

# 建構及檢測護理客觀結構式臨床技能測驗考題之 信效度、鑑別度及難易度

黃湘萍<sup>1</sup> 趙莉芬<sup>2</sup> 王瑜欣<sup>2</sup> 劉英妹<sup>3</sup> 倪麗芬<sup>1</sup> 簡淑慧<sup>4\*</sup>

## 摘要

- 背景** 客觀結構式臨床技能測驗 (objective structured clinical examination, OSCE) 是目前評估護理學生專業臨床能力的方式之一，其標準化流程之考題發展與考題信效度檢測值得關注。
- 目的** 本研究旨在，依標準化流程建構護理 OSCE 考站；分析 OSCE 考站專家效度及預測效度、評分者間一致性、內在一致性、考題相關性、鑑別度與難易度。
- 方法** 以描述性、橫斷式研究設計，自 2015 年 9 月至 2016 年 10 月依測驗藍圖發展 8 站式 OSCE 考試，應用於北部某科技大學 207 位四技四年級臨床選習前的護理系學生，並且測試 OSCE 考題的信效度、鑑別度與難易度。
- 結果** 本研究共建構 8 站 OSCE 考題，各題的專家效度指標為 .85–1.0，評分者間一致性為 .80–1.0，Cronbach's alpha 為 .53–.82，考題相關性為 .36–.63，鑑別度多為優良或非常優良，難易度多為容易到適中，且 OSCE 成績對臨床選習成績具良好的預測效度。
- 結論／實務應用** 透過標準化的考題建置流程，發展的 8 站式護理 OSCE 考題具有良好的信效度、難易度適中及良好的鑑別度，並可做為檢測學生最後一哩專業技能的方法，期許本研究的結果提供護理教育者於發展 OSCE 評量之參考。

**關鍵詞：** 客觀結構式臨床技能測驗、信度、效度、鑑別度、難易度。

## 前言

客觀結構式臨床測驗 (objective structured clinical examination, OSCE) 起源於 1975 年由英國學者 Harden 及其同儕提出，用於評估醫學系學生的臨床能力 (Rushforth, 2007)。1990 年代，美、加、澳洲等西方國家也逐漸普遍運用，2000 年代亞洲日、韓等各國也逐漸用於醫學生考試 (范, 2015)。在臺灣，2009 年考選部宣布將 OSCE 正式納入醫師國家考試，2011 年試辦、2012 年正式執行，牙醫師、中醫師、藥師等專業學門也考慮將 OSCE 納入國家考試應考資格 (范, 2015)。Miller (1990) 提出「臨床能力評量架構」以評量臨床能力 (competence) 的知識、技能與態度，並強

調臨床能力的展現應包含學習者應該知道 (knows)、瞭解如何做 (knows how)、呈現如何做 (shows how) 以及實際去做 (does)，而傳統的紙筆測驗是評估前兩個層次的能力，較忽略問題解決溝通與批判思考或同理心與關懷的護理專業素養 (Hodges, Turnbull, Cohen, Bi- enenstock, & Norman, 1996)。Miller 認為臨床技能測驗的設計，應該思考如何接近學習者未來將面對的專業情境，讓學習者展現如何做的能力。OSCE 藉由客觀的結構式方法達到一致性公平的模式，並透過標準化病人模擬臨床情境，以展現學習者其臨床能力，對醫護教育的評量方式或國家專業證照考試有莫大的變革 (李、張、顏、王, 2011; 盧、王, 2014; Brighton, Mackay, Brown, Jans, & Antoniou, 2017)。

接受刊載：106 年 10 月 13 日

doi:10.6224/JN.000084

<sup>1</sup>長庚科技大學護理系助理教授 <sup>2</sup>長庚科技大學護理系副教授 <sup>3</sup>長庚科技大學護理系副教授兼林口長庚紀念醫院兒童血液腫瘤科副研究員 <sup>4</sup>長庚科技大學護理系(所)副教授兼林口長庚紀念醫院血液腫瘤科副研究員

\* 通訊作者地址：簡淑慧 33303 桃園市龜山區文化一路 261 號 電話：(03) 2118999-3136；E-mail：swjane@mail.cgust.edu.tw

台灣自2006年起，專科護理師(nurse practitioner)將OSCE應用於證照考試，其包含：病史詢問、身體檢查、鑑別診斷健康問題，以評估專科護理師的護病溝通、專業態度、批判性思考等臨床能力(童、王，2009)。國外許多護理學校也將OSCE作為評量學生專業能力的方法且師生都認為OSCE是一項有效且可信的評量護理學生臨床能力的方法(Mitchell et al., 2014; Rushforth, 2007; Selim, Ramadan, El-Gueneidy, & Gaafer, 2012)。國內自2010年起，有護理學校率先建置OSCE考場、發展教案並運用OSCE於學生的形成性評量(formative evaluation)或總結性評量(summative evaluation; Fan, Chao, & Jane, 2014; Fan, Wang, Chao, Jane, & Hsu, 2015)。近年來，在教育部經費挹注，多所護理技職學校陸續建置OSCE考場並開始應用於學生最後一哩實習前的能力檢定(王等，2015; 楊、李、楊、陳，2016)。雖然OSCE可以檢測複雜的臨床能力，然而對教案撰寫的嚴謹度及OSCE測驗的品質控制之信效度檢定之探討仍較缺乏(Fuller, Pell, Homer, & Roberts, 2012)，其可能會增加評量結果之誤差。同樣，目前醫學教育OSCE其內部一致性信度差異仍頗大，介於.48(Wessel, Williams, Finch, & Gémus, 2003)或.87-.95(Sturpe, Huynh, & Haines, 2010)。有鑑於目前仍缺乏本土性護理OSCE考題信效度之研究，故本研究旨在探討：(一)依標準化流程建構護理OSCE考題；(二)分析OSCE考題的信度、效度、鑑別度及難易度。

## 文獻查證

本文文獻探討的重點，包括：OSCE信效度、鑑別度以及難易度檢測，因受限於篇幅，效度包括：專家內容效度與預測效度；信度部分包括：評分者間一致性、考題內在一致性及考題相關性。

### (一)OSCE常見之效度檢測方式

#### 1.內容效度(content validity)

由專家判斷評分項目是否能測量到期望測量的構面，如此又稱為專家效度(expert validity)。通常請專家針對各項評分內容的重要性、語意簡明性與適當性進行判定，給分通常使用四分法，範圍為1~4分，1分表示完全不正確或完全不適合，2分表示不正確或不適合，有列入的價值但需大幅修正，3分表示正確或適合，有列入的價值但需做文辭修改，4分表示非常正確或非常適合。依據每位專家給分3分以上的題數除以總題數，得到的值(商)即為內容效度指標

(content validity index, CVI)，CVI值.78以上時表示內容效度良好(Polit, Beck, & Owen, 2007)。

#### 2.預測效度(predictive validity)

是效標關聯效度(criterion-related validity)之一，指測驗分數與未來的效標資料之相關性。常用的效標資料包括：測驗的成績或實際工作的成果等。此種效度之檢測是運用追蹤的方法，對受試者將來的行為表現，作長期觀察與追蹤再與當初的測驗分數做相關分析，以衡量測驗結果對未來成就的預測效應(簡、余、劉，2015)。

### (二)OSCE常見之信度檢測方式

#### 1.內在一致性

主要是檢測同一組評分項目是否確實測量同一特質，常用的測量的方法如：庫李法(KR-20)、折半信度或Cronbach's alpha，庫李法用於二分法的評量工具，非二分法工具的信度檢測中以Cronbach's alpha為目前最常用的方法，不須折半計算而可直接分析其係數值。一般認為高階-OSCE(例如，專業執照考試)Cronbach's alpha最好大於.90，決定通過與否之考試則大於.80，中階OSCE(例如，課程結束後之測驗).80-.89，基礎型的OSCE(例如，單元小考、測驗)則在.70-.79(Roberts, Newble, Jolly, Reed, & Hampton, 2006)。

#### 2.評分者間一致性

當OSCE應用於大量的考生時，同一考題可能有多位考官，在正式測驗前，評分者間信度更顯重要。在單一題目評分信度方面，組內相關係數(intra-class correlation coefficient, ICC)通常用來檢測二到多位評分者間的一致性(inter-rater reliability)，ICC係數介於.6-.9表示評分者間信度良好(Swift, Spake, & Gajewski, 2013)。當評核結果是等級或次序型的評分資料時，可以選擇Kendall和諧係數(Kendall's coefficient of concordance)來評估各評分者評分的相關性(李，2012; 簡等，2015)。

#### 3.考題相關性

為了瞭解OSCE整體測驗的內在相關性，可以將每一站學生的得分與所有站別的總分進行考題相關性分析(item-to-total correlation)。當該站與總分之相關係數低於.3，則表示該站與其他站別的相關性低，也就代表在整體測驗為了能區分考生程度的情況下，該站應該刪除(簡等，2015)。針對偏低的考題，宜進行檢視是否該題與其他考題及整體測驗的本質不同。

### (三)OSCE考題之鑑別度

鑑別度(discrimination)與難易度也是OSCE測驗

中需要關注的問題，其中鑑別度也屬於考題建構效度的一部分(簡等, 2015)。鑑別度主要是比較高分群與低分群成績的差異，高分群及格率越高而且低分群及格率越低則該考題的鑑別度高，意即兩組及格率差異越大表示鑑別度越佳。Kelley (1939)建議在常態分配下最適當的高低分群比率是各佔27%，因此將各題所有考生的分數排序後，各取27%之高分群及低分群計算及格率，計算鑑別度的步驟如下：(1)將考生依原始分數的高低排列，(2)從最高分往下，取總人數的27%為高分組，(3)從最低分往上，取總人數的27%為低分組，(4)計算高分組與低分組及格人數與及格率，(5)依公式計算每題的鑑別指數(李, 2012)。

鑑別度計算公式為 $D$ (鑑別度指數) $=P_H$ (高分組該題之及格率) $-P_L$ (低分組及格率)。鑑別度指數為兩組及格率的差異，越高表示該題鑑別度越高，指數大於.4表示試題「非常優良」；.3-.39表示試題「優良」(可能需要修改)；.2-.29表示試題「尚可」(通常需要修改)；.19以下表示試題「鑑別度低」(淘汰或修改；余, 2011)。

#### (四)OSCE考題之難易度

難易度(difficulty)仍然是計算高分群與低分群的及格率進行估計。依照Kelley建議高低分群各佔27%，計算難易度的步驟如下：(1)將考生依原始分數的高低排列，(2)從最高分往下，取總人數的27%為高分組，(3)從最低分往上，取總人數的27%為低分組，(4)計算高分組與低分組及格人數與及格率，(5)依公式計算每題的難易度指數。依據難易度係數計算公式 $P$ (難易度係數) $= [P_H$ (高分組該題之及格率) $+P_L$ (低分組及格率)] $/2$ ，以估計各題的難易度係數。難易度指數越小表示該考題越難，指數小於.25則表示該題「極難」；介於.25-.39表示「難」；介於.4-.7表示「適中」；介於.71-.9表示「易」；指數大於.9則表示「極易」(余, 2011)。

## 方 法

### 一、研究設計與對象

本研究是一描述性、橫斷式研究設計，探討北部某科技大學護理系即將進行臨床選習的四技四年級學生，接受8站OSCE考試後檢測各考題之信效度、鑑別度及難易度。研究期間自2015年9月-2016年10月。測驗完成後將施測對象個人資料去連結再進行分

析，測驗成績並不影響學生學習權益，本研究計畫經長庚紀念醫院倫理委員會評定免審查案件(案號：103-5195B)。

### 二、研究步驟

本研究由課程組與臨床技能中心主管，以本系所發展之十大核心架構為主軸，並參考醫學(包括，病史獲得、理學檢查、判讀技巧、醫療處置技巧、溝通技巧、態度等；Byrne, 2007/2008; Dornan & O'Neill, 2006/2008)及護理(包括，病史獲得、理學檢查、判讀技巧、衛教或健康促進、技術如無菌技術及溝通技巧等；Khattab & Rawlings, 2008; Lauder et al., 2008; Rentschler, Eaton, Cappiello, McNally, & McWilliam, 2007)OSCE文獻，建立本研究OSCE測驗藍圖(blue print)，期許能透過OSCE檢定以強化護理學生技能、溝通及批判性思考能力。再依據高(2013)提出教案發展六步驟(1.界定教案的目標；2.確定教案角色的特點；3.設定臨床場景與情境；4.發掘與建立臨床病例資訊；5.決定須回饋的方式、項目與來源；6.評分表的制定原則)及林、黃、蔡、余(2015)提出護理OSCE教案撰寫的原則，召集已完成8小時考官培訓並曾撰寫OSCE考題的教師參與本項計畫，說明研究目的，經同意後進入本研究工作小組，透過工作會議凝聚共識，擬定OSCE教案的主題及面向(表一)，後進行專家內容效度、及格標準設定及評分者一致性。於2015年9至11月共發展8站OSCE教案考題，完成後用於2016年1月21-22日四技選習前OSCE考試。

每站考試時間為看題1分鐘，操作8分鐘，換站1分鐘，每位學生共計10分鐘/站 $\times$ 8站=80分鐘。專家效度：種子教師著手撰寫教案，完成初稿後先進行成員同儕審閱(peer review)，各題撰寫教師依意見修改後，由兩位與考題領域相符且具備OSCE考官資格的臨床專家進行專家效度審查，估計專家效度指標(CVI)，並依專家意見修改形成考題的確定內容。

#### (一)及格標準

考題完成後，以Angoff法建立各考題及格標準，此法是目前常用的及格標準設定方法(范、蔡、賀, 2015)。邀集所有考題撰寫老師擔任及格標準的設定人員，會議中逐題進行設定，先讓撰寫老師說明該考題測驗目的與面向，再讓設定人員估計受試者可接受的最低能力表現並設定為該題及格分數，將所有設定人員的設定分數輸入，然後估算所有設定者評分

數的平均值與標準差。結果每一題皆經過設定人員三次的設定後，符合標準差小於2的設定原則，各試題及格標準如表一。

### (二) 評分者間一致性

各考題由負責老師找兩位學生試考，並拍攝「優」與「不良」兩段技術試考影帶，做為建立考官一致性評分之用。考官一致性會議在OSCE進行2週前舉行，各考題撰寫的老師擔任考題負責人，會議中透過考題解說、試評、討論以達成共識，並計算考官評分一致率。

### 三、資料分析

考試結束後，以SPSS 20.0版執行信度分析，分析Cronbach's alpha值以估計各題的內在一致性，以皮爾遜積差相關(Pearson correlation)分析各題與整體測驗總分的考題相關性(item-to-total correlation)，並排序各題得分前27%的學生為高分組，各題得分後27%的學生為低分組，依公式計算各題的難易度與鑑別度，再以獨立樣本t檢定(independent t test)與皮爾遜積差相關(Pearson correlation)分析OSCE成績與臨床選習成績之差異及相關性，以估計OSCE之預測效度。

## 結 果

本研究依OSCE考題發展標準化流程，發展8站教案考題，並運用於四技(N = 207)臨床選習前OSCE測驗。實施OSCE測驗前先進行專家效度及評分者間一致性之分析，測驗後則進行預測效度、考題內在一致性、考題相關性、難易度及鑑別度分析，各項分析結果詳述如下。

表一

OSCE測驗藍圖與各試題及格標準

面向	測考試題							
	抽痰	單次導尿	靜脈注射帽給藥	一般換藥	肌肉注射	糖尿病衛教	腹痛檢查	周邊淋巴結檢查
護病溝通				V		V		
技術	V	V	V	V	V			
學理應用			V			V		
理學檢查							V	V
病史詢問							V	
倫理								V
總分	34	32	28	30	34	28	34	34
及格分數	20	21	17	18	21	16	18	20

註：OSCE = objective structured clinical examination；V表該題測驗的面向。

### 一、效度

#### (一) 專家內容效度

每一考題教案撰寫完成後，先經由同儕檢視，再由兩位臨床專家審查，評分項目(以腹痛檢查及糖尿病衛教為例)如附件一。結果專家內容效度分析CVI值介於.85至1.00之間，以「一般換藥」最高(CVI = 1.0)，「糖尿病衛教」最低(CVI = .85)，「抽痰」、「單次導尿」、「腹痛檢查」、「周邊淋巴結檢查」四題CVI為0.99，「肌肉注射」為.98，「靜脈注射帽給藥」為.96，整體平均為.97。

#### (二) 預測效度

共計206位學生參加該梯次臨床選習，首先，將學生依OSCE通過與否分為兩組，以獨立樣本t檢定分析兩組學生的選習成績是否有異，結果發現通過OSCE學生(n = 159)之選習成績(平均值84.97，標準差4.01)顯著高於未通過OSCE而需補考者(n = 47；平均值82.96，標準差4.35；t = 2.858, p = .006)。再以皮爾遜積差相關分析OSCE總分與學生選習成績之關聯，發現學生之OSCE總成績與選習成績呈顯著正相關(r = .285, p < .01)，表示OSCE總分越高者之選習成績越高，由此可知，OSCE對於選習成績具有良好的預測效度。

### 二、信度

#### (一) 評分者間一致性

考題一致性會議中說明該題測驗目的與評分要點後，由考官針對「優」與「不良」的兩段試考影帶進行評分，並針對不一致之處澄清以建立共識。結果考題的ICC介於.96至.99之間(p < .001)，平均.98，表

示達到良好的評分者間一致性。由於選習前OSCE的結果是以通過與否區分，因此針對考官們對於通過與否的一致率進行分析，結果各題的考官一致性介於80–100%，平均93%，以「靜脈注射帽給藥」、「腹痛檢查」、「肌肉注射」三題的一致率較低（表二）。

#### （二）內在一致性

以Cronbach's alpha檢測考題的內在一致性，各題的係數介於.53–.82（平均.64；表二）。以「單次導尿」、「一般換藥」與「周邊淋巴結檢查」等三題的內在一致性較高，係數皆大於.7，而「抽痰」、「靜脈注射帽給藥」、「糖尿病衛教」、「腹痛檢查」等四項考題的係數低於.6，其中又以「抽痰」最低，係數為.53。

#### （三）考題相關性

將每一考站的學生得分與總分進行相關性分析，結果各題得分與總分之相關係數介於.36至.63，平均.46（表二）。以「肌肉注射」最低，而「周邊淋巴結檢查」最高，各題都大於.3，並且達到統計顯著性

（ $p < .01$ ），表示此次OSCE測驗中所有考題與總分間具有良好的相關性。

### 三、鑑別度

試題的鑑別度係數介於.08至.65，平均.40，表示整體鑑別度「非常優良」（表三）。其中「靜脈注射帽給藥」、「腹痛檢查」、「周邊淋巴結檢查」三項的鑑別度「非常優良」，「單次導尿」、「一般換藥」與「糖尿病衛教」三項的鑑別度為「優良」，「肌肉注射」鑑別度為「尚可」，「抽痰」鑑別度為「低」。

### 四、難易度

試題的難易度係數介於.40至.96，平均.67，整體難易度為「適中」（表三）。「靜脈注射帽給藥」、「一般換藥」、「肌肉注射」、「腹痛檢查」、「周邊淋巴結檢查」等五項题目的難易度「適中」（介於.4至.7間），「單次導尿」與「糖尿病衛教」兩題為「容易」，而「抽痰」的難易度大於.9，對四技考生而言「極易」。

表二

OSCE考題的信度分析

試題	ICC	評分者間一致率(%)	Cronbach's $\alpha$	item total score
抽痰	.96***	100	.53	.44**
單次導尿	.99***	100	.72	.47**
靜脈注射帽給藥	.97***	80	.57	.42**
一般換藥	.98***	100	.74	.43**
糖尿病衛教	.98***	100	.54	.40**
腹痛檢查	.97***	80	.58	.51**
肌肉注射	.97***	80	.60	.36**
周邊淋巴結檢查	.99***	100	.82	.63**
平均	.98	93	.64	.46

註：OSCE = objective structured clinical examination；ICC = intra-class correlation coefficient。

\*\* $p < .01$ ，\*\*\* $p < .001$ 。

表三

OSCE考題的鑑別度與難易度分析

試題	高分組及格率(%)	低分組及格率(%)	難易度	難易度說明	鑑別度	鑑別度說明
抽痰	100	92	.96	極易	.08	低
單次導尿	100	69	.85	易	.31	優良
靜脈注射帽給藥	87	35	.61	適中	.52	非常優良
一般換藥	88	52	.70	適中	.37	優良
糖尿病衛教	96	63	.80	易	.33	優良
肌肉注射	54	27	.40	適中	.27	尚可
腹痛檢查	83	19	.51	適中	.63	非常優良
周邊淋巴結檢查	88	23	.56	適中	.65	非常優良

註：OSCE = objective structured clinical examination；難易度係數： $< .25$ 「極難」， $.25-.39$ 「難」， $.4-.7$ 「適中」， $.71-.9$ 「易」， $> .9$ 「極易」；鑑別度係數： $< .19$ 「低」， $.2-.29$ 「尚可」， $.3-.39$ 「優良」， $> .4$ 「非常優良」。

## 討 論

### 一、考題及格標準的建立

本次共發展8站考題，每一考題分別針對1-2項能力進行評估，並透過Angoff及格標準的設定方式(李, 2012)建立各考題的及格標準，經由三次的設定確立了各題的及格標準，經過測驗也發現雖然某些題目及格標準低於總分的50-60%，仍在測驗中屬於難易適中且鑑別度良好的考題，由此可知，OSCE及格標準不適合再以傳統總分的60%做為及格標準。本研究透過各科老師建立OSCE全新考題，並透過多項信效度檢測的方法，以達到較嚴謹的專業臨床能力評估。

### 二、考題的效度

本研究的專家效度指標CVI值介於.85-1.0間，文獻建議CVI值.78以上時表示整份檢核量表或該題內容效度良好(Polit et al., 2007)，由此可見本研究各題具良好的內容效度。本研究所有寫題老師均接受完整的8小時考官訓練，多位老師過去已有撰寫經驗，且交由業界專家審閱前均已完成同儕檢視，因此考題有較好的內容效度。

在預測效度方面，四技OSCE總成績與選習成績呈現顯著正相關，且通過OSCE之學生其選習成績顯著高於OSCE未通過者，表示OSCE具良好的預測效度，Bartfay、Rombough、Howse和Leblanc(2004)認為OSCE能夠評估學生達成「show how」的層次，這是學習理論的較高層次，表示OSCE較能評值學生「如何做」的專業技能，當學生具備了「如何做」的能力時，則於選習時能有較好的臨床表現。

### 三、考題的信度

Swift等人(2013)主張ICC係數介於.6-.9是表示評分者間信度良好，本研究各題的ICC介於.95-.99，顯示有良好的評分者間一致性，這項結果可能由於本研究在考官一致性會議時，會先說明各題評分指引，再針對試評影帶給分，安排考官時也盡量按照考官的專長背景安排，國內學者也肯定這樣建立共識的方法(李, 2014)。

綜合文獻查證，小考或一般性OSCE考題Cronbach's alpha值應該在.7以上較理想(Downing, 2004; Roberts et al., 2006)。本研究中「抽痰」、「靜脈注射帽給藥」、「糖尿病衛教」、「腹痛檢查」等四項之係數皆低於.6。這樣的結果與國內護理教育OSCE的數值相去不遠(李, 2014)。Tavakol及Dennick(2011)曾經提到Cronbach's alpha值較低可能是由於題數太

少，或評分項目與項目間相關性低所致，可以刪除與其他評分項目低相關性(接近零)的項目。此外，文獻也提到如果評分項目包含大於一個面向能力的評量，則評量表的內在一致性可能偏低，這些題目中「靜脈注射帽給藥」、「糖尿病衛教」、「腹痛檢查」等三題是評核兩個面向的能力，可能因此而呈現較低的Cronbach's alpha值。「抽痰」僅針對技術面向的評值，Cronbach's alpha值偏低並非評估多項能力所致，Silva、Lunardi、Mendes、Souza和Carvalho(2011)認為技術操作的考站可能有較大的變化導致內在一致性偏低。國內學者李(2014)也發現抽象概念(如無菌技術)內在一致性較低，若為具體的概念且步驟清楚者的內在一致性較高，可能由於抽痰評分表中涵蓋無菌概念、抽痰程序等較多樣的概念，因此內在一致性較低。

### 四、考題的鑑別度

本研究八項考題中「抽痰」的鑑別度低，「肌肉注射」的鑑別度尚可，其餘六題屬於鑑別度優良或非常優良。「抽痰」是屬於難度極易且鑑別度低的考題，也許我們低估了考生執行此項技術的能力，如果將及格分數提高，提高了本題的難度，很可能也提高本題的鑑別度，不過，這項考題一直是學生在臨床較怯於執行的技術，也可能考題設計的評分內容忽略了某些關鍵能力，值得未來繼續探討。此外，「單次導尿」、「糖尿病衛教」雖都屬於難度偏易而鑑別度良好的考題，因此都屬於適當的考題。「靜脈注射帽給藥」、「腹痛檢查」、「周邊淋巴結檢查」等三題之鑑別度非常優良，整體而言，本研究的考題鑑別度堪稱優良。

### 五、考題的難易度

考題的難易度與鑑別度也是影響考題信效度的重要因素，本研究的考題難易度多為適中，「單次導尿」、「糖尿病衛教」的難易度屬於容易，只有「抽痰」的難度屬極易，全數的高分組考生及九成以上低分組的考生「抽痰」這題皆及格，由於這項技術是該年度內外科護理實驗加強練習的技術，所以可能考生加強練習的效應且及格標準訂得過低所致。「單次導尿」、「糖尿病衛教」的難易度屬於容易，高分組考生及格率皆大於90%，而低分組考生及格率則約60-70%，但是考量其他五項屬於難易適中的考題，搭配此兩題屬於容易的考題也不失為良好的組合，倘若未來希望提高難度，則可考慮略為調高這兩題的及格標準。

## 結 論

本研究經由測驗藍圖透過標準化流程建構8站OSCE考題、Angoff及格標準設定法、專家效度與預測效度、考官評分者間及考題內在一致性、考題相關性分析，及難易度與鑑別度分析，確定整體考題具有良好的信效度、難易度與鑑別度。未來仍須針對部分內在一致性較低與鑑別度不佳的考題進行檢討與修訂。本研究只探討OSCE考題與考官信效度，而考試過程的環境動線安排、標準化病人的訓練與其表現等，也都是影響OSCE考試信效度的因素，未能探討這些議題是本文的限制。目前OSCE也逐漸被醫院用來評估新進護理師的臨床技能，因此，建議未來OSCE相關之研究，除了探討測驗的品質與信效度等議題外，可以探討在校OSCE成績對就業後臨床表現及職場OSCE表現之影響及探討OSCE測驗對護理教育核心能力的影響。此外，OSCE需要充分的人力與物力才得以進行，成本的考量使OSCE的普及化受到限制，因此，建議未來的研究可以探討不同OSCE總站數之外在效度的差異。希望藉本研究結果，可提供各臨床機構或護理學校建構OSCE考題信效度、難易度與鑑別度之參考。

## 參考文獻

王秀禾、孫嘉玲、陳美慧、洪兆嘉、周培萱、賴玉敏  
 …陳權豐(2015)·客觀性結構式臨床測驗(OSCE)應用於護理系課程之成效·*技術及職業教育學報*, 6(2), 1-15。[Wang, H. H., Sun, J. L., Chen, M. H., Hung, C. C., Chou, P. H., Lai, Y. H., ... Chen, Z. F. (2015). The effectiveness of an objective structured clinical examination (OSCE) training program in nursing. *Journal of Technological and Vocational Education*, 6(2), 1-15.] doi:10.6235/TVE.2404

余民寧(2011)·*教育測驗與評量：成就測驗與教學評量*(第三版)·台北市：心理。[Yu, M. N. (2011). *Education test and evaluation: Achievement testing and teaching assessment* (3rd ed.). Taipei City, Taiwan, ROC: Psychological.]

李選、張婷、顏文娟、王郁琇(2011)·由全球人才競爭思維探討護理師教考用制度之現況、困境與趨勢·*國家菁英季刊*, 6(4), 29-45。[Lee, S., Chen, T., Yen, W. J., & Wang, Y. S. (2011). Study on current situation, difficulties, and trends of RN teaching, testing, and employing from global competition perspectives. *National Elite*, 6(4), 29-45.]

李石增主編(2012)·*客觀結構式臨床測驗—從基礎建立到*

*執行與品質監測*·台北市：台灣愛思唯爾。[Lee, S. T. (Ed.). (2012). *The objective structured clinical examination establishment, implementation, quality control*. Taipei City, Taiwan, ROC: Elsevier Taiwan.]

李國箴(2014)·*臨床護理技術能力—OSCE之應用*·台北市：華杏。[Lee, K. C. (2014). *OSCEs in nursing*. Taipei City, Taiwan, ROC: Farseeing.]

林淑君、黃翠媛、蔡麗紅、余文彬(2015)·護理OSCE教案撰寫原則與範例·於范君瑜主編，*護理客觀結構式臨床技能測驗*(初版，73-88頁)·台北市：台灣愛思唯爾。[Lin, S. C., Huang, T. Y., Tsai, L. H., & Yu, W. P. (2015). Principles and examples of writing OSCE teaching plan. In J. Y. Fan (Ed.), *Objective structured clinical examination in nursing education* (1st., pp. 73-88). Taipei City, Taiwan, ROC: Elsevier Taiwan.]

范君瑜(2015)·OSCE測驗於台灣護理領域之現況與應用·於范君瑜主編，*護理客觀結構式臨床技能測驗*(初版，1-13頁)·台北市：台灣愛思唯爾。[Fan, J. Y. (2015). The current situation and application of nursing OSCE in Taiwan. In J. Y. Fan (Ed.), *Objective structured clinical examination in nursing education* (1st., pp. 1-13). Taipei City, Taiwan, ROC: Elsevier Taiwan.]

范君瑜、蔡青青、賀倫惠(2015)·OSCE及格標準設定·於范君瑜主編，*護理客觀結構式臨床技能測驗*(初版，177-194頁)·台北市：台灣愛思唯爾。[Fan, J. Y., Tsai, C. C., & Ho, L. H. (2015). Standard settings for the OSCE. In J. Y. Fan (Ed.), *Objective structured clinical examination in nursing education* (1st., pp. 177-194). Taipei City, Taiwan, ROC: Elsevier Taiwan.]

高聖博(2013)·*精通標準化病人*·台北市：力大圖書。[Kao, S. P. (2013). *Mastering standard patients*. Taipei City, Taiwan, ROC: Leader Book.]

童恒新、王采芷(2009)·身體評估之OSCE臨床能力鑑定簡介·於蔡秀鸞總校閱，*身體評估之OSCE臨床能力鑑定*(1-6頁)·台北市：台灣愛思唯爾。[Tung, H. H., & Wang, T. J. (2009). A brief introduction to OSCE clinical ability identification of body assessment. In S. L. Tsay (Ed.), *Physical assessment of OSCE clinical competence* (pp. 1-6). Taipei City, Taiwan, ROC: Elsevier Taiwan.]

楊政議、李麗紅、楊惠娟、陳淑齡(2016)·台灣護理專科學校OSCE之發展現況與困境·*澄清醫護管理雜誌*, 12(1), 40-47。[Yang, C. I., Lee, L. H., Yang, H. C., & Chen, S. L. (2016). The OSCE and five-year junior nursing colleges in Taiwan: Development and challenges. *Cheng Ching Medical Journal*, 12(1), 40-47.]

- 盧瑞華、王瑜欣(2014)·客觀結構式臨床測試於護理教育之應用·*長庚護理*, 25(3), 274–282。[Lu, J. H., & Wang, Y. H. (2014). The applications of objective structured clinical examination in nursing education. *Chang Gung Nursing*, 25(3), 274–282.] doi:10.3966/102673012014092503003
- 簡淑慧、余怡珍、劉英妹(2015)·OSCE測驗信效度分析·於范君瑜主編, *護理客觀結構式臨床技能測驗*(初版, 207–236頁)·台北市:台灣愛思唯爾。[Jane, S. W., Yu, I. C., & Liu, Y. M. (2015). The reliability and validity analysis of OSCE. In J. Y. Fan (Ed.), *Objective structured clinical examination in nursing education* (1st., pp. 207–236). Taipei City, Taiwan, ROC: Elsevier Taiwan.]
- Byrne, G. (2008)·*外科OSCE臨床技能實戰測驗*(蔡淳娟、陳德芳審閱, 廖翊舒編譯)·台北市:江山。(原著出版於2007)[Byrne, G. (2008). *Core clinical skills for OSCEs in surgery* (T. C. Tsai & D. F. Chen, Eds., Y. S. Liao, Trans.). Taipei City, Taiwan, ROC: Jiangshan. (Original work published 2007)]
- Dornan, T., & O'Neill, P. (2008)·*內科OSCE臨床技能實戰測驗*(簡竹君審閱, 張亦嚴編譯)·台北市:江山。(原著出版於2006)。(Dornan, T., & O'Neill, P. G. (2006). *Core clinical skills for OSCEs in medicine* (C. C. Chien, Eds., I. Y. Chang, Trans.). Taipei City, Taiwan, ROC: Jiangshan. (Original work published 2006).]
- Bartfay, W. J., Rombough, R., Howse, E., & Leblanc, R. (2004). Evaluation. The OSCE approach in nursing education. *The Canadian Nurse*, 100(3), 18–23.
- Brighton, R., Mackay, M., Brown, R. A., Jans, C., & Antoniou, C. (2017). Introduction of undergraduate nursing students to an objective structured clinical examination. *Journal of Nursing Education*, 56(4), 231–234. doi:10.3928/01484834-20170323-08
- Downing, S. M. (2004). Reliability: On the reproducibility of assessment data. *Medical Education*, 38 (9), 1006–1012. doi:10.1111/j.1365-2929.2004.01932.x
- Fan, J. Y., Chao, L. F., & Jane, S. W. (2014). Application of the objective structure clinical evaluation in evaluating clinical competence for a BSN program. *Open Journal of Nursing*, 4(12), 868–877. doi:10.4236/ojn.2014.412092
- Fan, J. Y., Wang, Y. H., Chao, L. F., Jane, S. W., & Hsu, L. L. (2015). Performance evaluation of nursing students following competence-based education. *Nurse Education Today*, 35(1), 97–103. doi:10.1016/j.nedt.2014.07.002
- Fuller, R., Pell, G., Homer, M., & Roberts, T. (2012). Comments on 'A systematic review of the reliability of objective structured clinical examination scores'. *Medical Education*, 46(3), 337. doi:10.1111/j.1365-2923.2011.04198.x
- Hodges, B., Turnbull, J., Cohen, R., Bienenstock, A., & Norman, G. (1996). Evaluating communication skills in the objective structured clinical examination format: Reliability and generalizability. *Medical Education*, 30(1), 38–43. doi:10.1111/j.1365-2923.1996.tb00715.x
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30(1), 17–24. doi:10.1037/h0057123
- Khattab, A. D., & Rawlings, B. (2008). Use of a modified OSCE to assess nurse practitioner students. *British Journal of Nursing*, 17(12), 754–759. doi:10.12968/bjon.2008.17.12.30293
- Lauder, W., Holland, K., Roxburgh, M., Topping, K., Watson, R., Johnson, M., ... Behr, A. (2008). Measuring competence, self-reported competence and self-efficacy in pre-registration students. *Nursing Standard*, 22(20), 35–43. doi:10.7748/ns2008.01.22.20.35.c6316
- Miller, G. E. (1990). The assessment of clinical skills/competence/ performance. *Academic Medicine*, 65(9, Suppl.), S63–S67. doi:10.1097/00001888-199009000-00045
- Mitchell, M. L., Jeffrey, C. A., Henderson, A., Glover, P., Nulty, D. D., Kelly, M. A., ... Knight, S. (2014). Using an objective structured clinical examination for bachelor of midwifery students' preparation for practice. *Women and Birth*, 27(2), 108–113. doi:10.1016/j.wombi.2013.12.002
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459–467. doi:10.1002/nur.20199
- Rentschler, D. D., Eaton, J., Cappiello, J., McNally, S. F., & McWilliam, P. (2007). Evaluation of undergraduate students using objective structured clinical evaluation. *Journal of Nursing Education*, 46(3), 135–139.
- Roberts, C., Newble, D., Jolly, B., Reed, M., & Hampton, K. (2006). Assuring the quality of high-stakes undergraduate assessments of clinical competence. *Medical Teacher*, 28(6), 535–543. doi:10.1080/01421590600711187
- Rushforth, H. E. (2007). Objective structured clinical examination (OSCE): Review of literature and implications



- for nursing education. *Nurse Education Today*, 27(5), 481–490. doi:10.1016/j.nedt.2006.08.009
- Selim, A. A., Ramadan, F. H., El-Gueneidy, M. M., & Gaafer, M. M. (2012). Using objective structured clinical examination (OSCE) in undergraduate psychiatric nursing education: Is it reliable and valid? *Nurse Education Today*, 32(3), 283–288. doi:10.1016/j.nedt.2011.04.006
- Silva, C. C. B. M., Lunardi, A. C., Mendes, F. A. R., Souza, F. F. P., & Carvalho, C. R. F. (2011). Objective structured clinical evaluation as an assessment method for undergraduate chest physical therapy students: A cross-sectional study. *Brazilian Journal of Physical Therapy*, 15(6), 481–486. doi:10.1590/S1413-35552011005000033
- Sturpe, D. A., Huynh, D., & Haines, S. T. (2010). Scoring objective structured clinical examinations using video monitors or video recordings. *American Journal of Pharmaceutical Education*, 74(3), 44. doi:10.5688/aj740344
- Swift, M., Spake, E., & Gajewski, B. J. (2013). Student performance and satisfaction for a musculoskeletal objective structured clinical examination. *Journal of Allied Health*, 42(4), 214–222.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. doi:10.5116/ijme.4dfb.8dfd
- Wessel, J., Williams, R., Finch, E., & Gémus, M. (2003). Reliability and validity of an objective structured clinical examination for physical therapy students. *Journal of Allied Health*, 32(4), 266–269.

## 附件一

## OSCE 考題評分表 (以腹痛檢查及糖尿病衛教為例)

腹痛檢查				糖尿病衛教			
項目	沒有做到	部分做到	完全做到	項目	沒有做到	部分做到	完全做到
01. 症狀 + 時間				01. 詢問個案有無不適症狀感覺			
02. 疼痛的位置及輻射的位置				02. 向個案說明要做評估			
03. 疼痛的性質及疼痛的嚴重度				03. 詢問平時血糖值的程度			
04. 疼痛對日常生活的影響				04. 詢問今天早餐進食內容與時間			
05. 疼痛頻率與持續的時間				05. 詢問平日用餐進食內容與時間			
06. 什麼情況下會減輕疼痛及增加疼痛				06. 詢問今日早餐前後有無服用降血糖藥			
07. 慢性疾病或重大疾病				07. 詢問早上時段運動情形			
08. 意外事件及住院記錄				08. 告知飯前與飯後血糖的正常值			
09. 目前藥物治療				09. 告知個案此次高血糖的原因至少 2 項			
10. 旅遊史				10. 衛教個案飲食原則至少 2 項			
11. 聽診腹部四象限腸音				11. 衛教個案運動原則 2 項			
12. 叩診腹部四象限				12. 衛教個案藥物作用與使用原則 2 項			
13. 輕觸診腹部四象限				13. 個案提出糖尿病造成的影響，能回答相關合併症			
14. 深觸診腹部四象限				14. 使用個案聽得懂的語言，避免艱澀醫學術語。			
15. 疼痛象限最後檢查							
16. 反彈痛檢查							
17. 急性膽囊炎徵象檢查							

註：OSCE = objective structured clinical examination。

**引用格式** 黃湘萍、趙莉芬、王瑜欣、劉英妹、倪麗芬、簡淑慧 (2017) · 建構及檢測護理客觀結構式臨床技能測驗考題之信效度、鑑別度及難易度 · *護理雜誌*, 64(6), 67–76。[Huang, H. P., Chao, L. F., Wang, Y. H., Liu, Y. M., Ni, L. F., & Jane, S. W. (2017). The establishment and examination of the reliability, validity, discrimination, and difficulty of nursing objective structured clinical examination (OSCE). *The Journal of Nursing*, 64(6), 67–76.] doi:10.6224/JN.000084

# The Establishment and Examination of the Reliability, Validity, Discrimination, and Difficulty of Nursing Objective Structured Clinical Examination (OSCE)

Hsiang-Ping Huang<sup>1</sup> • Li-Fen Chao<sup>2</sup> • Yu-Hsin Wang<sup>2</sup> • Ying-Mei Liu<sup>3</sup>  
Lee-Fen Ni<sup>1</sup> • Sui-Whi Jane<sup>4\*</sup>

**Background:** The objective structured clinical examination (OSCE) is one of the evaluation strategies used to examine the clinical competence of nursing students. More attention needs to be paid to establishing an examination that is based on the standardized protocol and on reliability and validity.

**Purposes:** The aims of this study were to develop an OSCE that is based on a standardized protocol and to analyze the content validity and predictive validity, inter-rater consistency, internal consistency, item-to-total correlation, and the degree of discrimination and difficulty of this OSCE.

**Methods:** Based on the blueprint for OSCEs that use an 8-station format, this study used a descriptive, cross-sectional design and a sample population of 207 senior students who were preparing for their last-mile practicum at a nursing technology university in northern Taiwan. The study period lasted from September 2015 to October 2016. Furthermore, the validity, reliability, degree of discrimination, and difficulty of the OSCE were examined.

**Results:** An OSCE with an eight-station format was developed. The content validity index (CVI) was .85–1.0, inter-rater reliability was .80–1.0, Cronbach' alpha internal consistency was .53–.82, and item-to-total correlation was .36–.63. In addition, the level of difficulty of the OSCE was easy to moderate and the discrimination of the OSCE ranged from fair to excellent. Meanwhile, the OSCE scores of the participants appeared to be a valid predictor of scores on the last mile practicum.

**Conclusion/Implications for Practice:** According to the standardized protocol of OSCE development, an 8-station nursing OSCE format is reliable and valid, providing fair levels of difficulty and discrimination. The developed OSCE may be applied to evaluate the clinical competence of nursing students prior to beginning their last mile practicum. The results from the present study may provide a reference for nursing educators who are responsible for developing OSCE evaluation strategies.

**Key Words:** objective structured clinical examination (OSCE), reliability, validity, difficulty, discrimination.

---

Accepted for publication: October 13, 2017

<sup>1</sup>PhD, RN, Assistant Professor, Department of Nursing, Chang Gung University of Science and Technology; <sup>2</sup>PhD, RN, Associate Professor, Department of Nursing, Chang Gung University of Science and Technology; <sup>3</sup>PhD, RN, Associate Professor, Department of Nursing, Chang Gung University of Science and Technology, and Associate Research Fellow, Division of Pediatric Hematology and Oncology, Linkou Chang Gung Memorial Hospital; <sup>4</sup>PhD, RN, Associate Professor, Department of Nursing and Graduate Institute of Nursing, Chang Gung University of Science and Technology, and Associate Research Fellow, Division of Hematology/Oncology, Department of Internal Medicine, Linkou Chang Gung Memorial Hospital.

\*Address correspondence to: Sui-Whi Jane, No. 261, Wenhua 1st Rd., Guishan District, Taoyuan City 33303, Taiwan, ROC. Tel: +886 (3) 211-8999 ext 3136; E-mail: swjane@mail.cgust.edu.tw