心目中的優越程度；最後，評定作品的順序，每一位評定者應有不同，以避免順序造成的效果。

對於受評作品，Amabile（1996）亦指出了幾個要件，第一，該作品必須是具體可觀察，

足以讓評定者產生具體感受而評定之。第二，作品的評量的結果具有相當的開放性，足以讓評定者去評估其感受強度與內容。第三，被評量的作品應非受限於特定能力或技術的作業產品，例如特定類型的手工藝作品或一篇學術論文，因為這類作品深受特定能力與技術而決定其產出的可能，此時評定者的評分可能只是被動的反應作品產生的困難度與可行性。

根據一系列的研究，CAT技術被證實是一個可以用來評估各種不同類型作品是否具有創造力的評量技術，例如說故事（Stubbs & Amabile, 1979），美術繪畫（Getzels & Csikszentmihalyi, 1976; Berglas, Amabile, & Handel, 1981; Amabile, 1982; Amabile, Goldfarb, & Brackfield, 1990），詩詞創作（Amabile, Goldfarb, & Brackfield, 1990; Amabile, 1985），而且對象包括了成人或兒童的作品。不論是專業評審或一般大學生評審，都有相當理想的評定共識，評分者間信度多介於.70至.80之間。因素分析的結果發現，雖然這些研究所使用的指標分成三大類，與創造力有關、與作品技術有關（technical）、以及美感（aesthetic）（例如個人的喜好度、作品的美感）。但是實際數據只跑出創造力有關與技術有關兩個主要因素，而美感指標與兩個因素的關聯均很微弱（因素負荷量很小），顯示前兩項指標是構成創造力判斷的主要內容，而美感評定則屬個人的內在喜好評定。

二、文化影響與評定者偏誤

Sternberg及其學生最近一項有關於創造力共識衡鑑的研究（Niu & Sternberg, 2001），發現中國評定者較西方評定者有較大的共識，也就是說，華人受試者在評定一件作品是否具有創意時，彼此間的評分差異較小。進一步的，研究者指出，兩種不同文化背景的評定者對於評定標準的運用的差異雖然不大，但是美國大學生在評定創意時，較中國大學生有較為嚴格的傾向，顯示文化因素造成共識衡鑑的歧異性，除了反映在標準使用的寬嚴差異上，更會反映在共識形成的難易程度。若以統計學的術語來說，寬嚴的程度反應在集中量數（平均數）上的差異，共識的高低則為離散量數（變異量）的差異。另一篇類似的研究，也發現了華人評定者的評定分數高於西方評定者，但是研究者（Chen, Kasof, & Himself, 2003）認為此一現象是華人評定者受到自我效能與自我涉入（self-serving）的情況所造成，而非基於社會互動的考量。顯示華人評定者的評定歷程的確有別於西方評定者之處。

前述研究可以說為創造性的共識衡鑑議題，開啟了另一個研究的空間，也就是文化差異的比較。傳統上，創造力的共識衡鑑研究，多以西方社會的樣本來進行研究，因此研究者所

關心的問題多以個體的評量歷程為中心，例如共識衡鑑結果是否受到外在環境的影響，而造成高估或低估（Hennessey & Amabile, 1988），或是關心外在酬賞、涉入程度對於個體進行評定時的影響（cf., Amabile, 1996）。但是，如果在一個社會文化特性截然不同的母體中，探討影響共識衡鑑的影響因素時，重要的影響變項即可能與該文化體系當中特有的個人或環境變項有所關聯。

以華人的社會環境為例，人們的社會互動緊密，人與人之間的互動關係影響了許多社會行為的模式，在這種強調「關係」或「社會取向」的社會模式（Triandis, 1996；楊國樞，1997；何友暉、陳淑娟、趙志裕，1991）下，人們的互動方式會有強調和諧、重視共識、抑制歧見的現象，對於創造力的表現，顯而易見的會有負面的影響（吳靜吉，2002），邱皓政（2003a）稱此種因為人際互動型態所導致的創意行為的負面作用為「人情困境」現象。從此一邏輯來看，由於華人評定者的社會取向與關係特性，使得共識衡鑑似乎顯得更容易，這或許是造成Niu與Sternberg（2001）與Chen, Kasof, & Himself（2003）等人能夠觀察到華人的共識較西方評定者為高的一個主要原因。也就是說，原本華人社會的行為風格對於創造力的發展是一個不利的因素，但是在共識衡鑑過程中，反而成為一個促成共識達成的正向因素（林欣鈺、邱皓政，2003）。

在心理計量上有一個重要的概念，是高信度不代表高效度，高效度則必須要有信度為基礎。將此一概念應用到本研究課題，即成為「高共識是有效評估的基礎，但是高度的共識並不一定表示此一共識是正確有效的，也可能是一種系統性偏誤」。在共識衡鑑上，評定者之間的意見歧異被視為是一種隨機誤差（random errors），代表每一位評定者獨特的意見或喜好，將各評定者的評定結果加以平均即得到集體的共識評分；但是如果評定者之間的意見歧異呈現一種系統性波動，也就是一種系統性誤差（systematic errors），例如一致性的高估或低估，此時意見的歧異部分的加總將造成共識的高估或低估，形成無效的（invalid）共識，稱為評定者偏誤（rater bias），例如慈悲或寬大效果（leniency errors）（Kneeland, 1929）。

三、評定偏誤的影響

在實務領域，共識衡鑑不僅是創造力評定的常用技術，同時也廣泛被應用在其他各種人類特質與能力的評定過程中。例如人事甄選時，主試官以面談技術評定應徵者的工作態度與能力；研究所入學考試的推薦甄選，審查委員除了審查申請者的備審資料，另以口試來評斷

學生的學習性向與能力；還有各式各樣的比賽、得獎作品的決定，都是倚賴評審委員的主觀評量的共識程度，來決定一個人或作品的命運。

共識衡鑑的基本假設，是每一個評審對於他人的評定都有獨特的意見，而大家看法的集中點（平均數）就是對此一受評者最公允的評斷，因此Sternberg（1999）指出，一個人或作品是否有創意，重點並不在於這個人或作品本身，而是取決於他人如何評斷。在實際生活中，我們經常可以聽聞此類倚賴評定者進行評定以決定入學資格或得獎與否的批評論調，或是同一個司法案件，不同的法官基於不同的自由心證，使得判決的結果南轅北轍，令眾看官迷惑不已。這些爭議，除了因為主觀評定的主觀性濃厚，造成他人對於評定公平性的質疑，更重要的是共識衡鑑的正確性，的確深受個人特質與環境變項影響（Saal, Downey, & Lahey, 1980）。

（一）系統性偏誤與隨機性偏誤

共識衡鑑最大的威脅，是評定者主觀因素所造成的系統性偏誤（systematic bias），而非隨機性誤差（random error）。從測量的觀點來看，任何評量皆有可能產生誤差，隨機性誤差雖然影響評定效果的正確性，但是由於造成誤差的來源並沒有一致性因素，因此對於平均數的計算不會產生影響，共識性的評定結果不會產生特定的高估或低估現象，此時，誤差所反應的是測量的信度（reliability）問題，而非評量的效度（validity），此時，共識的決定仍可以說是公平或正確的，因為誤差效果可以被統計平衡。

然而，如果造成評定分數變動的誤差原因是一致性的使分數提高或降低，此時稱為系統性誤差，所造成的影響即已不是信度的測量一致性或穩定性問題，而涉及了評量的有效與否，也就是涉及了公平正確與否的問題，此時，由於系統性偏誤造成的評定者偏誤（rater bias）效果，被視為測量分數中，一種基於評量方法造成的變異量（稱為方法變異，method variance）（Campbell & Fiske, 1959），是共識衡鑑正確運用的最大殺手，因此研究者無不費盡心思來克服此一障礙（e.g., Cronbach, 1995; Kenny, 1994）。

（二）個別評定者偏誤（RS bias）與對偶評定者偏誤（DS bias）

在最近的一篇文獻中，Hoyt（2000）指出評定者偏誤主要來自於兩種評定者間的不一致意見：第一，是評定者對於評定標準的解釋上的歧異，第二，是評定者對於被評定物的知覺的一致或波動。這兩類因素可能發生於同一個評定者對於所有的被評定物，也可能發生於特

定的被評定物。前者稱為個別評定者偏誤（*rater-specific bias*; RS bias），後者稱為對偶評定者偏誤（*dyad-specific bias*; DS bias）。根據此一分類，當追求共識衡鑑技術來評定創造力或其他特質的正確性與可比較性時，特定評定者偏誤（RS）的影響較容易獲得控制，因為RS偏誤是一般性的評定者間變異，而不涉及特定的對應關連（Raymond & Viswesvaran, 1993），然而對偶性的偏誤則涉及不同的評定者與受評對象之間的對應關係，在控制上必須先行找出對偶的因素，例如特定類型的老師（評定者）欣賞特定類型的學生（對象），才可能針對偏誤的系統變異來源加以控制（Hoyt, 2000）。

在對偶性的評定偏誤問題中，除了評定者與被評對象之外，另一個混淆因子是所評定的向度，當評定的內容牽涉到不同的向度時，對偶偏誤的發生即可能針對不同的向度而有所不同，例如創造力評量中，被評定的內容除了新奇性，還包括了適切性與吸引力等等，而對偶偏誤除了因為評定者與對象之間的不同對應關係之外，亦可能因為不同的評定向度或效標而有不同變動，添加了評定偏誤的控制的難度。

（三）文化差異議題

由前面的討論可以得知，對偶性的評定偏誤是威脅共識衡鑑的主要問題，尤其對於創造力的評量，因為創造力的抽象特質與多元性內涵，使得特殊性的評定者對偶偏誤更容易發生。Niu & Sternberg（2001）的研究指出中國評定者較西方評定者有較大的共識，如果從Hoyt（2000）的觀點來看，此一論點所意味著華人評定者的共識衡鑑的差異，是一種RS型個人性評估偏誤，而非對偶性偏誤。事實上，西方研究者對於創造力共識衡鑑的探討，多將共識衡鑑界定在一種RS個人性偏誤的層次，並進而討論如何受到外在機制的影響，造成偏誤作用的變化。例如Amabile（1996）在其創造力的脈絡（*Creativity in context*）一書中，即已經完整的將外在因素（例如獎賞、競爭、合作因素）如何影響創作者的創造歷程加以說明，並據以建構了創造力的社會心理學（*the social psychology of creativity*）的理論體系。

Niu & Sternberg（2001）的研究證明了RS偏誤的確存在於兩個文化中，造成了平均數的差異，但是仔細檢視研究數據，會發現其實兩個國家的大學生評定者，對於創造力的本質與定義看法並無差異，而評定分數的平均數差異，是因為不同文化評定者，對於標準的運用方式有所不同，Niu & Sternberg（2001）的描述如下：

「……rating criteria used by these two groups of judges was that Chinese judges tended to award higher grades on average to all products than did American judges……this difference to

American judges applying a higher standard to judge art products due to the apparently higher artistic ability of Americans compared to Chinese. …… This specific results seems to reveal that people in different cultures may have a similar understanding of artistic creativity.」(p.239)

他們的研究數據支持了他們的論點，也就是共識衡鑑是一種方法效應的展現，且具有文化差異。也就是說社會與環境的脈絡，造成了創造力表現水準（平均數）高低的差異。然而，此類研究取向並無法深入探討共識衡鑑的偏誤的對偶性特質，也就是到底是什麼原因，影響特定文化下的評定者偏誤產生變異，這些問題尚待我們進行深入的探討。

伍、結語

人之所以為萬物之靈，即因為人類具有其他物種所沒有的心智能力（intellectual abilities），使得人類可以快速有效的處理外界的訊息，進行高階的思考，並進而轉換成實際的行為能力，解決各種問題，其中的核心概念就是智力（intelligence）與創造力（creativity）。在學術領域，很早便對智力問題產生興趣，各種學理觀點令人眼花撩亂，測量的工具與方法不一而足，例如在1921年，美國的教育心理學刊即以Intelligence and its measurement為主題進行專題研討。但是，對於創造力的研究，直到1950年代才獲得學術領域的重視。到了2005年的今天，雖然有關創造力的內涵、機制與相關議題的爭議仍然很多，但五十年來的探討，至少確立了一個結論：就是創造力與智力是不同的，而且在不同的知識領域，創造力的表現形式有所不同，重要的影響要素也有差異（Runco, 2004）。

人類的創造能力及其表現是一個非常抽象且複雜的現象。經過半世紀來的研究與討論，學者對於「創造力」雖然已有較多的了解，但仍然無法以一個簡單的概念來界定「創造力」，對於一個作品或一個人是否具有創造力，更難以單一的標準或量尺來評估之。創造力的研究，極端倚賴一套適切的方法與工具。另一方面，但我們也要瞭解Simonton（2003）的主張，創造行為的研究不能脫離真實世界，而具有創造力的個體或產品，也必須對於我們所身處的世界產生具體的影響才具有意義。事實上，Simonton（2003）的觀點與Csikszentmihalyi（1996）的系統理論有密切關連，Csikszentmihalyi（1996）認為創造力不單只是一種認知歷程的產出，而是一種文化與社會的產物。

本文花費了相當的篇幅介紹了有關創造力的測量工具，主要的分類依據是筆者的指導教授Dennis Hocevar於1981年即已提出，當時即已經指出創造性測驗的各種缺點。然而20多年

來，創造力的測量方法並沒有特出的發展或革新，在沒有更有效、方便的研究工具的窘狀之下，創造性研究者除了更積極的發展新的測量工具之外，必須不斷地修正舊有的測驗常模，使得測量工具本身不致造成太大的偏誤。同時，創造性測驗為了能夠涵蓋更完整的範疇，亦有必要發展整合性工具。

另一方面，本文耗費了相當篇幅介紹了共識衡鑑技術，雖然CAT技術涉及主觀的評斷，並非客觀評量技術（Hocevar, 1981），但是從近年來共識衡鑑普遍受到學者的重視，以及實際運用的成效來看，CAT技術確實是客觀評量方法之外的一個替代選擇，事實上，共識衡鑑更重要的是能夠反應當代專業小眾對於創造力的共同看法與價值取向，亦即反應作品的社會與文化脈絡，因此具有其他測量技術更理想的生態效度（ecological validity）（Amabile, 1996），在應用到創造力的文化比較議題或社會心理現象的探討時，CAT技術恰可以用來理解背後的社會文化脈絡的影響與作用。

但是水可載舟亦可覆舟，共識衡鑑技術的使用雖然具有社會生態的評價意義，但也可能因為外在影響力量的作用而造成排擠與偏誤作用。例如當一個學門普遍採取特定觀點來評價學術的創造力之時，可能會抑制了另類或新觀點的提出，流於一言堂式的創作。再如當入學制度逐漸改以非統一考試的推薦甄選與申請入學之際，學生進入大學或研究所就讀的途徑越來越依賴評審委員對於一個學生多元能力、學習作品、創意潛力的共識評定，此時，若無對於評審進行共識衡鑑的心理歷程與相關因素的系統研究，提供一套積極有效的甄選策略，則難避免評量偏誤現象的發生。Csikszentmihalyi（1996）即指出，創造力深受「學門」的影響與宰制，某一領域中之一群人的共同意見與行為模式，往往會決定了何種作品是創意或傑出之成就，此一「守門人」現象說明共識評定對於創造力的界定與產出，雖可超越單一個人主觀判斷因素與認知因素的主導，但是卻帶來另一種社會文化常規的宰制作用，是共識衡鑑技術最大的限制與隱憂。

不論是客觀評定或共識衡鑑，獲得客觀正確的創造力評量分數是學術工作者與實務工作者的共同期望。近年來，心理計量理論與統計分析技術發展甚為快速，利用項目反應理論（Item Response Theory; IRT）或結構方程模式（Structural Equation Modeling; SEM）等高階的測量與分析技術，得以更有效的處理測量誤差與潛在構念的評量，是創造力評量的一個具有研究潛力的領域。可惜的是，這些高階統計技術的進入門檻甚高，必須要有深厚的統計測量知識背景才得以巧妙運用。即使進入了SEM的分析領域，多數研究者僅利用基本的驗證

性因素分析來檢驗因素結構的關係。事實上，CFA可以擴充至高階、多層次的因素結構的檢驗，（也就是higher-order/multilevel confirmatory factor analysis），也可以檢驗不同族群間的比較問題（也就是multi-sample structural equation modeling），以及不同情境或對象的效度複核的複核效化研究（cross-validation）（Noar, 2003；邱皓政，2003b；黃芳銘，2004），潛力非常可觀，可以說是創造力評量的進一步的研究主要方向（Hocevar & Bachelor, 1989），也讓創造力與計量研究者看到了一絲共同的未來。

參考文獻

- 丁興祥（1991）。中國傑出思想家的成長背景研究：以宋元明思想家為例。論文發表於中央研究院族學研究所行為研究組小型專題研討會系列之六，台北南港。
- 丁興祥（1992）。中國文學創造的社會脈絡：一種跨歷史的世代分析。輔仁學誌（文學院），21，53-79。
- 丁興祥（1993）。中國傑出科技人才創造發展的環境脈絡。輔仁學誌（文學院），22，137-163。
- 丁興祥、陳明穗、蔡啟通、邱皓政、李宗沅（1991）。羅賽浦創造傾向量表修訂報告。測驗年刊，38，189-200。
- 何友暉、陳淑娟、趙志裕（1991）。關係取向：為中國社會心理方法論求答案。楊國樞、黃光國（主編），中國人的心理與行為。台北市：桂冠。
- 林欣鈺、邱皓政（2003）。涉入程度與外在酬賞因素對創造力評定之影響。論文發表於中國心理學會第四十二屆年會，輔仁大學，台北縣。
- 邱皓政（2002c）。創造力的人情困局：人際關係的遠近親疏對於創意表現之影響。論文發表於中國心理學會第四十一屆年會，成功大學，台南市。
- 邱皓政（2003a）。個人開放性、人際關係型態對於創意表現之影響。論文發表於中國心理學會第四十二屆年會，輔仁大學，台北市。
- 邱皓政（2003b）。結構方程模式：LISREL的理論技術與應用。台北市：雙葉書廊。
- 王木榮，林幸台（1986）。威廉斯創造力測驗修定研究。師大特殊教育研究學刊，2，231-250。
- 吳靜吉（2002）。華人學生創造力的發掘與培育，應用心理研究，15，17-42。

- 吳靜吉、高泉豐、王敬仁、丁興祥（1981）。拓弄思圖形創造思考測驗指導及研究手冊（甲式）。台北市：遠流。
- 吳靜吉、高泉豐、王敬仁、丁興祥（1981）。拓弄思語文創造思考測驗指導及研究手冊（乙式）。台北市：遠流。
- 黃奕光（2003）。Asian創造力－為什麼西方人比東方人有創造力。台北市：培生。
- 黃芳銘（2004）。結構方程模式在教育資料應用之研究。台北市：五南。
- 楊國樞（1992）。中國人的社會取向：社會互動的觀點。楊國樞、余安邦主編，中國人的心理與行為-理念及方法篇。台北市：桂冠。
- Amabile, T. M. (1979). Effects of external evaluation on artistic creativity. *Journal of Personality and Social Psychology*, 37, 221-233.
- Amabile, T. M. (1983). *The Social Psychology of Creativity*. New York: Springer-Verlag.
- Amabile, T. M. (1985). Effect of motivational orientation on creative writers. *Journal of Personality and Social Psychology*, 48, 393-399.
- Amabile, T. M. (1996). *Creativity in context*. Boulder, Colorado: Westview Press Inc.
- Amabile, T. M., Goldfarb, P., & Brackfield, S. C. (1990). Social influences on creativity: Evaluation, coactions, and surveillance. *Creativity Research Journal*, 3, 6-21.
- Amabile, T. M., Hennessey, B. A., & Grossman, B. S. (1986). Social influences on creativity: The effects of contracted-for reward. *Journal of Personality and Social Psychology*, 50, 14-23.
- Bastos, L.D.(1974). *An evaluation of the Torrance tests of creative thinking*. Dissertation Abstracts International, January, 34, 3976-3977.
- Berglas, S., & Amabile, T. M., & Handle, M. (1981). *Effects of evaluation on children's artistic creativity*. Unpublished manuscript, Brandeis University.
- Bruner, J.S. (1962). The conditions of creativity. In H. Gruber, G. Terrell, & M. Wertheimer (Eds.), *Contemporary approaches to creative thinking*. New York: Atherton Press.
- Barron, F. (1995). The disposition toward originality. *Journal of Abnormal and Social Psychology*, 51, 478-485.
- Barron, F.(1969). *Creative Person and Creative Process*. New York : Holt, Rinehart and Winston.
- Barron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annual Review of*

Psychology, 32, 439-476.

Barbara, K. and Camea, G. (2003). Measuring creativity in research and practice. In Lopez, S.J. & Snyder, C.R. (Eds.), *Positive psychological assessment: a handbook of models and measures*. Washington, D C: American Psychological Association.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The Concept of Validity. *Psychological Review*, 111(4), 1061-1071.

Brittain, W. L., & Beittel, K. R. (1964). A study of some tests of creativity in relationship to performance in the visual arts. In W. L. Brittain (Ed.), *Creativity and art education*. Washington DC: National Art Education Association.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.

Chen, C., Kasof, J., & Himself, A. J. (2003). Creativity in drawings of geometric shapes: A cross-cultural examination with the consensual assessment technique. *Journal of Cross-Cultural Psychology*, 33, 171-187.

Cronbach, L. J. (1951). Coefficient alpha and internal structure of tests. *Psychometrika*, 16, 297-334.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.

Cronbach, L. J. (1995). Giving method variance its due. In P. E. Shrout and S. T. Fiske (eds.), *Personality research, methods, and theory: A festschrift in honor of Donald Fiske* (pp. 145-157). Hillsdale, NJ: Erlbaum.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. New York : Harper Collins Publishers.

Csikszentmihalyi, M., Getzels, J. W. (1970). Concern for discovery: An attitudinal components of creative production. *Journal of Personality*, 38, 91-115.

Davis, G. A. (1989). Testing for creative potential. *Contemporary Educational Psychology*, 14, 257-274.

Davis, G., Rimm, S. (1982). (GIFFI) I and II: Instruments for identifying creative potential in the junior and senior high school. *Journal of Creative Behavior*, 16, 50-57.

- Feldhusen, J. F. (1995). Creativity: A knowledge base, meta-cognitive skills, and personality factors. *Journal of Creative Behavior*, 29, 255-268.
- Foster., J. (1970). *Creativity and the teacher*. London: Macmillan.
- Getzels, J., & Csikszentmihalyi, M. (1976). *The creative vision. A longitudinal study of problem-finding in art*. New York: Wiley-Interscience.
- Ghiselin , B.(1963). Ultimate criteria for two levels of creativity. In C. Taylor & F. Barron(Eds.) , *Scientific creativity ; Its recognition and development*. New York : Wiley.
- Gough, H. (1957). *The California Psychological Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Goolsby , T.M. & Helwig , L.D.(1975). Concurrent validity of the Torrance tests of creative thinking and the Welsh figural preference test. *Educational & Psychological Measurement* , 35, 507-508.
- Harrington, D. M., Block, J. H. , & Block, J. (1987). Testing aspects of Carl Rogers's theory of creative environments: Child-rearing antecedents of creative potential in young adolescents. *Journal of Personality and Social Psychology*, 52, 851-856.
- Hennessey, B. A., & Amabile, T.M. (1988). The conditions of creativity. In R.J. Sternberg (Eds.), *The nature of creativity: Contemporary psychological perspectives* (pp. 11-38). Cambridge, England: Cambridge University Press.
- Hocevar, D. (1981). Measurement of creativity: Review and critique. *Journal of Personality Assessment*, 45, 450-464.
- Hocevar, D., Bachelor, P. (1989). A taxonomy and critique of measurement used in the study of creativity. In J. A. Glover, R. R. Ronning, C. R. Reynolds (Eds.), *Handbook of creativity*. NY: Plenum Press.
- Hocevar, D. & El-Zahhar, N. (1992). Cross-cultural differences in test anxiety: Establishing transliteral equivalence. In K. A. Hagtvet and T. B. Johnsen (Eds.), *Advances in Test Anxiety Research*, 7, 48-61, Lisse: Swets & Zeitlinger.
- Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups.

- Psychological Bulletin*, 97, 562-582.
- Holland, J.L.(1968). Test reviews. *Journal of Counseling Psychology*, 15, 297-298.
- Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5(1), 64-86.
- Hudspith, S. (1985). *The neurological correlates of creative thought: A comparison of the EEG activity of high and low creative subjects with an ergonomic presentation of the results for the lay person*. Unpublished doctoral dissertation, University of Southern California., Los Angeles.
- Isaksen, M. C., Murdock, R. L., Firestien, & D. J. Treffinger (1993). *Understanding and recognizing creativity: The emergence of a discipline*. Norwood, NJ: Ablex.
- Jackson, P., & Messick, S. (1965). The person, the product and the response: Conceptual problems in the assessment of creativity. *Journal of Personality*, 33, 309-329.
- Jordan, L.A.(1975). Use of canonical analysis in Cropely's "A Five-year longitudinal study of the validity of creative tests." *Developmental Psychology*, 11, 1-3.
- Kazelskis, R.(1972).The convergent, divergent and factorial validity of the Torrance tests of creativity. *Southern Journal of Educational Research*, 6, 123-129.
- Khatena, J. & Torrance, E.P. (1976). Manual for Khatena Torrance creative Perception Inventory. Chicago : Stoelting.
- Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefutte, J. (1995). Stability of rater leniency: Three studies. *Academy of Management Journal*, 38, 1036-1051.
- Keller, R. T., & Holland, W. E. (1978). A cross-validation study of the Kirton Adaptation-Innovation inventory in three research and development organizations. *Applied Psychological Measurement*, 2, 563-570.
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guildford.
- Kneeland, N.(1929). That lenient tendency in rating. *Personnel Journal*, 7, 356-366.
- Lance, C. E., LaPointe, J. A., Fiscaro, S. A. (1994). Tests for three causal models of halo rater error. *Organizational Behavior and Human Decision Processes*, 57, 83-96.
- Mayer, R.E.(1999). Fifty years of creativity research. In Sternberg R. J. (Eds.), *Handbook of creativity*, (pp.449-460.) N Y: Cambridge University Press.

- Niu, W., & Sternberg, R. J. (2001). Cultural influences on artistic creativity and its evaluation. *International Journal of Psychology, 36*(4), 225-241.
- Raudsepp, E.(1981). *How Creativity Are You ?* New York : Academic Press.
- Raudsepp, E.(1987). Establishing A Creative Climate. *Training and Development Journal* , 41(4) , 50-53.
- Runco, M. A. (2004). Creativity. *Annual Review of Psychology, 55*, 657-687.
- Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin, 88*, 413-428.
- Schaefer, C. (1970). *Manual for the Biographical Inventory: Creativity (BIC)*. San Diego, CA: Educational & Industrial Testing Service.
- Simonton,D.K.(1976). Biographical determinate of achieved eminence: A multivariate approach to the Cox data. *Journal of Personality & Social psychology, 32*, 1119-1133.
- Simonton,D.K. (1999). Creativity and genius. In L. Pervin & O. John (Eds.), *Handbook of personality theory and research* (2nd ed., pp.629-652). New York: Guildford Press.
- Simonton, D. K. (2003). Scientific creativity as constrained stochastic behavior: The integration of product, person, and process perspectives. *Psychological Bulletin, 129*(4), 475-494.
- Specher, , T. B. (1964). Creativity and individual differences in criteria. In C. W. Taylor (Ed.), *Widening horizons in creativity*. New York: Wiley.
- Sternberg, R. J. (1988). *The triarchic mind: A theory of human intelligence*. New York: Viking.
- Sternberg, R. J., Lubart, T. I. (1996). Investing in creativity. *American Psychologist, 51*(7), 677-688.
- Sternberg, R. J., Lubart, T. I. (1999). The concept of creativity: Prospects and paradigm. In Sternberg, R. J. (Eds.), *Handbook of creativity*, (pp.3-15). N Y: Cambridge University Press.
- Stubbs, M. L., & Amabile, T. M. (1979). *Explaining the relationship between fantasy and creativity*. Paper presented at the Third Annual Conference on the Fantasy and Imaging Process, New York.
- Taylor, C.W., Smith, R. W., & Ghiselin, B. (1963). The creative and other contributions of one sample of research scientists. In C.W. Taylor & F. Barron (Eds.), *Scientific creativity: Its recognition and development*. NY: Wiley.

- Torrance, E. P. (1974). *Torrance Test of Creative Thinking: Norms Technical Manual*, Research Edition personnel press, Inc., Princeton, N.J.
- Triandis, H. C. (1996). The psychological measurement of cultural syndromes. *American Psychologist*, 51(4), 407-415.
- Wallach, M. A., & Kogan N. (1965). *Modes of thinking in young chikdren: A study of the creativity-intelligence distinction*. New York: Holt, Rinehart & Winston.
- Walberg, H. J., & Stariha, W. E. (1992). Productive human capital: Learning, creativity, and eminence. *Creativity Research Journal*, 5, 323-340.
- Weisberg, R. W. (1986). *Creativity*. Yew York: Freeman.
- Yamamoto, K. (1963). Relationships between creative thinking abilities of teachers and achievement of pupils. *Journal of Experimental Education*, 32, 3-25.
- Zarnegar, Z., Hocevar, D., & Michael, W. (1988). Components of orginal thinking in gifted children. *Educational and Psychological Measurement*, 48, 5-16.