

從心境邁向情境：情境判斷測驗在人事甄選與心理特質評量上的發展與應用

邱皓政* 歐宗霖**

摘要

本文的主要目的在簡述近年來受到測驗學者重視的情境判斷測驗的基本原理與格式特性，情境判斷測驗的主要特色在於擬真情境的試題設計與生態效度，有別於傳統上由受測者依據其個人的心理感受、情緒狀態或態度立場進行表達的自陳式量表，而是要求受測者在特定的情境前提下，就個人的經驗或態度對於不同的選項進行判斷。基於多年來各方學者的努力，對於情境判斷測驗的試題建構與計分模式，已經有了初步的操作模式。本文除了詳細介紹情境判斷測驗的基本題型、發展程序、計分方式、信效度議題之外，另以空服員的職能系統的情境判斷測驗發展經驗為例，說明情境判斷測驗的建置過程，提供相關人士參考。

關鍵詞：自陳式量表、情境判斷測驗、量表發展、空服員、職能測驗

* 國立臺灣師範大學管理學院副教授（Associate Professor, College of Management, National Taiwan Normal University）

** 國立中央大學企業管理學系博士候選人（Doctoral Candidate, Department of Business Administration, National Central University）

From mind to situation: Application and development of situation judgment test in personnel selection and psychological testing

Hawjeng Chiou
Tsung-Lin Ou

Abstract

The main purpose of the present article is to introduce the principles and formations of the Situational Judgment Test (SJT). The advantages of situational judgment test are scenario design of item and ecological validity. As a results, the nature of SJT differs from self-report scale in terms of the process of personal judgment of testers on the specific situation risen by the test item, instead of the self-reported feeling or attitudes. In the recent years, scholars devote to develop suitable approaches for SJT construction. The scoring issues along with the reliability and validity has been discussed in the article. An example of SJT for measuring the competency of cabin crew was also introduced. Finally, the article concluded the significance of SJT and provided practical and methodological suggestions on the application.

Keywords: self-report scale, situational judgment test, scale development, competency testing

壹 前言

一、「操守」能考嗎？

日前，在一項關於國家考試興革的會議中，與會人士熱烈討論一個議題：文官的「操守」是否可以測出來？是否可以在考選過程中加以評估，並作為考試錄取標準或與人員進用參考？事實上，當主席把這個議題拋出之初，會場一片靜默，主席為了打破冷場，特意詢問身旁的測驗界宿老與業界專家，是否學術界或業界有類似的測評工具或技術可以加以借鏡。基本上，在人事甄選測驗當中確實是有所謂的誠實測驗（integrity test），可以用來評估候選人的誠信特質與職業倫理與道德信念強度（見周麗芳、林守紀、曾春榮，2012; Sackett & Wanek, 1996），但是這裡的誠實、正直是不是主席所謂的「操守」或「廉潔」還需要斟酌，更重要的是，這種測驗多半只能作為輔助參考，或讓受測者感受到主試者對於誠實廉潔的重視，測驗效度頗有疑義（Muchinsky, 2006），因此不宜作為甄選標準，更遑論用來「考」。

與會的學者專家中雖建議嘗試利用情境判斷測驗（situational judgment test; SJT）來取代自陳報告測驗（self-reported test），避免社會讚許（social desirability）與測驗作假（faking）的問

題，提高測驗的情境效度，但是 SJT 身為新近發展流行的測驗形式，對其特性、原理、乃至於優劣與應用範圍需要進一步的討論。最後，主席結論說道，文官考試的目的是在選「賢」與「能」，「賢」，就是人格特質與素養，「能」，就是能力與能耐，國家考試不論怎麼考，都只能針對知識或技術來加以評估，然而影響處事態度與任事方法最直接的個人特質與道德素養，卻無法透過紙筆或術科測驗來評估，因此，如何發展適切有效的工具、導入新興的測驗技術來評估文官的人格特質與態度素養，成為國家考試興革的重要議題。

二、「心境」與「情境」的對比

前面提到的「自陳報告測驗」與「情境判斷測驗」兩個名詞，就測驗技術來說，可以簡單的以「心境自陳」與「情境判斷」兩者來區分對比。所謂心境自陳，是指受測者反思個人的狀態或意見看法，透過測驗題目的強弱選項來反映，例如測驗量表最常使用的 Likert-type（李克特式）五等級量表，或是描述行為的頻率量尺（由「從未如此」到「總是如此」的不同等級），都是個人自我心境的陳述報告。很明顯的，這類型的測驗量表的信效度都必須建立在受

測者有充分的動機與意願，甚至於有能力去表達自我狀態的前提之上，同時也會受到受測者測量當下的心理狀態而有所影響，因此可以稱之為一種「心境的描述」。

相對的，情境判斷則是要求受測者在特定的情境前提下，就個人的經驗或態度對於不同的選項進行判斷，此時不僅題目中加入情境背景，選項也是不同的情境條件，藉由情境題幹的促發、引導，乃至於情境選項的挑選、匹配，來評估受測者的特質類型或是能力強度。換言之，情境判斷測驗試圖透過情境元素的刺激選擇，來改善人類心境狀態的主觀感受的測量偏誤與操弄意圖問題，立意雖好，但是也有編製難度與有效計分的各種技術問題，以及成本高昂、普及性低、編製期程冗長等缺點。儘管如此，情境判斷測驗仍不失為傳統自陳測驗的一種替代方法，同時在諸多學者專家的努力下，近年來在其理論與技術上的發展相當迅速，頗受學術界與實務界人士的關注，因此本文特意撰文加以簡介¹。

三、情境判斷測驗的興起

情境判斷測驗（SJT）是近年來諸

多測量工具中受到重視的評量工具之一，（Sackett & Lieven, 2008），在人事甄選與教育測驗上的應用逐漸開展。2006年的專書「情境判斷測驗：理論、測量與應用」（Situational Judgment Tests: Theory, Measurement, and Application）（Weekley & Ployhart, 2006, Eds.）可視為SJT邁向成熟的指標。該書的第一章，主編指出SJT的興起係源自於實務工作者手中所累積的珍貴經驗，然後在學術工作者手中獲得實證證據。其中SJT吸引眾人目光的主要原因在於工作情境的生態效度（ecological validity）與優於其他測驗的績效預測能力，此一實務優勢使得SJT的應用早於學術研究的步調。

SJT是在二次世界大戰期間，由美國Office of Strategic Service（OSS）所倡議發展，當時的目的是為了甄選擔任特種軍事任務人員。他們設計了多種測驗，其中包括將受測者安排在模擬某種戰鬥情境中，來觀察其處理或應付該情境的能力。他們所發展的「情境壓力測驗」（situational stress test），用來觀察受測者在受到壓力、困難或挫折情境下的行為反應，評估人格特質，藉以挑選出各種任務的適當人選。很多重要的軍

¹ 若要更完整的瞭解情境判斷測驗與傳統自陳測驗的對比，可以參考歐宗霖、邱皓政、孫國勛等人於2013年6月發表於「測驗學刊」的專文「當測驗碰上情境：情境判斷測驗與自陳式量表在量表發展上之比較」。

事任務，亦因這樣有效的甄選而得以圓滿完成（McDaniel & Whetzel, 2005）。

事實上，早在1800年代末期，一些類似的SJT測驗就已應用在美國專利局的商標審查員的考試，或是軍官指揮統御能力的評估。而學術界對於SJT本質與受測者判斷歷程的研究，主要是

Motowidlo等人（1990）對SJT本質的理論建構，並進而探討SJT不同形式與試題本身的設計、計分與資料分析等方法議題，以及重新回到SJT的實務面，審視應用與施測的議題。如此學術與實務間的相互切磋，促進了SJT發展的成熟。

貳 情境判斷測驗的特性

一、情境判斷測驗的特性

SJT從其名稱即可知，其主要特色在於試題當中加入了情境的要素，而個體在特定情境下的判斷反應就是SJT欲蒐集的主要資料。由於人們在特定情境下進行判斷時，會涉及認知與決策歷程，因此SJT的基本理論涉及決策判斷理論（Judgment of decision making; JDM）（Brooks & Highhouse, 2006），當測驗的型態是透過受試對情境脈絡的體認而判斷行為的優劣時，研究者所要面對的測驗複雜度就比傳統自陳式量表僅須自我心境的表態要高出許多。

從測驗形式來看，SJT的問題係模擬真實工作場域的各種問題情境，要求受測者從不同的選項中選擇一個最佳答案，從而反應出受測者所具備的知識、能力、或是人格特質或態度傾向。範例如下：

國科會計畫是大學教授研究經費的主要來源之一，而且通過案件多寡與金額高低都是教授升等的重要參考指標，未獲通過者有申覆的機會。如果您申請未獲通過，第一時間您會怎麼做？

- A. 詳細審閱審查意見，準備詳細的答辯書，立刻提出申覆爭取時效。
- B. 自我反省檢討，分析未獲通過原因，作為來年改善的依據，如有必要才會申覆。
- C. 國科會計畫只是眾多資源的一種，視為資源分配或競賽失利，不會太在意得失，不會提出申覆，而把重點放在其他的研究資源的申請。
- D. 花心思準備半天結果失利總是會有失落，可能會先放下工作去散心旅遊一下，不去想他，留得青山在，不怕沒柴燒。不會提出申覆。
- E. 不一定，都有可能，每一年狀況不一樣，可能申覆也可能不申覆。

在此一例題中，可以很清楚的看出，受測者的回應方式可能有下列兩種型態：根據受測者所認為的考官所要的答案來回答，或是根據受測者所認為的自己可能會採取行為來回答。在SJT的設計原理中，前者稱為「應然反應」(should do response)，後者稱為「實然反應」(would do response)。「應然反應」一般用於認知測驗，用來評估受測者在某項能力特質上的正確判斷或能耐程度，測驗題目具有正確答案，在SJT當中的作答不必然是面臨情境時可能的表現，而得分高低反映受測者的能力強弱，又稱為最大反應測驗(maximum performance test)。「實然反應」則多用於非認知測驗，用來評估受測者個人風格或習慣態度，測驗題目沒有正確答案，又稱為典型反應測驗(typical performance test)，在SJT當中的作答較能預測受測者面臨情境時的真實表現。

換言之，SJT不僅可以用來評估常見於自陳量表的人格、態度等非認知層面的心理構念評估，也可以應用在能力高低強弱的認知測驗診斷。尤其是題目的內容可以就受測對象的實際工作狀況或生活現場，設計具有特定意義與內涵的題目，具有高度的實務應用價值。但是相對的，也正因為特殊性高，因此測驗的類化性就會受到限制，特定的SJT僅能應用在特定的對象與測驗目的，而不容易應用到其他的情況下。以前述

範例而言，題目針對台灣大專院校的教師而設計，每位學者基於個人的經驗差異，學門特性的不同，對申複成功的預期也不同，而會做不同的選擇，因此可作為個人特質評估之用，此時，SJT題目的反應與計分效度是測驗應用成敗的重要環節。

二、情境判斷測驗的格式

SJT的基本題型係由三種歷程所組成：刺激成分(stimulus component)、反應成分(response component)與交互作用(interactivity)。其中刺激成分與反應成分兩部分與其他類型的測驗題目(例如李克特Likert-type題型)類似，亦即由一個主要題幹提供情境線索與條件，再由一組數值選項所組成的量尺來蒐集受測者的反應情形。不同的受測者針對題幹部分所指出的情境(作為標準刺激)有所瞭解後，然後在後續的選項中逐一進行判斷選出一個最佳答案。在題幹的呈現上，除了以文字型態來描述情境條件之外(例如McDaniel & Nguyen, 2001)，近來的新發展則是利用數位化影音或電腦動畫來呈現SJT的主要「情境」，藉以提高受測者作答動機與興趣(例如Kanning, 2005)。

至於反應成分部分，雖然SJT的選項部分也是由數值來代表受測者的不同選擇，但是SJT與其他各種測驗題型的主要差異之一是這些選項的數值並非作

為量尺之用，亦即他們並非如同 Likert 量尺（1 為非常不可能，5 為非常可能）或強迫選擇問題的「是」、「否」，來作為強度大小或得分高低的測度。基本上，SJT 的各個選項都可視為是一個個別的題目，需要受測者仔細研判加以評估判斷，每一個選項所反應的是不同的行為狀態或行動結果，每一則選項各有不同的意涵與行為結果，可能有著不同的計分數值，各選項之間具有對照比較之消長關係，因此屬於題組設計（item-set design）。受測者必須從這些題組式的選項中進行「判斷」或「決策」，藉以反映出受測者的能力、態度或性格傾向。這也就是 SJT 稱為情境判斷測驗的主要原因。至於這些選項的呈現方式，也如同前述刺激成分所呈現的「情境」一樣，可以是文字形式，也可以是影音或電腦動畫形式。

SJT 除了反應成分所具有的特殊計分型態之外，SJT 題型的第三個成分「交互作用」更與傳統測量方式不同。這裡所謂的交互作用係指受測者在進行 SJT 題型測驗時，其判斷過程與結果會受到前後題目鋪陳的影響。這就好比入學考試可能會有筆試與口試兩項考試，筆試時受測者僅需針對考卷的題目作出直觀的反應，但口試時受測者則必須面對面地與主考官對話，從其問題與答案的往復對話中而有所修飾，亦即一種像連環漫畫或連續劇一樣的交互影響動態

歷程。

以前述的國科會申請結果申覆 SJT 例題為例，如果接著該題繼續問道：「如果真的提出申覆後又得到不利的結果時，你會採取什麼反應？…」，此類「承上題」之類的連鎖題型具有高度交互影響性，從而可能產生順向干擾而減損 SJT 計分的客觀性，因此多數 SJT 測驗會避免使用連鎖題等具有高度互動性的題型，降低前後題型的交互影響，使交互性得以標準化（standardized interactivity）（Bauer, Truxillo, Paronto, Weekley, & Campion, 2004; Drasgow et al., 1999; Olson-Buchanan, Drasgow, Moberg, Mead, Keenan, & Donovan, 1998）。

對 SJT 而言，由於試題結構包括題幹的情境描述與題項的行為反應，因此試題設計的考量因素較自陳式量表來的多，包括試題的指導語、複雜度、真實性、結構性與防偽性。其中指導語牽涉受試如何作答的認知，不同的指導語會引導著不同的作答暗示，亦即前述的應然（should do）反應格式，例如，在上述的情境中，你應當採用何種方式因應？其次是實然（would do）反應格式，例如，在上述的情境中，你將會採用何種方式因應？這兩種格式是 SJT 的技術研究與實務應用的重點。一般而言，應然格式所測得的構念偏向知識與認知。相對之下，實然格式所測量的構念偏向個人偏好與反應風格的測量。

三、情境判斷測驗的選項

SJT最複雜之處在於選項的設定與計分方式。一般而言，SJT的選項數目介於3到12之間，這些選項多半由測驗發展者基於檔案、文獻或是訪談所得到的資訊所自行編定，至於針對從實務場域直接產出選項的可能性，Motowidlo, Hanson 與 Crafts (1997) 提出了一種以關鍵事件技術 (critical incident method) 為基礎的案例專家評定法 (subject matter experts; SMEs) 來蒐集SJT情境的「好」答案選項與「壞」答案選項。關鍵事件的描述係由一組A-B-C的反應鏈來表述，A (antecedent) 表示前因 (某個事件)，B (behavior) 表示某個個案的行為方式，C (consequence) 表示行為後的結果。A為前因可作為題幹的內容，B則為選項，對於A-B-C的反應鏈交由測驗機構的專家來評定其行為反應模式優劣程度，作為選擇適當選項與計分的依據。此一策略不但可以得到足夠數量的選項反應作為編題的參考，更重要的是SMEs法所產生的行為選項具有代表性與真實性，實務應用上的價值相對較高。

另一種選項編定方式是基於測量目的與需求所發展的「構念反應式選項」 (construct-based response options)，例如Trippe與Foti (2003) 為了測量受測者的人格特質，以大五人格 (Big-Five) 理論為基礎，發展出測量謹慎性 (con-

scientiousness)、順服性 (agreeableness) 與開放性 (openness) 的SJT，實證數據的結果顯示SJT得分符合理論預期的因素結構。Porr與Ployhart (2004) 等學者的研究指出，構念反應選項所編定的性格取向SJT比起傳統方式有較高的增益效度 (incremental validity)，而且適用於人格測量優於認知能力測量。

至於SJT選項效果評定也有三種策略。第一種方式即是案例專家評定法 (SMEs) (Motowidlo, Hanson & Crafts, 1997)，運用前述介紹的關鍵事件技術法來描述一組A-B-C事件行為鏈，再交由資深實務專家來評定這些行為反應的適切強度，作為選項得分的依據，稱之為實務編碼法 (rational scoring)。第二種方法是實證編碼法 (empirical keying)，透過檔案資料，求出不同行為取向與效標間的加權關聯係數來獲得各選項的優劣資訊。第三種方法則是基於理論、文獻或合理的推理，直接對於選項判定優劣。這三種編碼方式中，前兩種均以實務現場的資訊作為計分的基礎，不論實務應用價值或實證資料證據都有較佳的表現。

四、情境判斷測驗的計分

SJT的另一個特色是特殊與多樣化的計分模式，亦即將受測者在每一個SJT題目的選擇結果合併成單一量表獲分量表得分方式。SJT最直接的計

分方法是採強迫選擇法 (forced-choice methods)，將其中一個或幾個選項視為有效，其他的選項視為無效，然後將各題的有效與無效選項區分後加總，就像一般的考試計分法。這一類的計分法通常所搭配的選項指導與是「請選出一個最佳的答案」，使得受測者可以專注於最佳的選項，將一些不太理想且似是而非的選項加以排除，又稱為單選型 (pick best)。如果每一個選項各自有不同的得分，則稱為部分給分型 (partial credit score) 或排序給分型 (ranking score)。

部分給分型的計分型態係延伸自單選計分題型，在題項評定權值方面同樣由 SMEs 定義題組內最好與最差的行為選項，受試則是要針對題組內的行為選項進行最好跟最差的選擇 (Motowidlo, Dunnette, & Carter, 1990)，其計分方式如下：

1. 若是受試者選擇到的最好跟最差的題項跟 SMEs 所評定的最好與最差題項相同時，則該題組的得分為 +2 分。
2. 若是受試者選擇到的最好跟最差的題項跟 SMEs 所評定的最好與最差題項完全相反時，則該題組的得分為 -2 分。
3. 若是受試者選擇的最好或最壞的題項其中有一命中 SMEs 所評定的最好與最差題項時，則該題組的得分

為 +1 分。

4. 若是受試者選擇的最好的題項命中 SME 所評定最差的題項，或受試者最壞的題項命中 SMEs 所評定的最好的題項時，則該題組的得分為 -1 分。
5. 若受試者選擇的最好與最差的選項都沒有命中 SMEs 所評定的最好與最差的選項時，則該題組的得分為 0 分。

至於排序計分型計分法，則是由 Weekly, Harding, Creglow 與 Ployhart (2004) 所提出，其計分方法為無論是受試者或者是 SMEs，都對於題組內的題項進行等級排序的評分，之後再利用斯皮爾曼的等級相關進行題組的計分。

除了以考試計分法的強迫選擇式計分法外，另一種計分策略則是李克特式量尺法 (Likert-type scale method)。此種計分方法的作法是以 SMEs 法評定各選項優劣的 Likert 分數 (由 1 至 5 分)，然後計算受測者對於各選項的評分與 SMEs 專家評分間的差異分數 (difference)，當差異分數高時表示該題得分較低，差異分數小時則表示該題得分較高 (Chan & Schmitt, 2002; Sacco, et al., 2000; Wagner & Sternberg, 1985)。

以上所提的 SJT 計分法是以有所謂正確答案進行分類的。但對於某些 SJT，由於是基於特定理論下所發展的，因此研究者的重點並不在於判斷受

試者的得分是否正確，而是想要了解不同受試者作答的組型差異，例如Stemler與Sternberg（2006）所設計的SJT即為無正確答案且為Likert式的計分方式，這類型無正確答案的計分方法同樣有著單選型、排序型與Likert式的受試作答型態。

在SJT計分上的最新發展，則是由本文筆者所組成的研究團隊所發展之模糊計分系統，試圖導入模糊語意變數（fuzzy linguistic variables）與模糊複合分數（fuzzy composite score, FCS）兩種模糊觀點，來處理情境判斷測驗被忽略的判斷不確定性，讓研究者能夠探討在不同語意模糊度下的SJT題組分數差異，並比較兩種不同模糊（fuzzy）取向與明確（crisp）取向的計分變化。最近的一篇論文中，Chiou與Ou（2012）模擬不同語意模糊程度的受試反應作答，進行SJT的題組計分，並比較兩種不同模糊取向與明確取向的計分差異，發現語意模糊度的增加，其模糊題組計分的分散性會逐漸縮小。未來在實務上，模糊測量將可提供SJT計分模組更多的選擇，提升SJT的效用。

五、情境判斷測驗的信效度

在量表的信度部分，SJT因為試題可能包含多向度且異質的構念，因此除非SJT的題目夠多，否則傳統Crobach's α 信度並非理想的信度指標。研究指

出，SJT大約需要35個題項以上，Crobach's α 的值才容易大於0.7（Clevenger, et al, 2001）。然而由於SJT的試題複雜性不利於以單一構念為前提的信度檢驗，但仍可從平行測驗與受試反應一致性的觀點來尋求替代的解決方式，因此例如複本信度與再測信度則是SJT學者所認為適當的信度指標（Whetzel & McDaniel, 2009）。

在效度方面，文獻上對於SJT測驗最主要的效標關聯效度檢驗是與工作表現之間的關係的探討，效果相當良好。SJT之所以能夠用來預測工作表現，是因為SJT反映了工作知識的多寡（Motowidlo, Borman, & Schmit, 1997），實用智能（practical intelligence）的高低（Sternberg, Wagner, & Okagaki, 1993），或是一般認知能力的強弱（McDaniel et al., 2001）。有關SJT的構念效度的後設研究結果指出，SJT與人格特質中的謹慎性、情緒穩定性與順服性（McDaniel & Nguyen, 2001）與認知能力（McDaniel et al., 2001）都有明顯關聯。增益效度（incremental validity）的研究也證實了SJT可以作為認知測驗與人格測驗以外的補充性測量。例如McDaniel等學者（2007），在控制SJT不同形式指導語的情況下，利用階層迴歸（hierarchical linear regressions）檢測SJT相對於認知能力（g）、五大人格的增益效度（ R^2 change）介於.01到.02，McDaniel等學

者亦表示雖增益效度的值不大但達統計顯著，且大於認知能力(g)、五大人格最佳組合的增益效度。

在一項後設研究(meta-analysis)中(McDaniel, Morgeson, Finnegan, Campion & Braverman, 2001)，研究者整理了102個與SJT有關的效度係數，發現平均預測力(以相關係數表示)達.34。另一則實證研究將一般認知測驗、人格測驗與情境判斷測驗三種測量工具，來比較各種工具對工作表現的效標關聯效度(O'Connell, Hartman, McDaniel, Grubb III, & Lawrence, 2007)，研究結果指出，對於任務表現與脈絡表現的相關係數，認知測驗為.15與.07，人格測驗為.07-.14與.08-.13，而SJT亦達.14

與.10，顯示SJT與兩種工作表現的關係並不比傳統工具來得低。

綜合文獻上的討論可知，SJT的優勢在於其所測量的內容可以直接與員工所工作的情境相契合，因此不論從理論或實務上，SJT都是較傳統比較個別差異的認知測驗與人格測驗更能去預測為來的工作表現，尤其是利用增益效度的模式去探討SJT是否優於傳統策略，更能提供人力資源管理實務在人事甄選與績效評鑑上的實用知識。以下，僅就筆者所進行的一項三年期的國科會專題研究計畫：空服員實用智能情境判斷測驗的發展經驗為例，說明一個典型SJT量表的發展歷程。

參 情境判斷測驗的發展範例

航空服務業是一種與顧客有高度接觸的行業，這是由於在航空旅程中，空服員與旅客共同待在客艙的時間最久。因此，空服員的服務品質會直接形成旅客對於航空公司的服務印象(Hochschild, 1983)，也因此航空公司往往將空服員視為重要的人力資本，在進行甄選、考核、或評量工作時，評鑑者最基本的一個任務，是能夠利用飛航情境中所蒐集的有限資訊去正確預測空服員當下反應與未來行為表現之間的關聯，然

後做出正確的決定。而空服員在服務過程所具備的知能，除了透過航空公司安排的基本訓練獲得外，很重要的管道是經由工作情境中與顧客、同儕或主管的互動過程獲得經驗法則，此外，航空公司的脈絡文化、空服員本身的價值觀、性格與智力亦會有不等程度的影響效果，綜合上述，空服員服務知能的本質並非是一種單純的構念，而是牽涉一種深植在個人的經驗、判斷、聯想、創意，潛意識之心智模型內的知識，屬於

內隱知識的定義範圍 (Nonaka & Takeuchi, 1995)，因此對於空服員的甄選或評量而言，其關鍵就在於能夠發展適當的測驗工具與程序，測量空服員的內隱知識架構。

為了發展評估空服員人際實用智能的SJT量表，首要工作是確定量表的目的、屬性與內涵。由於量表定位在「人際實用智能」，因此評量所使用的情境為人際問題處理狀況，人際互動類型著重在空服員間的同儕關係、主管與空服員之間的主從關係，以及空服員與乘客之間的服務關係。問題處理的構念內涵主要包括不確定性 (uncertainty)、越權風險 (insubordination)、面臨質問挑戰 (status exertion) 與受迫非自願 (apathy) 的情況。

關於狀況問題與可行選項的來源主要是利用關鍵事件技術為基礎的案例專家評定法 (SMEs) 來蒐集，「狀況問題」的來源，是由研究團隊實地諮詢 20 位臺灣主流航空業的資深空服員作為主題專家的意見，諮詢題項著重在空服員面對人際相關情境時可能涉及的各种工作任務關鍵事件，其中不符合研究問題解決定義的情境皆加以排除，總共編擬出 38 個「狀況問題」。至於可行選項的來源，則是研究團隊實地諮詢 15 位臺灣主流航空業空服部主管，以及服務年資超過五年的資深績優空服員意見，歸納受訪者意見共發展出 152 個

可行選項，諮詢訪談的過程分成三個步驟：其一，先徵詢空服部主管以及相關空服員擔任主題專家的意願，以免違反研究倫理；其二，分別邀請主管與空服員兩組主題專家，分組進行為期一個月的理論介紹與工具發展說明；其三，分別邀請兩組主題專家，透過空服員工作內容分析後的結構性訪談問卷進行個別訪談，發展模擬情境題目以及各種可能的選項、並進行評鑑並決定其權重。

量表初稿完成後，研究者則請教空服部主管、資深空服員與測量專家的意見，就狀況題的代表性，可行選項的內容合理性進行評鑑。此外，立意抽樣 30 位空旅運管理學系的學生進行小規模的試探性預試，以探討量表的形式、文字的敘述是否能為受試者所完全理解無誤。

SJT 選項效果計分方法亦運用關鍵事件技術法，根據實用智能的構念向度與SJT的計分方式，分別處理受試對各選項的評量與得分，作為受試行為判斷與實用智能的測量。在計分層次上則分為題組與題項兩種得分的層次，為的是探討題組與題項得分間的一致性問題，採用李克特式量尺計分法，其計分方法同樣是由受試者或者是案例專家，針對題組內的題項，各別獨立評定其優劣程度，然後根據受試者與專家在李克特式計分上的差距為原則，計算題項或題組分數。

依據前述過程所完成的「台灣空服員人際實用智能量表」題本，包含空服人員在各個工作階段可能發生的「狀況問題」及解決該問題的「可行選項」。每一個狀況問題搭配一個以上的可行選項，讓受試的空服員，心理決策評斷這些選項用來解決實際問題的有效性。全題本共有38個題目，152個行為選項，可歸類成9種情境，分別為餐飲類、生

病類、續酒類、棘手要求、抗議罷機類、危安類、事件類、異常人事類、主管人際類與同儕人際類等事件。以餐飲類情境為例，共有9個題目，包括餐飲預應、餐飲不足、泡麵要求、訂餐訊息有誤、餐飲異物、餐飲異常、餐飲糾紛、餐飲失誤等事件。表1為危安類試題，茲以說明SJT試題設計的重要特徵。

表一 空服員人際實用智能量表例題

類別	情境	選項
起降危安	一段從香港返回台灣的航程，飛機開始進入降落的階段，空服員宋蕙喬瞧見了一位老伯伯急忙地往廁所的方向前去。然而飛機正在滑行，仍需一段時間才能降落於地。試問宋蕙喬應該如何處置呢？	1. 陪同老伯伯乘客一起去，並注意他的安危。
		2. 口頭制止老伯伯前往，但他堅持的話，仍然默許。
		3. 藉故移轉注意力，合理地忽略老伯伯上廁所的行動。
		4. 按照規定阻止老伯伯前往廁所，以避免失職受懲。

肆 結語

一、測評目標的達成

當生活或工作的情境本身就是個題材豐富的試題庫，而人們必須根據當下的情境脈絡條件進行抉擇時，我們可以發現情境判斷的工作在人生發展的歷程中總不斷上演著，因此SJT的概念著實

相當迷人並具吸引力。基本上，測驗與評量的要務在於確認測評的目的與需求，如此才能設計或選擇適當的程序與工具來執行測評工作，滿足測評的目的與需求。綜觀本文SJT的介紹，可以看出其重要優勢，在於對於生態效度的反

映，使得SJT測驗形式能夠滿足測評的消極與積極目標。

當我們利用測驗、量表甚至於考試來進行甄選、考核、或評量工作時，最基本的任務，是要能夠利用測評考試所蒐集的有限資訊，評估受測者的能力水準、性格或態度，進而能夠預期其未來的行為表現或工作成效，然後做出正確的決定。例如現階段的大學與研究所入學考試除了學測、指考的筆試成績之外，更仰賴面試來掌握考生的臨場反應、個人特質性格與學習態度。企業界的人才徵選過程，除了填寫記載各式資料的履歷表，參與特定內容的筆試或實作測驗，更有多階段的面試來證明自己是個最優候選人。這種以「測驗表現」來預測「行為表現」，以心理測驗的專業用語來說，就是要掌握測量評鑑的效標關聯效度。

適當有效的甄選評鑑程序與工具使用，不僅可以落實選賢與能的考選任務，有效運用資源與成本控制，避免錯誤任用與錯誤淘汰等最基本的消極目標，若以人力資源管理的用語來說，其積極目的在於可將適合的人放在適合的職位上，得以表現適切良好的行為，達成工作適配（person-job fit; PJ fit），同時這些適合的人若能找到適合的組織，更可創造良好的工作滿意與效忠忠誠，亦即組織適配（person-organization fit; PO fit）。這其中的成功關鍵就在適當的

測驗評量工具與程序的選用與開發。

二、理絡效度與實務意義的價值

人類史上第一份正式的智慧測驗，是1905年法國心理學家比奈（Alfred Binet）與西蒙（Simon）合作發展的「比西量表」，他們之所以開發這套測驗，是基於法國巴黎教育局的委託，發展一套能夠辨別需要幫助的特殊學童的測評工具。當時對於智能不足的評估並沒有什麼理論可以參考，因此他們走到小朋友的身邊，近身觀察那些表現優異與表現遲緩的小朋友差異，得出智愚有別的三個重要行為特性：第一，小朋友必須知道如何選擇哪些是必須要做的事情並安排步驟和方法（亦即指向性），第二是小朋友必須能在完成任務的過程中選擇有效的策略，並對策略執行適當的監控（亦即調控性），第三，小朋友必須能對自己或他人的想法和行動特性差異有所理解覺察、作出判斷、並進而批評和控制（亦即判斷性）。隨後比奈和西蒙編了58個題目，定義出行為表現的正常水準，記為100分，比一般水準表現優異者高於100，低於一般水準者低於100，在標準之下的一定的程度（例如兩個標準差）則視為是需要幫助的學生，這就是智力商數（intelligent quotient, IQ），至今仍廣泛為人所知所用。

很明顯的，測驗發展的背後動機，往往不是理論的需求，而是實務上的需

要，亦即一種由下而上的歸納演繹過程。但是如果測驗使用者能夠找出強而有力的論述，或能夠具體指出測驗所要衡量的具體內容與標準，那麼測驗就可以省略許多探索確認的過程，由上而下直接進入編製的程序，滿足測驗使用者的需求。值得一提的是，不論是由下而上的經驗歸納演繹，或是由上而下的理論需求指導，測評都有一個核心的主體：也就是受測的對象，以及特定的背景：也就是測驗的情境、內涵或脈絡，換言之，如果僅考慮到受測者本身（例如小朋友的智愚），而忽略了測驗的脈絡（例如小朋友學習與活動的內容），也就是測驗內容所反映的實務或理論意義，那麼測驗目標的達成可以說是言不及義，這也就是當代測驗理論所說的理絡效度（nomological network validity）。情境判斷測驗的最大優勢，即在於理絡效度落實。值得注意的是，由於SJT的優勢是能夠貼近工作場域，進而使得SJT測量能夠預測工作表現，因此SJT題目的研發，必須針對適用對象的工作環境來發展。

SJT的價值，除了和管理應用逐漸受到肯定，在心理與教育領域也獲得重視，尤其具有意義。例如主辦TOEFL與GRE考試的權威機構「教育測驗服務社」（Educational Testing Service; ETS）的研究員MacCann與Roberts（2008）指出，情緒智力（Emotional In-

telligence; EQ）的SJT測驗能夠有效預測考生的測驗焦慮與壓力臨床症狀，他們認為SJT將是EQ、IQ等認知情緒能力與相關行為測量的重要典範。在綜合考量測驗理論、測量技術、測驗偏誤、測驗倫理與公平正義、族群差異等議題下，近年來包括SJT等測量工具在教育考試與企業組織的發展成效是值得肯定與讚許，正說明了SJT的前景看好，值得學界與實務界持續加以關注。

最後，讓我們回到文首的破題「操守能考嗎？」，在本文陳述了SJT的特性與內容之後，對於SJT是否能夠應用於操守的評估這個問題，本文作者仍難給予正面的肯定答案。這是因為「廉潔」、「操守」不僅是個人特質，也會受到能力因素的影響，更涉及道德的判斷，實難利用少數的測驗題目來評估。SJT雖然具有良好的生態效度與高度特殊的命題優勢，但是仍受到測驗理論與技術的規範，必須明確指出測量的內容與對象究竟是能力還是特質，然後據以設計題幹與題項，進而規劃合適的計分與分析策略。即使我們認定「廉潔」、「操守」是個人的行事風格，屬於個人特質的測量而非能力診斷，但若進行判斷仍舊無法脫離認知之影響，因為道德判斷也是認知層次，與個人實際會表現出來的道德行為仍舊有段差距。本文雖有介紹並提倡SJT的積極目的，但是也必須指出SJT的本質與限制，避免測驗

工具的誤用。文首提出操守之議，目的或在引導議題，而非意指SJT有何特殊優勢，更是希望利用此一涉及爭議的特質測量，凸顯人員考選的嚴肅性質，反

映測驗技術的優劣限制。本文除了介紹討論SJT，更希望能夠促發更多關於考選測驗技術與測驗實務的深入討論，共同為考選事務的正向發展做出具體貢獻。

【參考書目】

1. 周麗芳、林守紀、曾春榮 (2012). 淺談心理測驗在企業組織中的應用：以訓練與發展為例。T&D 飛訊，146，1-19。
2. Bauer, T. N., Truxillo, D. M., Paronto, M. E., Weekley, J. A., & Campion, M. A. (2004). Applicant reactions to different selection technology: Face-to-face, interactive voice response, and computer-assisted telephone screening interviews. *International Journal of Selection and Assessment*, 12, 135-148.
3. Bergman, M., Drasgow, F., Donovan, M. A., & Juraska, S. E. (2003). *Scoring of Situational Judgment Tests*. Paper presented at the 18th annual conference for the Society of Industrial and Organizational Psychology, Orlando, FL.
4. Brooks, M. E., & Highhouse, E. S. (2006). Can good judgment be measured?. In J. A. Weekley and R. E. Ployhart (Eds.), *Situational judgment tests*. Mahwah, NJ: Lawrence Erlbaum Associates.
5. Chan, D., and Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15, 233-254.
6. Chiou, H., and Ou, T.-L. (2012). Application of fuzzy composite score for scoring situational judgment tests: A integrated operation viewpoint. In Luiz Moutinho and Kun-Huang Huang (Eds.), *Quantitative Modelling in Marketing and Management* (p. 333-355). NJ: World Scientific Publishing.
7. Clevenger J, Pereira, G. M., Wiechmann, D., Schmitt, N., & Schmidt, H. V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86, 410-417.
8. Drasgow, F., Olson-Buchanan, J. B., & Moberg, P. J. (1999). Development of an interactive video assessment: Trials and tribulations. In F. Drasgow (Ed.), *Innovations in computerized*

- assessment (pp. 177-196). Mahwah, NJ: Erlbaum.
9. Hochschild, A. (1983). *The managed heart: Commercialization of human feeling*. Berkerey: University of California Press.
10. Kanning, U. P. (2005). Computergestützte Personalauswahl von Polizeibeamten [Computer-based personnel selection of police officers]. *Zeitschrift für Arbeits- und Organisationspsychologie*, 49, 140-146.
11. MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion*, 8, 540-551.
12. MacLane, C. N., Barton, M. G., Holoway-Lundy, A. E., & Nickels, B. J. (2001). *Keeping score: Empirical vs. expert weights on situational judgment responses*. Paper presented at the 16th Annual Conference of the Society for Industrial and Organizational Psychology. San Diego, CA.
13. McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braveman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology*, 86, 730-740.
14. McDaniel, M.A. and Nguyen, N.T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9, 103-113.
15. McDaniel, M. A., & Whetzel, D. L. (2005). Situational judgment test research: Informing the debate on practical intelligence theory. *Intelligence*, 33, 515-525.
16. McDaniel, M. A., Whetzel, D. L., Hartman N. S., Nguyen, N. T., & Grubb, W. L. (2006). *Situational judgment tests: validity and an integrative model*. In J. Weekley and R.E. Ployhart (Eds.), *Situational Judgment Tests*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
17. McDaniel, M. A., & Whetzel, D. L. (2007). Situational judgment tests. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied Measurement: Industrial Psychology in Human Resources Management* (pp. 235-257). Mahwah, NJ: Lawrence Erlbaum and Associates.
18. Motowidlo, S. J., Borman, W. C., & Schmit, M. J. (1997). A theory of individual differences in task and contextual performance. *Human Performance*, 10(2), 71-83.
19. Motowidlo, S. J., Diesch, A. C., & Jackson, H. L. (2003). *Using the*

- situational judgment test format to measure personality characteristics.* Paper presented at the 18th Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
20. Motowidlo, S.J., Dunnette, M.D., & Carter, G.W. (1990). An alternative selection procedure: The low-fidelity simulation." *Journal of Applied Psychology*, 75, 640-647.
 21. Motowidlo, S. J., Hanson, & Crafts, J. L. (1997). Low-fidelity simulations. In D. L. Whetzel & G. R. Wheaton (Eds.), *Applied measurement methods in industrial psychology* (pp. 241-260). Palo Alto, CA: Davies-Black Publishing.
 22. Muchinsky, P. M. (2006). *Psychology Applied to Work*. Belmont, CA: Thomson Wadsworth.
 23. Nonaka, I., & Hirotaka, T. (1995). *The Knowledge Creating Company*. N.Y.: Oxford University Press.
 24. O'Connell, M. S., Hartman, N.S., McDaniel, M.A., Grubb, W.L., III, & Lawrence, A. (2007). Incremental validity of situational judgment tests for task and contextual job performance. *International Journal of Selection and Assessment*, 15, 19-29.
 25. Olson-Buchanan, J. B., Drasgow, F., Moberg, P. J., Mead, A. D., Keenan, P. A., & Donovan, M. A. (1998). Interactive video assessment of conflict resolution skills. *Personnel Psychology*, 51, 1-24.
 26. Paullin, C., & Hanson, M. A. (2001). *Comparing the validity of rationally-derived and empirically-derived scoring keys for a situational judgment inventory*. Paper presented at the 16th annual conference for the Society of Industrial and Organizational Psychology, San Diego, CA.
 27. Porr, W. B. & Ployhart, R. E. (2004). *The Validity of Empirically and Construct-Oriented Situational Judgment Tests*. Paper to be presented at the 19th annual conference for the Society of Industrial and Organizational Psychology.
 28. Sacco, J.M., Scheu, C.R., Ryan, A.M., Schmitt, N., Schmidt, D.B., & Rogg, K.L. (2000). *Reading level and verbal test scores as predictors of subgroup differences and validities of Situational Judgment Tests*. Paper presented at the 15th Annual Meeting of the Society for Industrial and Organizational Psychology. New Orleans, LA.
 29. Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of*

- Psychology*, 59, 419-500.
30. Sackett, P. R., & Wanek, J. E. (1996). New developments in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness and reliability for personnel selection. *Personnel Psychology*, 49, 787-829.
 31. Stemler, S. E., & Sternberg, R. J. (2006). Using *situational judgment tests* to measure practical intelligence." In J. Weekley and R.E. Ployhart (Eds.), *Situational Judgment Tests* (pp. 107-131). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
 32. Sternberg, R. J., Wagner, R. K., & Okagaki, L. (1993). *Practical intelligence: The nature and role of tacit knowledge in work and at school*. New Haven, CT: Yale University.
 33. Such, M. J., & Schmidt, D. B. (2004). *Examining the effectiveness of empirical keying: A cross-cultural perspective*. Paper presented at the 19th annual conference for the Society of Industrial and Organizational Psychology, Chicago, IL.
 34. Trippe, D. M., & Foti, R. J. (2003). *An evaluation of the construct validity of situational judgment tests*. Paper presented at the conference of the Society for Industrial and Organizational Psychology, Orlando, Florida.
 35. Wagner, R.K., & Sternberg, R. J. (1985). Practical intelligence in real world pursuits: The role of tacit knowledge," *Journal of Personality and Social Psychology*, 48, 436-458.
 36. Weekley, J. A., Harding, R., Creglow, A., & Ployhart, R. E. (2004). *Scoring situational judgment tests: Does the middle matter?* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.
 37. Weekley, J. A., & Ployhart, R. E. (2006). *Situational judgment tests: Theory, measurement, and application* (Eds.). Mahwah, NJ: Erlbaum.
 38. Whetzel, D.L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 1 , 188-202.