

# 「心理衡鑑初次晤談」客觀結構式臨床測驗 之初探研究\*

藍玉玲\*\*

張玉萱

陳畹蘭

國立東華大學諮商與臨  
床心理學系

元智大學管理學院及大數  
據與數位匯流創新中心

慈濟大學  
人類發展學系

在專業能力風潮影響下，心理學家反思心理師教育訓練模式，強調心理專業能力評量的重要性。在評量上，仍以紙筆測驗與自陳量表為主，較少採實作評量，故本研究擬採用常用的臨床專業能力評量模式「客觀結構式臨床能力測驗」(Objective Structured Clinical Examination, OSCE)，發展衡鑑目的為初次晤談之「心理衡鑑初次晤談 OSCE」(簡稱「初次晤談 OSCE」)，評估碩士生的衡鑑專業能力。初次晤談 OSCE 包含一個標準化病人測驗站(簡稱 SP 站)與一個紙筆站(分成選擇題與簡答題分測驗)。本研究招募 27 位臨床心理碩士生參與研究。研究結果顯示 SP 站各分量表與整體量表的係數皆大於 .8，代表 SP 站具有良好的內部一致性信度。兩評分者在紙筆站簡答題的評分結果也呈現高度正相關( $r = .75$ )。在效度方面，研究結果顯示 SP 站病史詢問與紙筆站簡答題間，呈現中度正相關( $r = .44$ )，顯示初次晤談 OSCE 具有良好的聚斂效度。同樣的，「紙筆站選擇題」與 SP 站各面向間呈現低度正相關，也反映初次晤談 OSCE 具有良好的區辨效度。以新發展測驗而言，研究結果顯示初次晤談 OSCE 具有一定的信效度。未來，有待更多研究的投入，以提供更完整的驗證與論述。

**關鍵詞：**客觀結構式臨床測驗、實作測驗、衡鑑能力評量

## 壹、緒論

### 一、從「專業能力導向觀點」建構「心理衡鑑專業能力評量」

美國心理學會與多個心理專業團體曾於 2002 年舉辦「專業能力研討會：專業心理師的教育與認證之未來方向」(簡稱「專業能力研討會」)之國際研討會，對於心理師專業能

\* 感謝行政院國家科學委員會專題研究計畫經費補助(編號：NSC101-2410-H-259-027)，李維倫與高之梅教授熱心協助宣傳受試者招募，亦感謝匿名審查委員的建議與指正。

\*\* 通訊作者：藍玉玲，花蓮縣壽豐鄉大學路二段 1 號，e-mail: ylan@mail.ndhu.edu.tw

力教育訓練與評量認證，廣邀來自全球不同領域的心理學者專家們，以分組方式進行討論，進而形成心理學界與實務界對於心理專業能力之共識（Kaslow, 2004），也同時帶動專業能力導向的學術研究與系列性專文評析（e.g., Collins, Kaslow, & Illfelder-Kaye, 2004; Kaslow, 2004; Rubin et al., 2007）。

與會專家學者們將心理專業能力分成八個領域，分別是「科學精神實踐」、「衡鑑／診斷／個案概念化」、「治療與處遇」、「諮詢和跨領域專業團隊合作」、「督導」、「專業發展」、「倫理與法律」，以及「個人與多元文化」。Krishnamurthy 等人（2004）更依據與會專家們對於「衡鑑／診斷／個案概念化」的討論結果，將此項專業能力細分成八個面向，分別是「具備基礎心理測驗背景」、「具有心理衡鑑基礎知識」、「具有可評估人們在認知、情感、行為與人格等面向心理特質的心理基礎知識與評估技巧」、「具有衡量治療／介入成效的能力」、「具備批判性評估心理師與個案在衡鑑過程中相互影響的能力」、「具有建立和維持專業關係以提供衡鑑服務的能力」、「了解衡鑑與治療間的關係」與「熟練的衡鑑專門技巧」。

同樣地，Rodolfa 等人（2005）也依據專業能力研討會的討論結果，提出「專業能力發展方塊模式（a cube model for competency development）」（簡稱「Cube Model」）。Cube Model 將心理專業能力評量分成三個向度，分別是「基礎性專業能力（foundational competency）」、「功能性專業能力（functional competency）」、與「專業發展階段（stages of professional development）」。在 Cube Model 中，「衡鑑／診斷／個案概念化」被歸類在「功能性專業能力」向度，是心理師執行心理專業工作所需具備的六項功能性專業能力之一。

Rodolfa 等人（2005）認為執行各項心理專業工作時，除了功能性專業能力之外，也需同時具備「科學知識與方法」、「人際關係」等六項基礎性專業能力。他們將心理師的「專業發展階段」分成「博士教育」、「博士前全職實習」、「博士後督導」、「博士後住院訓練」與「持續能力發展」五個階段，並認為置身於不同專業發展階段的心理師對其功能性與基礎性專業能力應有不同的期待。可惜的是，Rodolfa 等人的 Cube Model 只勾勒出心理專業能力評量的大方向，並未針對各種專業能力在各專業階段的評量內容提出更具體的評量向度與行為指標。因此，Fouad 等人（2009）參考 Cube Model 的理論架構，透過專家意見調查界定不同專業發展階段應具備之「基礎性專業能力」與「功能性專業能力」，以及這些專業能力的評量向度與行為指標，進而形成一份具專家共識之「Competency Benchmarks Document（簡稱 CBD）」供教育工作者與研究人員進行這些專業能力之評估與測量。

不論是 Krishnamurthy 等人（2004）的八大心理衡鑑能力評量向度，或是 Fouad 等人（2009）的 CBD 所包含之衡鑑能力評量向度與行為指標，若將其應用於國內心理師衡鑑能力評量之前，應先評估這些理論架構用於國內心理師群體之適用性。因此，藍玉玲（2010）

透過專家訪談，探索國內心理學界與實務界的專家學者們對於心理衡鑑能力評量範疇的看法，再透過「主題分析法 (thematic analysis)」(王國川、翁千惠譯，2005/1998) 將訪談內容整理成 13 個評量向度與 84 個能力指標，進而完成一份「心理衡鑑能力評量表 (Psychological Assessment Competency Scale, PACS)」。Lan 與 Chang (2014) 透過因素分析對於 235 位具心理師證照之台灣心理師在 PACS 的問卷調查結果加以分析，並提出一個國內心理衡鑑專業能力評量的潛在因素結構。

依據因素分析結果，Lan 與 Chang (2014) 萃取出九個心理衡鑑能力評量向度，每個向度的說明如下：面向 1「科學意識 (scientific mindedness)」代表心理師進行衡鑑實務工作時，需保持批判性思考、問題解決能力，和邏輯分析與組織等能力。面向 2「人際關係 (interpersonal relationship)」代表心理師面對個案提供衡鑑服務時，應懷抱著同理心、認真傾聽，並適時地對於個案的陳述提供適當的回應，而這些能力皆屬於建立維持專業關係的人際能力。面向 3「情感性能力 (affective skills)」代表心理師面對個案提供衡鑑服務時，應保持成熟穩定的情緒、具有耐心，以及挫折容忍度，才能夠維持良好的專業關係。面向 4「心理學科學基礎 (scientific foundation of psychology)」代表心理師提供衡鑑服務時，需具備基礎心理學知識，例如：精神藥理學、社會心理學、精神病理學、人格心理學等。面向 5「專業形象 (professionalism-department)」代表心理師提供衡鑑服務時，所必須保持的專業形象，如：良好的自我控制，適當的臉部表情（非語言行為）與衣著等。面向 6「心理測驗學知識 (knowledge of psychological testing)」代表心理師提供衡鑑服務時，必須具備心理測驗學的專業知識，才能提供良好的衡鑑服務。面向 7「跨領域間的合作 (interdisciplinary systems)」代表對個案提供衡鑑服務時，心理師往往必須與不同領域的專業人員合作，因此心理師必須了解如何和不同專業領域的工作人員溝通及互助合作的能力。面向 8「衡鑑技巧 (technical assessment skills)」代表心理師提供衡鑑服務時，必須具備界定轉介問題、系統性蒐集資料等衡鑑工作的實作技巧。面向 9「心理執業科學基礎 (scientific foundation of professional practice)」代表心理師執行衡鑑業務時，必須具備衡鑑知識、人格理論與人格測量等基礎心理學知識，才能提供良好的衡鑑服務。

這九個向度之中，有四個面向（面向 2、5、6 和 8）與 Krishnamurthy 等人 (2004) 的八大心理衡鑑專業能力評量向度相符，其他面向則與 CBD (Fouad et al., 2009) 的一部份基礎性專業能力相似。這些研究結果部份證實八大心理衡鑑能力評量向度的存在，並呼應 Cube Model 觀點（即，心理專業能力評量應同時包含功能性與基礎性專業能力的評估）。與 Krishnamurthy 等人或是 Fouad 等人 (2009) 提到的衡鑑能力評量模式相比，Lan 與 Chang (2014) 的九大衡鑑能力評量模式所涵蓋的專業能力以衡鑑工作的基本面為主，

較少涉及進階的衡鑑能力面向，例如：透過衡鑑評估治療成效，或是依據個人與文化因素影響來調整衡鑑方式。

若以評量方式而言，目前國內心理專業能力評量以紙筆測驗或自陳量表為主，較少透過實作測驗，評估個人是否具有有良好的實務能力與專業態度（藍玉玲、劉彥君、高聖博，2011）。雖然，紙筆測驗為一種客觀公正的評量方式，但往往僅測得認知層面的專業能力，難以評估行為層面的專業能力；自陳量表則依據個人主觀評分來評估專業能力。當專業能力評量結果涉及入學、升遷等重要決定時，自陳量表恐難有效評估個人的真實能力。因此，本研究擬使用目前醫事人員常用的實務能力評量模式「客觀結構式臨床測驗(Objective Structured Clinical Examination, OSCE)」，發展「心理衡鑑 OSCE」的實作評量評估處於「碩士教育」專業發展階段的心理師是否具有提供衡鑑服務所需之專業知識、技術與態度。

## 二、心理衡鑑專業能力評量—以「初次晤談 OSCE」為例

近四十年來，醫學（Walsh, Bailey, & Koren, 2009）、護理學（Rushforth, 2007），以及其他醫事相關領域（Fang et al., 2013），紛紛採用 OSCE，對於醫事人員進行知識、技術與態度的綜合性專業能力評量。OSCE 也是目前醫事人員專業能力評量的常用評量模式之一，用於評估臨床專業訓練成效、是否達到執業水準，以取得專業證照(Hodges, 2006)。因此，我們認為 OSCE 或許可用於心理衡鑑專業能力評量，評估個人是否具有衡鑑工作所需的專業知識、技術與態度。

由於 OSCE 是一種高度模擬臨床情境的實作評量（Performance Assessment, PA），難以透過單一 OSCE 測得心理衡鑑所需具備之各項專業能力。再加上，心理衡鑑是一個問題回答與問題解決的歷程，必須先界定一個待回答的問題／衡鑑目的時，才能夠進入此問題回答與解決的過程（許文耀譯，1997/1976）。因此，發展「心理衡鑑 OSCE」之前，本研究先將此測驗的衡鑑目的設定為「初次晤談（intake interview）」，擬發展一套以「初步鑑別診斷個案心理疾病狀態與個人現況」為衡鑑目的之「心理衡鑑初次晤談 OSCE」（簡稱「初次晤談 OSCE」）。關於「初次晤談 OSCE」的評量對象，本研究參考藍玉玲等人（2011）對於「Cube Model」之「專業發展階段」的修訂，選擇專業發展階段在「碩士教育」層級之碩士班學生為評量對象，衡量他們在「心理衡鑑初次晤談」的專業性。其他有關「初次晤談 OSCE」的測驗發展，將於下一章節逐一說明。

## 貳、研究方法

### 一、研究參與者

本研究以「尚未進入一年全職實習」且「未有心理諮詢全職工作經驗」為樣本篩選標準，透過網路發佈招募廣告至臨床心理相關系所與台灣臨床心理學會，招募臨床心理碩士生為研究樣本，共有 27 位來自全國七所心理相關系所之碩士生參與正式施測。表一為這些受試學生的基本資料描述。

表一 受試者背景變項與測驗情境之次數分配 (N = 27)

背景變項		人數	百分比
1. 性別	男性	9	33.3
	女性	18	66.7
2. 年齡	23歲	6	22.2
	24歲	7	25.9
	25歲	5	18.5
	26歲及26歲以上	6	22.2
	遺漏值	3	11.2
	平均年齡：25.1歲 標準差：2.42歲		
3. 就讀學校	A校	8	29.6
	B校	6	22.2
	其他五校	13	48.2
4. 年級	碩一	20	74.1
	碩二	7	25.9
5. 選修衡鑑相關課程數	1門	8	29.6
	2門	11	40.7
	3門及3門以上	8	29.6
測驗情境		施測（評分）人數	百分比
6. 考試日期	第1天	10	37.0
	第2天	10	37.0
	第3天	7	25.9
7. 標準化病人	A	7	25.9
	B	12	44.5
	C	8	29.6

表一 (續)

測驗情境		施測 (評分) 人數	百分比
8. 「SP站」評分者	1	9	33.3
	2	8	29.6
	3	4	14.9
	4	6	22.2
9. 「紙筆站」簡答題評分者	X	27	評分者信度
	Y	27	$r = .75^{**}$

註： $**p < .01$ , two-tailed.

## 二、研究程序

本研究所設計之「初次晤談 OSCE」包含一個 20 分鐘的「標準化病人 (Standardized Patient, SP) 測驗站」(簡稱「SP 站」)與一個 90 分鐘的「紙筆測驗站」(簡稱「紙筆站」)。為了提高測驗信效度並同時模擬心理衡鑑的部份歷程，本研究將這兩個測驗站設計成連貫站型式 (即，「SP 站」與「紙筆站簡答題」皆使用相同的個案內容與就診目的)，僅以不同測驗型式 (實作評量 vs. 紙筆測驗)，評估受試者不同面向之「心理衡鑑初次晤談」的專業能力。所有受試學生都先進入 SP 站，開始接受初次晤談 OSCE 的實作測驗，待 SP 站測驗完畢後，再進入紙筆站接受紙筆測驗。在考試前，所有受試學生都需接受試前說明與受試者同意說明，以了解本測驗各站的考試目的、試場規則、考試時間與動線說明，並了解本研究的研究內容。待簽署受試者同意書交給工作人員後，受試者再依試務人員之引導，依序接受考試。

## 三、「初次晤談 OSCE」之測驗發展

「初次晤談 OSCE」之測驗發展可分為「測驗內容設計」與「測驗實務規劃」兩部份。前者包含「測驗站設計」和「測驗向度與題項設計」，而後者包含「標準化病人招募訓練」、「評分者招募訓練」和「測驗實務規劃」。

### (一)「初次晤談 OSCE」之測驗內容設計

在「測驗站設計」部份，本研究希望模擬心理衡鑑的部份歷程，將「初次晤談 OSCE」分成「SP 站」與「紙筆站」兩站。SP 站將提供「心理衡鑑初次晤談」的模擬情境，由本研究所訓練之標準化病人 (SP) 擔任模擬病人的角色，由受試者扮演「心理師」的角色與 SP 互動，進而評估受試者是否具有「初步鑑別診斷個案心理疾病狀態與個人現況」之

「初次晤談」所需的專業知識、技術與態度。SP 站的評量內容包含變態心理與病理診斷的專業知識、心理衡鑑的專門技巧，以及與個案建立維持衡鑑專業關係之專業態度，並透過 SP 站評分表加以評估。此評分表的評量向度與題項設計將於「SP 站評量項目與題項設計」章節說明，而此評分表內容效度的驗證，也將於「初次晤談 OSCE 之內容效度」章節闡述。紙筆站透過一份紙筆測驗（包含「選擇題」與「簡答題」），評估受試者是否具有測驗與衡鑑基礎知識，是否能依據個案轉介目的，有效地應用與整合 SP 站所獲取的資訊（如：行為觀察與臨床晤談的各項資訊），使用適當的測驗工具來了解個案的心理狀況，進而驗證對於個案病況的初步假設／推論。

關於 SP 站模擬情境的設計，本研究採用臨床工作常見的「憂鬱症」（Kring, Davison, Neale, & Johnson, 2007）作為模擬病人的病症，設計此模擬情境的內容（如：個人成長史、婚姻與家庭等人際支持系統），使 SP 擁有充份資訊模擬個案的可能狀態。有關各站的測驗時間，SP 站與紙筆站的測驗時間分別為 20 分鐘與 90 分鐘。醫學 OSCE 的 SP 站之測驗時間通常是 4 至 15 分鐘（Walsh et al., 2009），本研究的 SP 站較一般醫學 OSCE 的 SP 站之測驗時間長，是因為心理晤談的實際看診時間通常較長，且心理晤談所需詢問的內容較多元，需要較多晤談時間來對於個案有一個較系統性的了解。本研究紙筆站的考試範圍與題型則是參考國內心理師高等考試的考試範圍與試題內容來設計紙筆站的試題內容，並依據紙筆站的試題內容訂定紙筆站的測驗時間。

在「測驗向度」部份，本研究除了參考 Krishnamurthy 等人（2004）的八大衡鑑能力評量向度，也依據藍玉玲（2010）的研究結果，從國內心理學界與實務界認為最重要的 20 項心理衡鑑專業能力中，選取十項最基礎的衡鑑專業能力，修改調整成「初次晤談 OSCE」的評量內容。表二為「初次晤談 OSCE」的雙向細目表，第一欄為本測驗的十項評量內容，分別屬於知識、技術與態度三個面向，將透過 SP 站與紙筆站加以評估。這些評量內容亦可納入 Krishnamurthy 等人的八大心理衡鑑能力評量向度（見表二第二欄）之中。以下將分別說明 SP 站與紙筆站的評量項目與題項設計。

表二 「初次晤談OSCE」雙向細目表

初次晤談OSCE 之評量內容	Krishnamurthy等人 (2004)之八大心理 衡鑑能力評量向度	初次晤談OSCE 之測驗工作站	
		SP站	紙筆站
<b>知識面</b>			
1. 選擇適當的測驗工具			簡答第四題
2. 測驗內容與適用範圍	向度1：心理測驗 之基礎知識		簡答第四題
3. 測驗分數的計算與解釋	向度2：心理衡鑑 之基礎知識		簡答第二題
4. 心理測驗與心理衡鑑的基本 概念			選擇題
5. 變態心理學及病理診斷	向度3：心理學基 礎知識與評估技巧	向度1：病史詢問	簡答第一題
6. 界定轉介問題	向度8：心理衡鑑 之專門技巧	向度2：溝通技巧 與專業態度	
<b>技術面</b>			
7. 有效地應用不同的衡鑑方法 進行資料蒐集（行為觀察、 臨床晤談與心理測驗）		向度1：病史詢問	簡答第三題
8. 資料整合分析與個案概念化		向度1：病史詢問	簡答第三題 簡答第五題
<b>態度面</b>			
9. 基本溝通技巧（互動能力、 同理心、傾聽）	向度6：與個案建 立維持衡鑑專業關 係之能力	向度2：溝通技巧 與專業態度	
10. 關係建立與維持		向度2：溝通技巧 與專業態度	

### 1. 「SP 站」評量項目與題項設計

受試者「SP 站」的表現將透過「SP 站評分表」，由受過 OSCE 評分者訓練之評分者評估。此 SP 站評分表包含 28 個三點量尺（即，「未做」、「已做，待改進」和「已做，表現良好」）的評分題項，這些題項可歸納成「病史詢問」和「溝通技巧與專業態度」兩向度。SP 站測驗總分界於 28 ~ 84 分之間，分數越高代表受試者 SP 站的表現越好。「病史詢問」向度包含 18 個評分題項，用於評估受試者與 SP 互動時，是否詢問一些重要問題來判斷

SP 的憂鬱狀況，藉以評估受試者對於憂鬱症在「變態心理學及病理診斷」的基礎知識。「溝通技巧與專業態度」向度包含 10 個評分題項，用來衡量受試者與 SP 互動時，是否具有良好的溝通能力與專業態度。這些題項將有助於評估受試者在「基本溝通技巧」和「關係建立與維持」的專業能力，也同時反映出受試者「與個案建立維持衡鑑專業關係」的專業能力。

除此之外，受試者與 SP 互動時，通常會依據 SP 的回答內容，提出進一步追問或是重覆詢問以確認清楚地了解 SP 欲表達的內容。這些詢問、追問與溝通確認的詢問內容與確認事項，都用來評估受試者進行初次晤談的實務技巧。即用來評估受試者在「界定轉介問題」、「有效地應用不同的衡鑑方法進行資料蒐集」與「資料整合分析與個案概念化」的專業能力。以「病史詢問」面向為例，評分者判斷受試者在變態心理學與病理診斷的知識水準是以受試者與 SP 互動時的「詢問內容」，判斷其「病史詢問」的表現；而評分者對於受試者衡鑑技巧的評估，則是依據受試者與 SP 互動時，能否觀察 SP 反應與答覆內容，適時地修正問題方向或是調整詢問方式，來衡量受試者能否有效地應用不同的衡鑑方法（即行為觀察與臨床晤談）進行資料蒐集，形成對個案的種種假設或推論。表二第三欄為 SP 站評量向度所對應之「初次晤談 OSCE」的評量內容。

## 2. 「紙筆站」評量項目與題項設計

「紙筆站」透過「選擇題」與「簡答題」評估受試者在心理衡鑑的專業能力。「選擇題」部份包含 30 題選擇題，以心理衡鑑與心理測驗的基礎知識為主，用來衡量受試者是否具有執行心理衡鑑所需具備的基本衡鑑知識與測驗學的基礎知識。「簡答題」部份包含五題簡答題，以及模擬病人在兩種測驗的測驗結果和這兩種測驗的指導手冊。「簡答題」的試題設計是模擬心理衡鑑的部份歷程，假設受試者已完成 SP 站的初次晤談，取得與衡鑑目的有關之重要訊息，擬進一步透過測驗工具評估個案的心理狀態，以驗證對於個案的種種假設／推論。因此，受試者必須整合「晤談資料」、「測驗結果」和「衡鑑相關的專業知識（如：變態心理學、心理測驗學與心理衡鑑的專業知識）」，才能回答簡答題中有關個案概念化、初步評估和處遇方案提供等問題。「紙筆站」的「選擇題」與「簡答題」各佔 60 分，紙筆站的測驗總分界於 0 ~ 120 分之間，分數越高代表受試者在紙筆站的表現越好。表二第四欄為紙筆站評量向度所對應之「初次晤談 OSCE」的評量內容。

### （二）「初次晤談 OSCE」之測驗實務規劃

「初次晤談 OSCE」的測驗實務規劃與一般 OSCE 的測驗實務規劃差異不大，故本研究僅針對此測驗與其他 OSCE 規劃不同之處加以說明。關於心理學 OSCE 的測驗實務規劃，

可參考藍玉玲等人（2011）一文。

### 1. 標準化病人招募訓練

本研究依據「初次晤談 OSCE」的憂鬱個案內容，設定憂鬱症標準化病人（SP）的招募條件，透過校園廣告公開招募。為避免志願者因扮演憂鬱症標準化病人時過於投入而導致個人情緒起伏過大或是促發心理疾病，本研究透過中文版貝克憂鬱量表（陳心怡譯，2000/1996）與個人心理狀態評估面試，排除曾患有心理疾患或憂鬱傾向較高之志願者後，共有六人參與本研究憂鬱個案之 SP 訓練。在 SP 訓練方面，本研究參考高聖博（2008）的標準化病人訓練方式，透過三次 SP 訓練工作坊，進行表演技巧訓練與 SP 扮演訓練。在表演技巧訓練方面，邀請具有多年劇場表演經驗與演出經驗之講師，訓練 SP 扮演者的臉部表情、肢體語言和角色揣摩。在 SP 扮演訓練方面，除了講述 SP 個案的性格、成長過程，以及婚姻與家庭等人際關係之外，也透過時間軸與重要事件來闡述腳本內容與個案的心理病史和促發原因，使 SP 扮演者有充份訊息以揣摩角色的性格、臉部表情和肢體動作。待完成 SP 扮演的初步訓練後，本研究透過一次分組模擬演練，使 SP 扮演者熟悉扮演 SP 的注意事項和訊息揭露原則（如：不自動提供答案、不用話來引導學生）。

最後，本研究透過預試，檢驗六位 SP 的表演品質與穩定度。預試時，每位 SP 連續扮演兩次標準化病人，接受兩位預試學生 20 分鐘「初次晤談」模擬詢問。本研究也透過錄影記錄這些晤談互動過程，再由本文第一與第三作者評估錄影資料中每位 SP 的表演品質、與學生互動溝通能力、訊息揭露合理性，和持續扮演 SP 的穩定度，並選出三位具有良好表演品質與表演穩定度的 SP，擔任正式施測的 SP。

### 2. 評分者招募訓練

本研究招募四位曾參與「心理學 OSCE 工作坊」且具有心理師證照的心理師擔任 SP 站的評分員，評估受試者在「SP 站」所展現的衡鑑專業能力。評分者訓練分成「心理學 OSCE 工作坊」與「評分者共識訓練」兩階段訓練。藍玉玲（2012）曾舉辦「心理學 OSCE 工作坊」，試圖建立「心理學 OSCE」之「評分者訓練」模式，並藉以培訓一批具有「心理學 OSCE」評分者訓練之心理師，協助「初次晤談 OSCE」SP 站的評分工作。此「心理學 OSCE 工作坊」包含「心理學 OSCE 之線上教學課程」與「評分者共識訓練工作坊」。所有參與者須事先閱讀「心理學 OSCE 之線上教學課程」，包含 OSCE 評量模式與此類實作測驗評分原則的介紹。待這些心理師皆已熟悉 OSCE 的基本概念與評分原則後，再開始接受 6~8 小時的「評分者共識訓練工作坊」。

「評分者共識訓練工作坊」包含兩階段評分演練與評分共識討論。第一階段評分演練

時，所有評分者先使用研究者提供的 SP 站錄影內容，獨立地使用 SP 站評分表，評估每段影片中受試學生與 SP 互動的表現。待所有評分者皆完成第一階段評分演練後，再進入第一階段評分共識討論。在第一階段評分共識討論時，研究者與所有評分者依據評分者對每位受試學生的評分結果，逐一討論評分者對於 SP 站評分表中每個評分項目的評量標準（如：哪些行為表現屬於「已做，表現良好」、哪些屬於「已做，待改進」或「未做」）。待評分者間已形成初步共識後，再展開第二階段評分演練，確認評分者對於相同學生在每題評分結果十分相似，再針對評分差異較大的題項，進行第二階段的評分共識討論，更明確界定哪些行為表現應歸屬哪個評量標準。待本研究展開「初次晤談 OSCE」的正式施測前，研究者再展開第二階段「評分者共識訓練」，使用「初次晤談 OSCE」預試參與者和 SP 互動的錄影內容，進行評分者共識訓練，確保評分者對於相似的行為表現皆採用相似標準評分，進而達到評分的公平客觀性。以「適當地回應個案的談話內容」題項為例，學生若能依據 SP 回答內容，適時提出進一步追問釐清病況，或是注意到 SP 欲言又止的態度，適時提供支持性言語安撫鼓勵 SP，讓 SP 感到被同理、願意回答較敏感的問題，將被評為「已做，表現良好」。若學生只以點頭或其他肢體動作／語言來展現傾聽的態度，卻未適時地提問或表達支持來進一步了解個案病況時，將被評為「已做，待改進」。若學生只詢問自己想問的問題，卻未注意到 SP 的回答內容，將被評為「未做」。

紙筆站的「選擇題」部份，有標準答案，故由本研究的研究人員協助計分，並未特別訓練評分人員；而紙筆站的「簡答題」部份，大量涉及 SP 站個案內容與紙筆站 SP 個案的模擬測驗結果，不易訓練本研究之外的心理師參與評分，故由本文的第二與第三作者先召開評分標準會議，擬定初步評分標準後，再使用預試資料進行評分，以建立評分者間的評分標準。待確立評分標準後，再開始評估正式研究時受試者在紙筆站「簡答題」的表現。若兩位評分者對於同一位受試者在某一題「簡答題」評分差異大於 10 分時，則由本文第一作者獨立評分後，再與兩位評分者討論，以確定受試者在此題的最終得分。

#### 四、資料分析

本研究使用 SPSS16.0 進行敘述性統計、項目分析、信效度檢驗、兩相依樣本 t 檢定，以及無母數統計之 Mann-Whitney 與 Kruskal-Wallis 檢定。由於本研究的樣本數僅有 27 人，並未達到母數統計樣本數的基本要求 ( $N \geq 30$ ) (Gravetter & Wallnau, 2012)，故後續多項統計分析大多採用無母數統計檢定，評估受試者背景變項與測驗情境對受試者「初次晤談 OSCE」測驗分數之影響。「初次晤談 OSCE」的信效度檢驗將透過內部一致性信度、皮爾森相關 (Pearson's correlation)、Mann-Whitney 與 Kruskal-Wallis 檢定檢驗。

## 參、研究結果

### 一、背景變項與測驗情境對於受試者「初次晤談 OSCE」測驗分數之影響

本研究首先透過 Mann-Whitney 與 Kruskal-Wallis 檢定，檢驗來自「不同背景」的受試者在「初次晤談 OSCE」各站的測驗分數是否具有顯著差異。研究結果顯示不論在 SP 站或是紙筆站，不同性別的受試者在「初次晤談 OSCE」兩測驗站的測驗分數上，皆未呈現顯著差異（SP 站部份，性別： $U = 46.50$ ， $p = .32$ ；紙筆站部份，性別： $U = 70.50$ ， $p = .59$ ）。同樣的，來自不同學校的受試者在 SP 站的測驗分數也未呈現顯著差異（ $\chi^2(2) = 2.04$ ， $p = .36$ ）；但是他們在紙筆站的測驗分數則已呈現出顯著差異（ $\chi^2(2) = 6.08$ ， $p = .05$ ），其中來自 B 校的受試者在紙筆站的測驗分數（ $M = 68.83$ ， $SD = 4.13$ ）明顯高於來自 A 校和其他五校的受試者（A 校： $M = 57.13$ ， $SD = 10.76$ ；其他五校： $M = 61.38$ ， $SD = 5.67$ ）。

進一步比較來自不同學校受試者的「年級」與「選修衡鑑課程數」時，發現與 A 校和其他五校受試者相比，來自 B 校的受試者以碩二學生為主，且這些受試者中有 66.7% 的人已修過至少三門的衡鑑課程。因此，我們猜測不同學校受試者在紙筆站所呈現的分數差異未必是因為學校訓練成效不同所致，而是受到「年級」與「選修衡鑑課程數」的影響。至於受試者在 SP 站測驗分數可能會受到受試者認知學習和實際操作能力的影響，以及選修衡鑑課程數和本站測驗分數間的關聯性，將於「初次晤談 OSCE」之「效標關聯效度」章節說明。

在「測驗環境」部份，本研究也透過 Kruskal-Wallis 檢定與兩相依樣本 t-test，檢驗不同測驗環境對於受試者「初次晤談 OSCE」測驗分數的影響。研究結果顯示考試日期與標準化病人的不同對於受試者在 SP 站與紙筆站的測驗分數皆無顯著影響（SP 站部份，考試日期： $\chi^2(2) = 2.95$ ， $p = .23$ ；標準化病人： $\chi^2(2) = 4.91$ ， $p = .09$ ；紙筆站部份，考試日期： $\chi^2(2) = 1.04$ ， $p = .60$ ；標準化病人： $\chi^2(2) = 0.57$ ， $p = .75$ ）。然而，比較評分者間的受試者測驗分數時，SP 站呈現顯著差異（ $\chi^2(3) = 9.32$ ， $p < .05$ ）；紙筆站「簡答題」雖呈現顯著差異（ $t(26) = 5.74$ ， $p < .01$ ），但其評分者信度達高度相關（ $r = .75$ ， $p < .01$ ）表示評分標準雖有差別但已達一致性。

研究結果顯示除了評分者效應之外，受試者在「初次晤談 OSCE」SP 站與紙筆站的測驗分數並未因為考試日期與標準化病人不同而呈現顯著差異。這些結果似乎反映出本研究不論在 OSCE 測驗實務規劃或是標準化病人訓練，已達到一定程度的標準化。有關評分者效應，將於討論章節闡述。

## 二、「初次晤談OSCE」之內容效度

本研究發展之「初次晤談 OSCE」包含一個 SP 站與一個紙筆站。前者用來評估變態心理與病理診斷的專業知識、心理衡鑑的專門技巧，以及與個案建立維持衡鑑專業關係之專業態度；後者著重於心理測驗與衡鑑基礎知識，和心理衡鑑專門技巧的評估。本文將逐一說明兩個測驗站的內容效度。

### (一) SP站評分表

本研究在「初次晤談 OSCE」SP 站所使用之評分表是採用藍玉玲(2012)所設計之「憂鬱個案評分表」。藍玉玲透過「專家意見調查」與「專家共識會議」，完成「憂鬱個案評分表」的題項修訂。在「專家意見調查」部份，研究結果發現專家們對於「溝通技巧與專業態度」面向的大多數題項具有中高度共識，並依據重要程度將各題分類至「重要且必要」或「重要」類別。在「溝通技巧與專業態度」面向，專家們僅對於第三題「稱呼個案(先生、小姐)」呈現較分歧的看法。近半數專家認為此題屬於「重要且必要」的類別，也有近半數專家將此題歸類於「重要」的類別。雖然此題無法取得過半數專家同意將其明確地分類至「重要且必要」或「重要」類別，但就此結果而言，應可將此題視為重要且應保留在評分表中的題項。因此，依據專家意見調查結果，「溝通技巧與專業態度」面向的所有題項皆為重要題項，皆應保留在 SP 評分表中。

在「病史詢問」面向，專家意見調查結果也呈現相似現象。大多數題項都具有中高度專家共識，但在第 10 題、第 15 題與第 19 題則呈現專家意見分歧的狀況。近半數專家認為第 10 題屬於「不重要」的類別，也有近半數專家將此題歸類於「重要」的類別。關於第 15 題與第 19 題，近半數專家認為屬於「重要且必要」的類別，也有近半數專家認為應歸類在「重要」的類別。對於專家意見分歧的題項，藍玉玲(2012)透過「專家共識會議」針對這些題項進行討論。依據「專家共識會議」的討論結果，再次修訂「憂鬱個案評分表」的題項描述，並對於部份題項加以整併。例如：「病史詢問」面向第 10 題「性生活」經專家討論後，建議可合併至第 9 題「親密關係或婚姻關係」中。

整體而言，藍玉玲(2012)透過「專家意見調查」與「專家共識會議」檢驗「憂鬱個案評分表」中每一個評分題項的重要性，並對於題項內容進行再次修訂與整併，進而建立一份具有專家效度之「憂鬱個案評分表」。此評分表包含 28 個三點量尺的評分題項，並可分成「病史詢問」與「溝通技巧與專業態度」兩個向度，以評估碩士班學生面對憂鬱個案進行「心理衡鑑初次晤談」時，是否具有有良好的專業知識、技巧與態度。

### (二) 紙筆站的紙筆測驗

本研究在紙筆測驗設計完成後，邀請六位教授心理測驗或心理衡鑑課程之大學教授評

估此紙筆測驗「選擇題」的試題品質、題項難度與重要性。依據專家們的評估結果，從原本四十道選擇題中，刪除太難或太簡單但不重要的題項，保留三十題難度分佈適中且題目重要性高的題項作為紙筆站選擇題考題。本研究紙筆測驗「簡答題」因為涉及 SP 站的個案內容，且包含 SP 個案之模擬測驗結果，較不易透過外部專家評估試題品質。因此，「簡答題」部份採用研究團隊內專家討論與預試測驗結果來進行試題品質與難度評估。在研究團隊內專家討論部份，本研究第二作者先依據「初次晤談 OSCE」憂鬱個案內容，設計「簡答題」題項描述，接著由本研究第一與第三作者進行題項內容修訂，完成「簡答題」初稿。本研究再進一步透過預試檢驗「簡答題」初稿的測驗品質，最後則依據預試結果再次修訂簡答題的題項描述與試題難度，進而完成「初次晤談 OSCE」紙筆站的簡答題題項。

### 三、「初次晤談 OSCE」之試題分析與信度分析

本研究發展之「初次晤談 OSCE」包含 SP 站與紙筆站，這兩個測驗站在評量內容或測驗型式皆有所差異，難以視為單一向度，故試題分析時，將分別計算各站的測驗總分，作為各站高低分組的分組依據，並以各站測驗總分在 73% 以上者為高分組，27% 以下者為低分組。

#### (一)「SP 站」憂鬱個案評分表

本研究透過極端組檢核法與同質性檢核法，並依據吳明隆與涂金堂（2005）建議的三項刪題標準，衡量「SP 站憂鬱個案評分表（簡稱 SP 站評分表）」的試題品質：（1）兩獨立樣本 t-test 評估高低分組學生在各題項之表現是否具有顯著差異，作為刪題標準。囿於樣本人數較少，本研究以無母數統計之 Mann-Whitney 檢定代替兩獨立樣本 t-test 進行統計分析；（2）計算各題項與其所屬分測驗間的關聯性，刪除相關係數未達 .3 之題項；（3）計算各題刪除後分測驗整體  $\alpha$  係數，刪除整體  $\alpha$  係數在刪題後提高之題項。

表三為 SP 站評分表的試題分析與信度分析摘要表。基於高低分組人數皆  $\leq 10$  人的情況下，本研究以符合上述三項刪題標準作為題項刪除的評估原則。依據研究結果，SP 站評分表共有兩題可考慮刪除，分別是「病史詢問」向度第 8 題「自殺危險評估」與「溝通技巧與專業態度」向度第 1 題「招呼並與個案自我介紹」。當檢視上述試題的題項描述後，我們認為「自殺危險評估」屬於憂鬱個案臨床晤談時必須評估項目，且此題項在藍玉玲（2012）的憂鬱個案評分表修訂研究中被大部份專家認為是必要題項，故不考慮刪除。在此評分表修訂研究中，也曾討論「招呼並與個案自我介紹」的重要性，專家們認為對於碩士班學生而言，此題應予以保留，以衡量學生是否具有「心理衡鑑初次晤談」的基本能力。未來，若此評分表應用於執業心理師群體時，則可因為題項過於簡單而予以刪除。

表三 「初次晤談OSCE」SP站之試題分析與信度分析摘要表

題號	題目	Mann Whitney U值	與分測驗 總分相關	刪題後分測 驗總 $\alpha$ 係數
病史詢問	1 壓力的可能來源	3.50**	.63	.78
	2 目前的情緒狀況（型態、期間）	5.50**	.62	.77
	3 平日興趣與活動參與狀況的改變	11.50	.49	.78
	4 生理狀況評估（睡眠習慣、食慾體重變化）	0.00**	.64	.77
	5 認知功能影響評估（記憶力、注意力）	18.50	.18	.80
	6 個案面對憂鬱症狀影響之因應方式	15.00	.28	.79
	7 目前用藥狀況（藥名、服藥次數、成效）	8.00*	.37	.79
	8 自殺危險評估（意念／行動、頻率、強度）	22.50	-.01	.81
	9 學業／工作狀況	10.50*	.47	.78
	10 與家人的關係（父母、手足、姻親）	6.00*	.45	.78
	11 親密關係或婚姻關係	9.50*	.38	.79
	12 社會支持系統（宗教、社團、朋友、人際關係）	12.50	.24	.80
	13 與主訴有關的過去病史	10.50*	.39	.79
	14 與主訴有關的家族病史	24.50	.16	.80
	15 創傷經驗或受虐經驗	21.00	.10	.80
	16 過去心理治療經驗	1.00**	.69	.76
	17 過去其他求助經驗（宗教、算命等非正式心理治療方式）	17.50	.29	.79
	18 個案對於治療效果的期待	6.50*	.49	.78
面向1：病史詢問				<b>.80</b>
溝通技巧與專業態度	1 招呼並向個案自我介紹（姓名、職稱）	17.50	.12	.91
	2 提出適當問題來詢問／確認個案的就診目的	3.00*	.81	.85
	3 適當地回應個案的談話內容	3.50*	.83	.85
	4 用個案能理解的語言與溝通方式	10.50*	.65	.86
	5 技巧性地詢問敏感問題（如：自殺與創傷經驗詢問）	18.00	.52	.88
	6 計畫性地提問來蒐集個案的重要資訊	0.00**	.63	.86
	7 呈現溫暖接納的專業態度	0.00**	.82	.85
	8 適時展現同理心	0.00**	.86	.85
	9 合宜的非語言／肢體的回應	0.00**	.71	.86
	10 簡單摘要本次晤談內容並安排後續晤談事宜	5.00**	.55	.87
面向2：溝通技巧與專業態度分測驗				<b>.88</b>
SP站「憂鬱個案評分表」				<b>.89</b>

註：\* $p < .05$ , two-tailed; \*\* $p < .01$ , two-tailed.

關於 SP 站評分表的內部一致性信度，SP 站評分表整體  $\alpha$  係數為 .89，「病史詢問」和「溝通技巧與專業態度」向度分別為 .80 與 .88。這些研究結果顯示 SP 站評分表與兩個分測驗皆達到 DeVellis (1991) 建議的新測驗內部一致性信度之下限 (i.e.,  $\alpha \geq .7$ )，表示 SP 站評分表具有良好的內部一致性信度。

## (二) 「紙筆站」紙筆測驗

紙筆站的紙筆測驗包含 30 題「選擇題」與 5 題「簡答題」，故試題分析將依據題型不同而分開分析。「選擇題」的試題分析將透過「難度」、「鑑別度」與「點二系列相關」加以評估。依據余民寧 (2011) 的建議，選擇題的難度最好界於 0.4 至 0.8 之間，而題項的「鑑別度」最好在 0.3 以上；若界於 0.2 ~ 0.29 之間，代表題項品質尚可接受，但要進行局部修改；若在 0.19 以下，代表題項品質不佳，需考慮刪除或進行大幅修改。與 SP 站評分表的試題分析相似，「簡答題」的試題分析將透過極端組檢核法與同質性檢核法，並依據吳明隆與涂金堂 (2005) 建議的三項刪題標準加以評估。

### 1. 「選擇題」的試題分析

就「難度」而言，30 題選擇題的難度界於 0.15 至 1.00 之間，平均難度為 0.72，表示這份選擇題的難度偏簡單。就「鑑別度」而言，選擇題題項的鑑別度界於 -0.13 至 0.55 之間。就「點二系列相關」而言，選擇題題項的點二系列相關界於 -0.17 至 0.56 之間。表四為紙筆站「選擇題」的試題分析摘要表。綜合上述研究結果，在 30 題選擇題中，有 9 題品質良好，12 題品質尚可有待局部修改，4 題品質不佳需考慮刪除，有 5 題(即，第 1、9、19、22 與 23 題) 雖然難度較低且鑑別度較低，但因為是心理測驗與心理衡鑑的重要基礎知識，故予以保留，以評估碩士班學生是否具有這些基礎知識。

表四 「初次晤談 OSCE」紙筆站「選擇題」的試題分析摘要表

選擇題題號	難度	鑑別度	Pbis	試題分析
1	0.93	0.10	0.37	重要基礎知識，保留*
2	0.96	-0.13	-0.17	刪除
3	0.74	0.18	0.14	品質尚可，有待修改
4	0.15	0.03	0.02	刪除
5	0.41	0.33	0.25	品質良好，保留
6	0.59	0.25	0.22	品質尚可，有待修改
7	0.41	0.53	0.44	品質良好，保留
8	0.67	0.48	0.45	品質良好，保留
9	0.96	0.10	0.12	重要基礎知識，保留*

表四 (續)

10	0.89	0.20	0.46	品質尚可，有待修改
11	1.00	0.00		刪除
12	0.41	0.20	0.17	品質尚可，有待修改
13	0.78	0.08	0.09	品質尚可，有待修改
14	0.78	0.40	0.48	品質良好，保留
15	0.82	0.40	0.56	品質良好，保留
16	0.30	0.53	0.47	品質良好，保留
17	0.52	0.55	0.46	品質良好，保留
18	0.41	0.43	0.38	品質良好，保留
19	0.96	0.10	0.19	重要基礎知識，保留*
20	0.96	0.00	0.05	品質尚可，有待修改
21	0.78	0.18	0.06	品質尚可，有待修改
22	0.96	0.10	0.47	重要基礎知識，保留*
23	0.93	0.10	0.17	重要基礎知識，保留*
24	0.82	0.18	0.11	品質尚可，有待修改
25	0.67	0.18	0.16	品質尚可，有待修改
26	0.89	0.10	0.34	品質尚可，有待修改
27	0.85	0.30	0.32	品質良好，保留
28	0.89	0.20	0.21	品質尚可，有待修改
29	0.93	0.10	0.17	品質尚可，有待修改
30	0.30	-0.05	0.00	刪除

註：第1、9、19、22與23題雖然難度低且鑑別度低，但因為題項內容為衡鑑／測驗的重要基礎知識，故予以保留。

## 2. 「簡答題」的試題分析

依據吳明隆與涂金堂（2005）的建議，本研究也使用 Mann-Whitney 檢定、題項與其所屬分測驗之相關，以及刪題後分測驗之整體  $\alpha$  係數來評估紙筆站「簡答題」的試題品質。本研究以吳明隆與涂金堂提出的三項刪題標準作為題項刪除的評估原則，依據研究結果，簡答題第2題可考慮刪除。當檢視此題的題項描述後，發現此題是用來評估受試者是否具有使用測驗工具指導手冊解釋測驗分數的能力，屬於衡鑑實務工作的基本能力，應予以保留，以評估碩士班學生是否皆已具備此項基本能力。在內部一致性信度部份，由於簡答題的評量內容（見表二）並非單一向度，包含測驗工具的選擇、測驗分數計算與解釋、變態心理學和病理診斷、整合各類衡鑑資料進行綜合分析、個案概念化，與介入方案提供。因

此，紙筆站簡答題部份或許具有一定的內容效度，但因為各題評量向度差異較大，而導致其內部一致性信度偏低（ $\alpha$  係數為 .25）。未來在設計紙筆站簡答題時，宜仔細考慮如何在信度與效度之間取得平衡，此部份將於討論章節說明。

#### 四、「初次晤談 OSCE」之構念效度分析

本研究採用多特質—多方法分析 (Multitrait-multimethod, MTMM) (Campbell & Fiske, 1959) 檢驗「初次晤談 OSCE」的構念效度。表五為「初次晤談 OSCE」之多特質—多方法矩陣。研究結果顯示「SP 站」的「病史詢問」與「溝通技巧與專業態度」之間，呈現高度正相關（ $r = .71, p < .01$ ），且這兩個分測驗的測驗分數皆與 SP 站測驗總分之間，呈現高度正相關（病史詢問： $r = .95, p < .01$ ；溝通技巧與專業態度： $r = .90, p < .01$ ）。「SP 站」的「病史詢問」與紙筆站的「簡答題」之間，呈現中度正相關（ $r = .44, p < .05$ ）。這些研究結果證實採用不同評量方式（SP 站 vs. 紙筆站）測量相同面向衡鑑能力時，「初次晤談 OSCE」具有良好的聚斂效度（convergent validity）。

表五 「初次晤談 OSCE」多特質—多方法矩陣（ $N = 27$ ）

方法	特質	SP站：實作評量		紙筆站：紙筆測驗	
		HT	CS+PS	MC	SA
SP站： 實作評量	病史詢問（HT） 溝通技巧與專業態度 （CS+PS）				
紙筆站： 紙筆測驗	選擇題（MC） 簡答題（SA）	.71**		.14	.03
		.44*	.32	.18	

註1：粗體數字為相同特質—不同方法，斜體數字為不同特質—相同方法，其他數字為不同特質—不同方法。

註2：\* $p < .05$ , two-tailed; \*\* $p < .01$ , two-tailed.

紙筆站的「選擇題」與 SP 站兩分測驗的測驗分數之間，皆呈現低度正相關（病史詢問： $r = .14, p = .50$ ；溝通技巧與專業態度： $r = .03, p = .89$ ）；其與「簡答題」之間，也呈現低度正相關（ $r = .18, p = .37$ ）。這些研究結果顯示「初次晤談 OSCE」不論使用相同方式評量不同構念，或是使用不同方式評量不同構念時，皆具有良好的區辨效度（discriminant validity）。

整體來說，「初次晤談 OSCE」的 SP 站與紙筆站分別以不同測驗型式（實作評量 vs. 紙筆測驗）評估受試者在「心理衡鑑初次晤談」的專業知識、技術與態度。研究結果顯示此測驗用於相同構念測量時，具有聚斂效度；用於相異構念測量時，具有區辨效度，證實

「初次晤談 OSCE」具有良好的構念效度。

### 五、「初次晤談OSCE」之效標關聯效度

在效標關聯效度方面，本研究採同時效度（concurrent validity）以「年級」與「選修衡鑑課程數」檢驗受試者在「初次晤談 OSCE」測驗分數的影響。囿於研究樣本數小於 30 人，本研究使用無母數統計之 Mann-Whitney 與 Kruskal-Wallis 檢定加以分析，其分析結果與各組之平均數與標準差皆列於表六。以年級來說，碩二學生在「紙筆站簡答題」的測驗分數明顯高於碩一學生（ $U = 14.50, p < .01$ ），他們在「紙筆站整體表現」也明顯優於碩一學生（ $U = 10.50, p < .01$ ）。在「選修衡鑑課程數」部份，選修「2 門」或「 $\geq 3$  門」者在紙筆站的平均分數明顯高於選修「1 門」者之平均分數。因此，以年級和衡鑑修課數為效標時，研究結果證實「初次晤談 OSCE」的「紙筆站」具有良好的效標關聯效度。然而在「SP 站整體表現」上，不同年級和不同衡鑑選修課程數的碩士生之間並未呈現顯著差異。可能因為本研究的受試者為碩一、碩二學生，大多缺乏面對個案進行臨床晤談的經驗，因此他們在「初次晤談 OSCE」之「SP 站」的表現並無顯著差異。此研究結果也說明，實際操作經驗在認知能力轉換為行為能力過程中的必要性。

表六 不同年級與選修衡鑑課程數之受試者在「初次晤談OSCE」測驗分數的平均數、標準差與Mann-Whitney和Kruskal-Wallis統計量

各站／各面向之測驗分數	年級				Mann-Whitney U值
	碩一		碩二		
	Mean	SD	Mean	SD	
面向一：病史詢問	33.90	6.47	34.67	5.09	55.50
面向二：溝通技巧與專業態度	23.42	4.65	24.43	4.61	60.00
「SP站」測驗總分	57.68	10.21	60.00	9.14	49.00
分測驗一：選擇題	21.30	2.64	23.00	3.16	49.00
分測驗二：簡答題	37.60	6.72	46.14	2.78	14.50**
「紙筆站」總分	58.90	7.50	69.14	3.20	10.50**
各站／各面向之測驗分數	選修衡鑑相關課程數			Kruskal-Wallis $\chi^2$ 值	
	1門	2門	$\geq 3$ 門		
	Mean (SD)	Mean (SD)	Mean (SD)		
面向一：病史詢問	33.00 (6.63)	34.70 (6.72)	34.38 (5.34)	0.34	
面向二：溝通技巧與專業態度	23.14 (5.70)	23.36 (4.54)	24.63 (3.96)	0.08	
「SP站」測驗總分	57.00 (11.85)	58.50 (10.63)	59.00 (7.96)	0.17	
分測驗一：選擇題	21.50 (2.20)	22.00 (3.03)	21.63 (3.38)	0.58	
分測驗二：簡答題	33.75 (7.72)	42.86 (4.87)	41.69 (5.41)	6.26*	
「紙筆站」總分	55.25 (8.52)	64.87 (6.62)	63.31 (6.15)	5.93*	

註：\* $p < .05$ , two-tailed; \*\* $p < .01$ , two-tailed.

## 肆、討論與建議

目前，國內正積極推動醫事人員專業能力評量的改革，不論醫師或專科護理師，皆試圖透過 OSCE 對於醫事人員進行綜合性臨床專業能力評估（張文蕙、謝素英、許麗齡，2014；蔡詩力、陳震寰、方基存、蔡景仁、張上淳，2008）。本研究期望透過「初次晤談 OSCE」的測驗發展與信效度初步檢驗，評估 OSCE 應用於「心理衡鑑初次晤談」衡鑑能力評量的可行性。本研究屬於初探性質，目前國內尚未提出心理學 OSCE 實徵研究之際，本研究的研究結果尚無法與國內其他心理學 OSCE 實徵研究相互比較。為了避免研究結果的過度推論，本文僅依據研究結果提出下列討論與建議。

### 一、「初次晤談 OSCE」的測驗品質

在「初次晤談 OSCE」正式施測前，本研究透過專家評估與預試結果對於 SP 站與紙筆站進行試題內容與題項修訂。在 SP 站部份，由於評分表內容是 SP 站測驗品質的重要關鍵，故藍玉玲（2012）透過「專家意見調查」與「專家共識會議」，完成「憂鬱個案評分表」的題項修訂並建立一份具專家共識之 SP 站評分表。本研究也透過試題分析再次證實此評分表的測驗品質。

在紙筆站方面，本研究分成「選擇題」與「簡答題」兩部份加以評估。前者透過六位心理測驗／衡鑑專家評估試題的重要性、難易度與試題描述品質挑選出適用於臨床心理碩士生的題項，作為「初次晤談 OSCE」紙筆站的選擇題分測驗。依據本研究試題分析結果，此分測驗的平均難度為 0.72，代表此份試題的難度較為簡單，與原訂之試題難度（難易適中）略有不同。然而，誠如余民寧（2011，頁 278）所言：「尤其是當拿來作為指標分析用的樣本數太少時，所獲得的試題分析指標更是屬於暫時性的，它們隨時都會改變。因此，如同效標參照測驗的試題分析一樣，試題只要確立它是否能夠測量到所欲的教學目標即可，無須太過看重它的試題難度和鑑別度的高低」。在本研究樣本數較小的情況下，或許可以不必過度強調難度略簡單的問題，而應重視此份試題是否測量到研究者預期的評量內容。就專家評估結果而言，此測驗適用於評估碩士生是否具有心理測驗與心理衡鑑的基礎知識。

有關紙筆站簡答題部份，本研究透過預試進行調整修訂，並透過正式施測的研究結果評估試題品質。研究結果發現或許因為簡答題各題的評估內容（見表二）並非單面向，包含鑑別診斷、測驗工具分數解釋、個案概念化等向度，進而導致各題與簡答題分測驗之間呈現低度相關與內部一致性信度偏低（ $\alpha = .25$ ）的現象。雖然，表六的研究結果顯示簡

答題分測驗具有效標關聯效度。但是，就整體而言，未來在 OSCE 紙筆站簡答題設計上，仍需仔細思考如何在信度與效度之間，取得一定的平衡。在信度方面，或許可從增加更多測量相同概念的題目著手，而效度方面除了評量多個不同面向的重要概念之外，還必須注意題數與測驗時間的掌握，以設計一份兼具信效度之 OSCE 簡答題分測驗。

## 二、「初次晤談OSCE」的信效度

在內部一致性信度方面，不論是 SP 站評分表的整體  $\alpha$  係數 ( $\alpha = .89$ )，或是兩個分測驗之  $\alpha$  係數（病史詢問： $\alpha = .80$ ；溝通技巧與專業態度： $\alpha = .88$ ）皆達到 DeVellis（1991）建議的新測驗信度之下限，代表 SP 站評分表具有良好的內部一致性信度。雖然，紙筆站簡答題的內部一致性信度偏低 ( $\alpha = .25$ )，但是，如表二所述簡答題各題並非測量相同的概念，故內部一致性信度不高也是必然的結果。未來在題項設計上，需取得信度與效度之間的平衡。在評分者信度方面，兩位評分者在紙筆站簡答題的評分結果呈現高度正相關 ( $r = .75, p < .01$ )，代表具有良好的評分者信度，唯評分寬嚴不一仍有待改善。SP 站方面，當排除評分者 3 的評分結果後，其他三位評分者對於受試學生在 SP 站的評分結果並未呈現顯著差異 ( $\chi^2(2) = 2.71, p = .26$ )，這代表本研究的 SP 站評分者訓練具有一定的效果。造成評分者 3 可能評分過低之因，將於「初次晤談 OSCE 的評分者效應」章節說明。

在效度方面，「初次晤談 OSCE」依據 Krishnamurthy 等人（2004）的八大心理衡鑑能力評估面向與藍玉玲（2010）提出的 20 項最重要的心理衡鑑專業能力，選取十個最基礎的評量向度並完成本測驗的雙向細目表（見表二），故「初次晤談 OSCE」的評量內容應具有一定的「內容效度」。關於「構念效度」方面，本研究透過 MTMM 驗證「初次晤談 OSCE」的構念效度。研究結果顯示 SP 站與紙筆站對於相同構念的評量具有聚斂效度，對於不同構念的評量則呈現區辨效度。此研究結果也和 Newble 與 Swanson（1988）的研究結果相似，SP 評量方式與簡答題評量方式之間呈現中度相關，但 SP 評量方式與選擇題評量方式之間則呈現低度相關。誠如 Newble 與 Swanson 所言，SP 評量方式與其他評量方式的組合有助於 OSCE 信效度的提昇，也可更有效地掌控測驗時間與測驗所需投入的人力物力。再者，SP 站的「病史詢問」與紙筆站的「簡答題」之間，呈現中度正相關，也反映出兩測驗間的關連性，證實本研究將這兩個測驗設計為連貫站的目的已達到。

在效標關聯效度方面，本研究證實「初次晤談 OSCE」的「紙筆站」已達到一定的效標關聯效度，適用於評估身處於「碩士教育」專業發展階段心理師之衡鑑專業能力。與「紙筆站」相比，「SP 站」則未呈現相似的結果。或許對於碩一、碩二學生而言，學校衡鑑教育以衡鑑專業知識與衡鑑技巧的教授為主，較少直接面對／接觸個案、與個案進行臨床晤

談的實作訓練。因此，對於尚未進入全職實習的碩二學生而言，他們或許比碩一學生擁有更多的專業知識與技巧，但是在實作方面，可能因為實作經驗不足或 SP 站評量時間較短（20 分鐘），而無法立即有效地呈現已習得的衡鑑專業知識與技巧，進而使他們在「SP 站」的表現並未明顯地優於碩一學生。未來可透過擴大施測對象至碩三生與新進心理師，進一步評估「初次晤談 OSCE」的「SP 站」之效標關聯效度。

### 三、「初次晤談 OSCE」的評分者效應

本研究發現評分者 3 在「SP 站」的平均評分（ $M = 46.75$ ）明顯地低於其他三位（評分者 1： $M = 64.56$ ，評分者 2： $M = 55.71$ ，評分者 4： $M = 59.60$ ）。造成此現象的可能原因，除了評分者 3 採用比其他三位嚴格的評分標準之外，也可能在本研究隨機分派不同評分者評估不同學生時，評分者 3 所評估的學生剛好皆屬於能力較差者，進而導致評分者 3 的平均評分明顯低於其他三位評分者的平均評分。未來在「SP 站」的評分方面，可採用 Newble 與 Swanson（1988）的建議，每個「SP 站」皆設置兩個評分者同時評分，以避免評分標準不一所產生的影響。除此之外，也可在每個「SP 站」皆設置錄影設備，當兩個評分者的評分者信度過低時，將可由第三位評分者以錄影帶評分，來提供較合理的評分結果。未來從事心理學 OSCE 研究時，可對於學生與 SP 互動的可能行為表現提出更完整的評分對照表，再依此表來訓練評分者以降低評分者效應，進而提高評分者信度與評分結果的正確性。

### 四、研究限制與未來方向

囿於本研究的研究樣本較小，僅能對於「初次晤談 OSCE」提出初步的信效度驗證結果，此為本研究的研究限制之一。未來在人力與研究經費許可下，應招募更多受試者以提供更完整的信效度評估。再者，本研究僅包含「尚未全職實習且並無全職臨床實務工作經驗」的碩一、碩二學生，這些學生大多缺乏面對個案進行臨床晤談的經驗，較難評估「初次晤談 OSCE」在「SP 站」的效標關聯效度，此為本研究的另一項研究限制。未來對於「SP 站」效標關聯效度的檢驗，可參考郭生玉（1994）的建議，採用對照團體法（contrasted groups）進行效標關聯效度的驗證。除了招募碩一、碩二學生之外，同時招募更有臨床晤談實務經驗的群體（如：已完成全職實習的碩三學生與執業心理師）進行施測，以比較不同臨床晤談能力群體在「SP 站」的表現是否具有顯著差異來檢驗「SP 站」的效標關聯效度。除此之外，「SP 站」效標關聯效度的驗證也可採用重覆施測方式，招募碩三學生在全職實習前後各接受一次「初次晤談 OSCE」，以了解全職實習對於學生晤談能力的影響性，

進而驗證「SP 站」的效標關聯效度。

整體來說，本研究雖然屬於初探性質，但研究結果似乎反映出「初次晤談 OSCE」用於「心理衡鑑初次晤談」衡鑑能力評量的可行性。基於 OSCE 評量模式在模擬臨床實作情境的限制，本研究僅針對衡鑑目的為「初次晤談」且心理病症為「憂鬱症」的情況設計模擬情境，進行「心理衡鑑初次晤談 OSCE」的測驗發展，並提出此測驗的信效度初探結果。對於 OSCE 應用在其他衡鑑目的或是其他心理病症之可行性，仍有待更多專家學者的參與或更多實徵研究的投入，進行各種 OSCE 的測驗發展與實徵研究，形成更完整的驗證與論述。

### 參考文獻

- 王國川、翁千惠譯（2005）：質性資料分析－如何透視質性資料。台北：五南。Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage.
- 余民寧（2011）：教育測驗與評量：成就測驗與教學評量（第三版）。台北：心理。[Yu, M. N. (2011). *Educational testing and assessment: Achievement test and teaching evaluation* (3rd ed.). Taipei, Taiwan: Psychological Publishing Co.]
- 吳明隆、涂金堂（2005）：SPSS 與統計應用分析（第二版）。台北：五南。[Wu, M. L., & Tu, C. T. (2005). *SPSS & The application and analysis of statistics* (2nd ed.). Taipei, Taiwan: Wu-Nan.]
- 高聖博（2008）：慈濟大學標準化病人實務工作坊手冊。花蓮：花蓮慈濟醫院標準化病人中心。[Kao, S. P. (2008). *Handbook for standardized patient trainer workshop*. Hualien, Taiwan: Standardized Patient Center, Buddhist Tzu Chi General Hospital.]
- 張文蕙、謝素英、許麗齡（2014）：運用客觀結構式臨床測驗以評量出院準備護理人員溝通能力之研究。護理雜誌，61，45-57。DOI: 10.6224/JN.61.3.45 [Chang, W. H., Hsieh, S. I., & Hsu, L. L. (2014). The use of OSCE to evaluate the competency of discharge nurses. *The Journal of Nursing*, 61, 45-57. doi: 10.6224/JN.61.3.45]
- 許文耀譯（1997）：心理衡鑑。台北：心理。Maloney, M. P., & Ward, M. P. (1976). *Psychological assessment: A conceptual approach*. Oxford, New York: Oxford University Press.
- 郭生玉（1994）：心理與教育測驗（第八版）。台北：精華。[Kuo, S. Y. (1994). *Psychological and educational testing* (8th ed.). Taipei, Taiwan: Gin-Hua.]

- 陳心怡譯 (2000)：貝克憂鬱量表第二版中文版指導手冊。台北：中國行為科學社。Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory-2nd edition (BDI-II)*. San Antonio, TX: Psychological Corporation.
- 蔡詩力、陳震寰、方基存、蔡景仁、張上淳 (2008)：客觀結構式臨床測驗 (OSCE) 執行指引。醫學教育, 12, 118-132。[Tsai, S. L., Chen, C. H., Fang, J. T., Tsai, J. J., & Chang, S. C. (2008). A guideline for the implementation of objective structured clinical examination. *Journal of Medical Education, 12*, 118-132.]
- 藍玉玲 (2010)：心理衡鑑能力客觀結構式臨床測驗的發展與前導研究。行政院國家科學委員會專題研究成果報告 (編號：NSC98-2410-H-259-016)。[Lan, Y. L. (2010). *Development of objective structured clinical examination in psychological assessment*. National Science Council in Taiwan (No. NSC98-2410-H-259-016).]
- 藍玉玲 (2012)：「評分表內容」與「評分者效應」對於心理衡鑑 OSCE 測驗品質的影響—以「初次晤談」為例。行政院國家科學委員會專題研究成果報告 (編號：NSC100-2410-H-259-024)。[Lan, Y. L. (2012). *The impacts of OSCE items and rater effects on the psychometric properties of psychology OSCE: Using the intake interview OSCE as an example*. National Science Council in Taiwan (No. NSC100-2410-H-259-024).]
- 藍玉玲、劉彥君、高聖博 (2011)：另一種心理師專業能力的評量模式：客觀結構式臨床測驗模式在心理學的應用。中華心理衛生學刊, 24, 209-245。[Lan, Y. L., Liu, Y. C., & Kao, S. P. (2011). The competency-based assessment model: An OSCE application in psychology. *Formosa Journal of Mental Health, 24*, 209-245.]
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105. doi: 10.1037/h0046.016
- Collins, F. L., Kaslow, N. J., & Illfelder-Kaye, J. (2004). Introduction to the special issue. *Journal of Clinical Psychology, 60*, 695-697. doi: 10.1002/jclp.20009
- DeVellis, R. F. (1991). *Scale development: Theory and applications*. Beverly Hills, CA: Sage.
- Fang, L., Bogo, M., Mishna, F., Murphy, L., Gibson, M. F., Griffiths, V., & Regehr, G. (2013). Development and initial evaluation of the cyber-counseling objective structured clinical examination. *Research on Social Work Practice, 23*, 81-94. doi: 10.1177/1049731512459966
- Fouad, N. A., Grus, C. L., Hatcher, R. L., Kaslow, N. J., Hutchings, P. S., Madson, M. B.,...Crossman, R. E. (2009). Competency benchmarks: A model for understanding and measuring competence in professional psychology across training level. *Training and Education in Professional Psychology, 3*,

S5-S26. doi: 10.1037/a0015832

- Gravetter, F. J., & Wallnau, L. B. (2012). *Statistics for the behavioral sciences* (9th ed.). Belmont, CA: Wadsworth Publishing Company.
- Hodges, B. D. (2006). The objective structured clinical examination: Three decades of development. *Journal of Veterinary Medical Education*, 33, 571-577. doi: 10.3138/jvme.33.4.571
- Kaslow, N. J. (2004). Competencies in professional psychology. *American Psychologist*, 59, 774-781. doi: 10.1037/0003-066X.59.8.774
- Kring, A. M., Davison, G. C., Neale, J. M., & Johnson, S. L. (2007). *Abnormal psychology* (10th ed.). New York, NY: John Wiley & Sons, Inc.
- Krishnamurthy, R., VandeCreek, L., Kaslow, N. J., Tazeau, Y. N., Miville, N. L., Kerns, R.,...Benton, S. A. (2004). Achieving competency in psychological assessment: Directions for education and training. *Journal of Clinical Psychology*, 60, 725-739. doi: 10.1002/jclp.20010
- Lan, Y. L., & Chang, Y. (2014). *Development and initial psychometrics of the psychological assessment competency scale*. Manuscript submitted for publication.
- Newble, D. J., & Swanson, D. B. (1988). Psychometric characteristics of the objective structured clinical evaluation. *Medical Education*, 22, 325-334. doi: 10.1111/j.1365-2923.1988.tb00761.x
- Rodolfa, E., Bent, R., Eisman, E., Nelson, P., Rehm, L., & Ritchie, P. (2005). A cube model for competency development: Implications for psychology educators and regulators. *Professional Psychology: Research and Practice*, 36, 347-354. doi: 10.1037/0735-7028.36.4.347
- Rubin, N. J., Bebeau, M., Leigh, I. W., Lichtenberg, J. W., Nelson, P. D., Portnoy, S.,...Kaslow, N. J. (2007). The competency movement within psychology: An historical perspective. *Professional Psychology: Research and Practice*, 38, 452-462. doi: 10.1037/0735-7028.38.5.452
- Rushforth, H. E. (2007). Objective structured clinical examination (OSCE): Review of literature and implications for nursing education. *Nurse Education Today*, 27, 481-490. doi: 10.1016/j.nedt.2006.08009
- Walsh, M., Bailey, P. H., & Koren, J. (2009). Objective structured clinical evaluation of clinical competence: An integrative review. *Journal of Advanced Nursing*, 65(8), 1584-1595. doi: 10.1111/j.1365-2648.2009.05054.x

收 件 日 期：103年09月22日

複 審 二 日 期：104年05月11日

複 審 一 日 期：104年01月27日

通 過 日 期：104年06月14日

## Development and Initial Validation of the Intake Objective Structured Clinical Examination

Yu-Ling Lan

National Dong Hwa  
University

Yu-Hsuan Chang

Yuan Ze University

Wan-Lan Chen

Tzu Chi University

Influenced by the competency-based movement, psychologists have begun to reconsider what has been missing in the education and evaluation of psychologists, and advocate the importance of competency-based evaluation in professional psychology. In the last four decades, Objective Structured Clinical Examination (OSCE) has been widely adopted in health-related disciplines to measure clinician's knowledge, skills and attitudes. The purpose of this study was to develop and validate the intake OSCE, a competency-based evaluation for measuring master students' competencies in psychological assessment when conducting an intake interview. The intake OSCE included a standardized patient test (SP test), a follow-up short answer test (S test), and a multiple-choice test (M test). In this study, 27 graduate students in training were recruited to examine the psychometric properties of this test. Results of this study indicated that the Cronbach's alpha values of the total and subscale scores of the SP test were all above .80; and the inter-rater reliability of the S test was .75. The moderate positive correlation between the history taking subscale scores of the SP test and the S test total scores suggested the convergent validity of the intake OSCE. The weak positive association between the M test total scores and the SP subscale scores also indicated the divergent validity of this test. As a newly developed test, intake OSCE demonstrates good reliability and validity. Since the present study only provided initial validation of this test, more research needs to be done to provide further psychometric evidence for the intake OSCE.

**Keywords:** objective structured clinical examination, performance assessment, psychological assessment competency