

從使用者觀點檢視虛擬實境技術 在衡鑑能力評量的可行性*

藍玉玲

國立東華大學
諮商與臨床心理學系

陳畹蘭

慈濟大學
人類發展與心理系

張玉萱

元智大學
管理學院

王櫻芬

國立臺灣師範大學
教育心理與輔導學系

本研究將虛擬實境 (virtual reality, VR) 技術用於「VR 版心理衡鑑初次晤談客觀結構式臨床測驗」(簡稱「VR 版初次晤談 OSCE」) 的開發, 並透過使用者觀點檢視 VR 技術在衡鑑能力評量的可行性。本研究包含測試階段與正式研究。在測試階段, 本研究聘任六位諮商/臨床心理碩士生與心理師擔任模擬考生, 檢驗「初次晤談 OSCE」在諮商與臨床心理的通用性。模擬考生在「初次晤談 OSCE」的測驗分數顯示, 兩組模擬考生在「病史詢問」和「溝通技巧與專業態度」面向, 皆未呈現顯著差異, 初步證實其在諮商與臨床心理的通用性。正式研究階段, 本研究招募 37 位研究參與者, 分成臨床組學生、諮商組學生與諮商心理師三組, 檢驗 VR 技術在衡鑑能力評量的可行性與 VR 系統的設計品質。三組參與者在「心理衡鑑能力評量表——簡明版」測量結果顯示, 「VR 版初次晤談 OSCE」用於「鑑別診斷」與「統合推論」的可行性優於「溝通技巧」與「專業態度」。在「虛擬實境使用性量表」測量結果顯示, 參與者在 VR 系統設計方面, 對於「使用者介面」與「求助功能」評價最高, 認為「語音辨識」與「互動性」設計仍有待改善。從使用者觀點而言, VR 技術在衡鑑能力評量的應用具有可行性, 待系統開發成熟後, 可進一步評估其在 OSCE 施測成本節約的效果。

關鍵詞：客觀結構式臨床測驗、心理衡鑑、虛擬實境技術

* 1. 通訊作者：藍玉玲, ylan@mail.ndhu.edu.tw。

2. 作者們感謝科技部經費補助 (計畫編號: MOST 108-2410-H-259-015), 王勇智、楊長青與周大為的技術協助, 審查委員們對本文的專業建議, 以及本計畫研究助理們的行政協助。

伴隨臺灣社會對於精神醫療服務的需求日增，以及社會大眾對於心理師專業性的重視。立法院三讀通過《心理師法》，從法律層面保障心理師的執業專業性，也同時透過法律條文規範心理師執照的應考資格與實習訓練要求，進而掌握心理師養成教育的專業性（「心理師法施行細則」，2011）。與醫師、護理師相比，國內心理師證照考試仍以紙筆測驗為主，偏重知識面專業能力評估，較少透過實作能力評量（performance-based assessment），檢視考生面對臨床個案的實務技巧與專業態度（許文耀、張素鳳，2008；藍玉玲等人，2011）。

近年來，考試院積極推動專技醫事人員國家考試革新。不論醫師、牙醫師，或是護理師國考，皆從專業知識為主的評量模式，逐漸轉型成兼具專業知識與實務能力的評量模式，並將「客觀結構式臨床測驗」（Objective Structured Clinical Examination, OSCE）評量模式納入醫事人員國家考試中，對醫事人員的專業知識、臨床技術與專業態度，進行綜合性專業能力評估（李選等人，2015；劉克明、曾慧敏，2010）。近五十年來，OSCE 不僅成為醫事人員專業能力評量最常用的評量模式之一（Hodges, 2006; Rushforth, 2007; Walsh et al., 2009）。多個統合分析（meta-analysis）也持續證實 OSCE 不但具有良好的信效度、可有效地評估醫事人員的臨床實務能力（Navas-Ferrer et al., 2017; Rushforth, 2007; Walsh et al., 2009）。OSCE 也不斷地納入各類臨床技能的教學訓練課程中，成為良好的教學輔助工具（Day et al., 2018; Yedidia et al., 2003）。

在臺灣，OSCE 已廣泛地應用於醫師與專科護理師專業能力評量（李選等人，2015；劉克明、曾慧敏，2010；蔡詩力等人，2008）；但 OSCE 在國內心理師專業能力評量的應用，仍有待推廣。藍玉玲等人（2011）提出四項影響 OSCE 用於心理專業能力評量的可能因素：（1）心理學專業能力較醫學、護理學抽象且不易評估；（2）不易形成客觀的評量準則；（3）心理學 OSCE 較不易編製，以及（4）心理學 OSCE 在施測上，有較多的成本考量。

以醫學 OSCE 為例，國內許多醫學院系與醫學中心已投注大量人力物力，定期舉辦大規模醫學 OSCE 評量測驗，其中有些機構不但設置標準化病人中心／臨床技能中心，提供充沛的試務行政資源（高聖博，2013），有些機構更挹注可觀經費興建 OSCE 考場，並取得國家認證，成為醫學 OSCE 國家級測驗中心（李選等人，2015）。即便醫學院系與醫療機構已投入大量的人力資源與經費支持，但對考生而言，OSCE 仍是一種費用高昂的測驗。依據 Walsh 等人（2009）提供的 OSCE 測驗費用估計，每位考生參加一次 OSCE 至少需要花費新臺幣 22,465 元到 26,813 元。

對比醫學 OSCE 的資源投入程度，心理學 OSCE 的資源挹注則相對不足，這也是影響國內心理學 OSCE 應用、推廣的關鍵因素之一。國內雖有不同研究證實 OSCE 用於心理專業能力評量的可行性（盧怡蓓等人，2017；藍玉玲等人，2015），也有效地克服上述前三項因素。但是，這些研究對於心理學 OSCE 的施測成本問題，尚未提供理想的解決方案。在資源相對有限之下，若能夠有效地降低心理學 OSCE 的施測成本，將有助於 OSCE 在心理專業能力評量的推廣與應用。

為克服心理學 OSCE 的施測成本問題，本研究以藍玉玲等人（2015）發展之「心理衡鑑初次晤談客觀結構式臨床測驗」（簡稱為「初次晤談 OSCE」）為基礎，透過虛擬實境（virtual reality, VR）技術，設計開發「VR 版初次晤談 OSCE」，並期待透過 VR 版本來有效降低心理學 OSCE 的施測成本與人力需求，以推廣 OSCE 在心理專業能力評量的應用。本研究擬透過「使用者觀點」，評估 VR 技術用於衡鑑能力評量的可行性。待瞭解使用者對此種新型態測驗的接受度後，未來再透過後續研究進一步評估「VR 版初次晤談 OSCE」的測量品質與此測驗用於衡鑑能力訓練的成效。

文獻探討

（一）能力本位之「衡鑑／診斷／個案概念化」專業能力評量

2002 年，美國心理學會、心理學博士後與實習中心，以及多個心理專業團體於美國亞利桑那州，共同籌辦「專業能力研討會：專業心理師教育認證之未來方向」（簡稱為「專業能力研討會」），廣邀全球各地的學者專家和實務工作者，從「能力本位觀點」（competency-based perspective），重新界定心理師應具備的專業能力範疇。與會學者專家們認為心理師專業能力應包含八個面向：「科學精神實踐」、「心理衡鑑」、「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「個人和多元文化」（Kaslow, 2004）。他們也在八大專

業能力理論基礎下，展開一系列能力本位的教育訓練、測驗發展與實徵研究（Kaslow et al., 2007; Krishnamurthy et al., 2004; Leigh et al., 2007; Rubin et al., 2007）。

國內心理學界與實務界也受到能力本位風潮影響，重新思考國內心理師的養成教育與實務訓練模式，並再次檢視國內心理師專業能力的評量範疇。林家興與黃佩娟（2013）透過實徵研究，提出六項諮商心理師應具備的專業能力：「衡鑑／診斷／概念化、介入、督導、諮詢、研究與評鑑能力、以及管理」。Lan 與 Chang（2016）透過混合研究法，提出九種國內碩士級心理師在衡鑑專業能力的評量內涵：「科學精神、人際關係、情感能力、心理學的科學基礎、專業形象、心理測驗知識、跨領域合作、衡鑑技巧，以及心理執業的科學基礎」。

不論國內、國外，不分臨床心理或諮商心理，學界與實務界都將「衡鑑／診斷／個案概念化」（簡稱為「心理衡鑑」）視為心理師必備的專業能力之一（林家興、黃佩娟，2013；Kaslow, 2004），也是心理師實務工作的主要項目之一（吳英璋等人，2001；許秋田等人，2009；Norcross & Karpiak, 2012）。心理衡鑑不僅是心理師養成教育的重要環節（黃政昌，2008；許文耀、張素凰，2008；American Psychological Association, 2009），也是心理師國考的考試科目之一（「專門職業及技術人員高等考試心理師考試規則」，2018），更屬於心理師法所規範的執業項目（「心理師法」，2020）。因此，從能力養成、工作需求和證照考試來看，心理衡鑑都是心理師專業能力評量的重要向度（吳英璋等人，2001；林家興、黃佩娟，2013；林丞增、林家興，2019；許育光、刑志彬，2019；許秋田等人，2009；Norcross & Karpiak, 2012）。如何依據能力本位觀點，重新開發適用於心理衡鑑的能力評量工具，也是心理學的關注焦點之一。

1. 能力本位之心理衡鑑能力評估面向

依據能力本位觀點，參與「專業能力研討會」的「心理衡鑑工作小組」（psychological assessment work group, PAWG）重新界定心理衡鑑專業能力的評量範疇，包含八個向度：（1）瞭解心理測驗基礎知識；（2）瞭解心理衡鑑基礎知識；（3）具備心理學基礎知識與評估技術；以衡量個案在認知、情感、行為與人格面向的情況；（4）具備評估治療或介入效果的能力；（5）瞭解在心理衡鑑的過程中，個案與心理師互動所產生的相互影響因素；（6）具備建立與維持心理衡鑑工作所需要之專業關係能力；（7）瞭解心理治療和心理衡鑑之間的關聯性；以及（8）熟悉心理衡鑑的專業技巧（Krishnamurthy et al., 2004）。但林家興（2013）認為臺美兩地心理師在訓練層級（碩士級 vs. 博士級訓練）、實習要求（一年全職實習 vs. 一年全職實習與一年實務工作訓練）與執業獨立性（半獨立 vs. 獨立執業）皆不同，並對於直接使用國外博士級心理師能力評量面向，評估國內碩士級心理師的專業性，提出質疑。

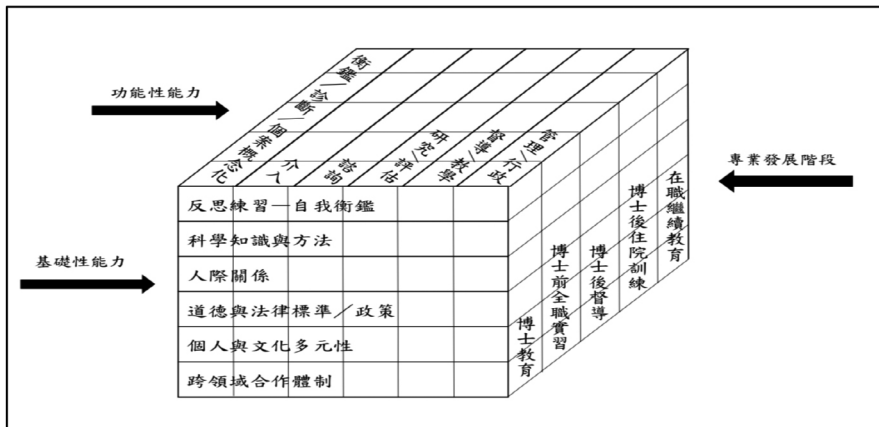
2. 國內心理衡鑑能力評估面向

為檢驗 PAWG 八大衡鑑能力評估面向用於國內心理師衡鑑能力評量的適用性，Lan 與 Chang（2016）透過混合研究法加以驗證，並發展出適用於國內碩士級心理師的能力本位評量工具：「心理衡鑑能力評量表」（Psychological Assessment Competency Scale, PACS）。依據其因素分析結果，PACS 萃取出九個國內碩士級心理師衡鑑能力的評量向度：（1）科學精神；（2）人際關係；（3）情感能力；（4）心理學的科學基礎；（5）專業形象；（6）心理測驗知識；（7）跨領域合作；（8）衡鑑技巧；以及（9）心理執業的科學基礎。Lan 與 Chang 發現，這九個衡鑑能力評量向度中，有四個面向（面向 2、5、6 和 8）與 PAWG 提出的面向一致，顯示不論碩士級或博士級心理師的衡鑑能力評量，都應包含「人際關係」、「專業形象」、「心理測驗知識」與「衡鑑技巧」。他們也額外發現，其他五個向度（如，科學精神、跨領域合作）與 Rodolfa 等人（2005）提出的「專業能力發展方塊模式」的一部份基礎性能力面向不謀而合，進而部份證實「專業能力發展方塊模式」的觀點。

圖 1 為 Rodolfa 等人（2005）提出的「專業能力發展方塊模式」，將心理師專業能力評量分成「基礎性能力」、「功能性能力」與「專業發展階段」三個維度加以評量。「基礎性能力」涵蓋六項心理師執業所需的基礎性能力，如，人際關係、科學知識與方法、反思練習——自我衡鑑等；「功能性能力」也涵蓋六項心理師執業所需的機能性能力，如，衡鑑／診斷／個案概念化、介入、諮詢等；「專業發展階段」則是將心理師專業發展階段分成五個階段，如，博士教育、博士實習、

在職教育等。Lan 與 Chang (2016) 的研究結果也呼應「專業能力發展方塊模式」的理論模式，證實國內碩士級心理師的衡鑑能力評量不只包含執行衡鑑業務的功能性能力，也包含執行衡鑑業務的基礎性能力。

圖 1
Rodolfa 等人 (2005) 的「專業能力發展方塊模式」



註:引自“A cube model for competency development: Implications for psychology educators and regulators,” by E. Rodolfa, R. Bent, E. Eisman, P. Nelson, L. Rehm, & P. Ritchie, 2005, *Professional Psychology: Research and Practice*, 36(4), p. 350 (<https://doi.org/10.1037/0735-7028.36.4.347>). Copyright 2021 by the American Psychological Association.

3. 自陳量表在心理衡鑑能力評量的應用與反思

為了進一步檢驗「專業能力發展方塊模式」所強調的不同專業發展階段應採用不同評量標準，許雅涵等人 (2016) 使用 PACS 的三個面向 (即，人際關係、心理測驗知識、衡鑑技巧) 修訂並完成適用於全職實習生的「心理衡鑑評量表——全職實習版」(PACS-I)。PACS-I 已證實適用於實習心理師自評與實習機構督導他評，皆具有良好的內部一致性信度 (實習生自評：Cronbach's $\alpha = .85 \sim .92$ ；督導他評：Cronbach's $\alpha = .83 \sim .97$)，與聚斂效度 (實習生自評： $r = .56 \sim .64$ ；督導他評： $r = .44 \sim .80$)，並能夠有效地評估實習心理師在實習前後的衡鑑能力變化。然而，林家興等人 (2015) 與許雅涵等人都認為使用自陳量表評估心理師專業能力需注意測量上的潛在問題：(1) 使用自陳量表進行自評與他評專業能力，兩者的關聯性可能不高；(2) 以自陳量表評估專業能力時，個人評分標準可能影響其自評/他評分數。例如，評分標準較寬鬆者容易產生仁慈偏誤 (leniency bias)，而提供較高的自評/他評分數。但評分標準較嚴格者則容易產生謙虛偏誤 (modesty bias)，而形成較低的自評/他評分數。因此，他們建議除了自陳量表之外，也可透過其他評量方式來評估心理專業能力。

4. OSCE 評量模式在心理衡鑑能力評量的應用

在兼顧基礎性能力與功能性能力評估，並達到同時評估心理師專業知識、臨床技能與專業態度的要求，學者專家們不斷審視 OSCE 用於心理專業能力教育訓練和測驗評量的可行性。在教育訓練方面，已有不同研究團隊將 OSCE 評量模式的「標準化病人」(standardized patient, SP) 概念用於心理師臨床技能的教育訓練上，並成為良好的教學輔助工具 (沈煜棠, 2015; Melluish et al., 2007)。

在測驗評量方面，藍玉玲等人 (2015) 以衡鑑能力評量為目標，發展「心理衡鑑——初次晤

談 OSCE」，並證實 OSCE 用於臨床心理碩士生在「心理衡鑑——初次晤談」專業知識、晤談技巧與專業態度評量的可行性。盧怡蓓等人（2017）透過行動研究法，分享馬偕紀念醫院 OSCE 用於諮商心理師訓練評量的實務經驗。Meghani 與 Ferm（2021）也將此評量模式納入美國舊金山大學（University of San Francisco）的 PsyD 學生資格考，評估 PsyD 學生在臨床實務能力的專業性。

陳皎眉等人（2015）從國家考試的角度，思考 OSCE 納入心理師國考的可行性，並建議合併臨床與諮商心理師考生，把心理師考試從一年兩次調整成一年一次，來降低試務成本，進而促進心理師 OSCE 國考的試務推動。未來，心理學 OSCE 評量測驗的發展也可參考陳皎眉等人的建議，朝向諮商與臨床心理師通用型 OSCE 為目標，以促進 OSCE 在心理專業能力評量的普及性。

（二）虛擬實境技術在醫療照護與醫事人員訓練的應用

虛擬實境（virtual reality, VR）透過電腦軟硬體技術，打造三維（three-dimensional, 3D）虛擬空間環境，讓使用者透過頭戴設備或感應裝置（如，3D 眼鏡、感應式手套），與環境中的虛擬物件互動，或是透過即時視聽感官刺激，產生身歷其境的臨場感受（Burdea & Coiffet, 2003）。伴隨 VR 技術的日新月異，VR 技術的應用已從軍事訓練，推廣至醫療照護、教育訓練，以及娛樂遊戲產業（李喬芳等人，2018；賴崇閔等人，2009；Fertleman et al., 2018）。

1. VR 技術在醫療照護與心理治療的應用

在醫療照護方面，VR 技術已大量應用於手術模擬與醫療專業人員訓練，且證實具有極大的貢獻，不僅提高醫師培訓效能，更大幅降低手術風險、確保醫療品質（李喬芳等人，2018；Lu et al., 2005）。在心理治療方面，VR 技術也用於不同類型恐懼症（如，懼高症、飛行恐懼症、蜘蛛恐懼症等）皆有良好療效（Power & Emmelkamp, 2008; Rothbaum et al., 2006）。VR 介入治療法可協助思覺失調症個案改善社會認知功能與社交技巧（Rus-Calafell et al., 2018）、進行職涯復健（Tsang & Man, 2013），並提高生活自理能力。Fodor 等人（2018）選擇 39 個使用 VR 介入治療法的實徵研究，進行統合分析，並證實 VR 介入治療法在焦慮症與憂鬱症的治療效果不但顯著，且治療效果已達到中到大效果量（effect size）。

Powers 與 Emmelkamp（2008）的統合分析更進一步發現，VR 介入治療法在治療效果方面與傳統治療法的療效相當，但是在成本與便利性考量方面，卻遠勝於傳統治療法。雖然心理師大多肯定 VR 介入治療法的療效，也認為與投入成本相比，是值得的投資（Segal et al., 2011）。但是，Segal 等人（2011）也發現，心理師的理論取向、對 VR 技術的了解與接受度，都影響他們對於 VR 介入治療法的成效評估。因此，未來研究 VR 介入治療法的成效評估時，宜將心理師個人變項（如，教育程度、訓練背景、年資、理論取向等）的影響，納入考慮。

2. VR 技術在醫事人員教育訓練的應用

VR 技術在醫事人員教育訓練應用，已逐年遞增。Kononowicz 等人（2019）對「虛擬標準化病人」（virtual standardized patient, VSP）在醫事人員訓練應用，提出文獻回顧專文。他們從 MEDLINE、Web of Science、PsycINFO 等 7 個資料庫，選取 1990 年到 2018 年間的實徵研究共 51 篇，進行統合分析。研究結果顯示，VSP 教學與傳統教學在知識面的教學效果相似，無顯著差異；但是，VSP 教學在技術面的教學效果，略優於傳統教學，其效果量介於小到中效果量之間。

在精神科教學應用方面，VSP 已用於憂鬱症（Shah et al., 2012）、自殺風險評估（Foster et al., 2015）、行為規範障礙症（Parsons et al., 2008）與失智症（Matsumura et al., 2018），衡量醫學院學生對於精神疾患個案的鑑別診斷能力、溝通技巧與治療處遇能力。Shah 等人（2012）認為 VSP 不僅可用於醫學生病史詢問、身體檢查、醫療倫理等專業能力評估，也可用於臨床教學上，透過 VR 互動式教學，對於醫學生的實際表現，提供即時且客觀的回饋，並可以提供學生課後反覆練習的機會，協助學生自主強化個人的專業知能。

(三) VR 版初次晤談 OSCE 的系統開發與挑戰

Burdea 與 Coiffet (2003) 認為成功虛擬實境設計需具備三種特質：沉浸性、互動性與想像力。「沉浸性」指使用者在虛擬環境中，透過感應裝置配戴，產生視覺、聽覺等感官擬真刺激，達到身歷其境、信以為真的知覺感受。「互動性」指使用者在虛擬環境中，透過感應裝置操作電腦程式介面以形成反應，VR 系統再依據使用者反應，提供即時回應，達到人機互動性。「想像力」指使用者在虛擬環境中，運用個人想像力，形成與現實環境不同的想像空間。依據上述特質，本文以表 1 所列三個研究團隊設計的 VSP 系統為例，分析 VR 版 OSCE 在 VSP 系統的開發現況與挑戰。

表 1
三個 VSP 系統在精神科教學應用之系統設計特色與改善方向

| 研究團隊 | Shah 等人 | Foster 等人 | Matsumura 等人 |
|------------|--------------------------|-------------------------------|-----------------------------|
| 疾病 | 憂鬱症 | 自殺風險評估 | 失智症 |
| VSP 系統設計特色 | 文字互動為主，且 VSP 表情變化不大 | 文字互動為主，且 VSP 靜態圖像呈現 | 選單點選為主，且 VSP 臉部有表情變化 |
| VSP 系統改善方向 | 增加表情與聲音對白，且增加同理心對白評估專業態度 | 運用多媒體增加情境互動真實性，且增加 VSP 與學生互動性 | 使用語音辨識技術，讓學生可自由詢問，而非預設選單式詢問 |

1. 沉浸性

比較上述三個 VSP 系統的「沉浸性」設計，Matsumura 等人 (2018) 的設計最佳，其 VSP 不再只以靜態圖像呈現，而是具有臉部表情變化。在多媒體技術日新月異下，VR 版 OSCE 的 VSP 系統在「沉浸性」設計，可透過細緻的 3D 動畫人物繪製，輔以生動的臉部表情和肢體動作，並透過 VSP 聲音表情的強化，提高 VSP 模擬真人標準化病人（簡稱為「真人 SP」）的效果，使學生更容易產生信以為真的知覺感受。

2. 互動性

在「互動性」方面，上述三個 VSP 系統採用「打字輸入」或「選單點選」方式，形成學生與 VSP 的互動性模式。這兩種互動性模式的優點是容易預先建置問答資料庫，讓學生與 VSP 形成互動，但缺點是擬真性有限，難以達到替代真人 SP 的效果。Matsumura 等人 (2018) 建議，未來可透過「自動語音辨識」（Automatic Speech Recognition, ASR）技術，設計 VSP 系統的互動性模式。本研究所設計的 VSP 系統使用 ASR 技術，讓學生透過「自由提問」方式與 VSP 形成互動，達到更貼近真人 SP 與學生互動的效果，以改善「互動性」問題。

3. 想像力

在「想像力」方面，當 VSP 系統的「沉浸性」與「互動性」都大幅提高時，將使 VSP 更近似真人 SP、更有效地達到替代真人 SP 的效果。學生在高度擬真的虛擬實境環境中，將更容易發揮想像力，順利進行各種類型的專業能力訓練與評量測驗。

除了「沉浸性」、「互動性」與「想像力」，Shah 等人 (2012) 額外提到 VSP 腳本內容多以病史評估為主，建議可增加同理心對白，來評估學生的專業態度。因此，「VR 版初次晤談 OSCE」的 VSP 系統之初期開發，包含下列三項：

(1) 在互動性方面。「VR 版初次晤談 OSCE」擬採用 ASR 技術，讓學生如同面對真人 SP 一樣，透過自由提問，瞭解 VSP 的預設病況，形成貼近真實臨床現場的互動性晤談模式。

(2) 在沉浸性方面。「VR 版初次晤談 OSCE」擬透過精緻 3D 虛擬人物、環境與物件的建置，搭配頭戴裝置與感應式手套，讓學生更容易產生接近真人 SP 互動的擬真狀態，來增加學生「身歷其境」的沉浸性感受。為提昇 VSP 的擬真性，以利學生進入想像的虛擬實境環境，本研究在 VSP 系

統開發時，先培訓真人 SP，透過模擬考生測試，確認真人 SP 的演出品質。待真人 SP 的演出品質確認後，再展開 VSP 對白的錄音工作，並透過聲音表情強化，提高 VSP 擬真性。

(3) 在想像力方面。「VR 版初次晤談 OSCE」的 VSP 個案內容設計除了病史詢問之外，也包含溝通技巧與專業態度的內容設計，以貼近真實臨床個案狀態，讓學生更容易發揮想像力，順利展開與 VSP 的初次晤談工作。圖 2 為「VR 版初次晤談 OSCE」在 VSP 系統的人物、場景設計圖。

圖 2

「VR 版初次晤談 OSCE」在 VSP 系統的人物、場景設計圖



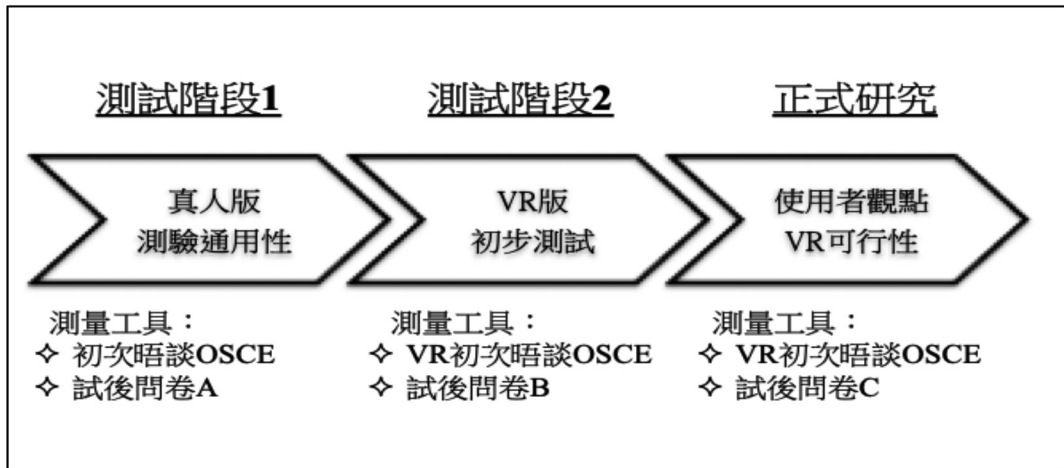
註：引自《虛擬實境技術在心理衡鑑能力評估之應用》（計畫編號：MOST108-2410-H-259-015）（頁 7），藍玉玲，2021，科技部補助研究計畫成果報告。

綜言之，本研究的研究目的有二：（1）完成「VR 版初次晤談 OSCE」的初期開發；（2）以使用者觀點，評估 VR 技術在衡鑑能力評量的可行性與 VSP 系統的設計品質。待本測驗發展成熟，能使用 VSP 替代真人 SP，進行一部份衡鑑能力評估工作時，再進一步評估本測驗用於心理衡鑑能力評量的測量品質和其在衡鑑教育訓練應用的成效。

研究架構

本研究包含「測試階段」與「正式研究」。「測試階段」包含「真人版測驗通用性評估」與「VR 版測驗初步測試」兩階段。「真人版測驗通用性評估」階段，考量「初次晤談 OSCE」用於臨床心理衡鑑能力評量，已獲得實徵研究證實（藍玉玲等人，2015），但尚未用於諮商心理。依循陳皎眉等人（2015）觀點，本研究團隊擬致力開發適用於臨床與諮商心理之心理衡鑑 OSCE 通用型測驗。因此，在導入 VR 技術、完成「VR 版初次晤談 OSCE」之前，本研究擬在「測試階段」，先評估「初次晤談 OSCE」用於臨床與諮商心理的測驗通用性。待確認真人版「初次晤談 OSCE」的測驗通用性之後，再展開通用型「VR 版初次晤談 OSCE」的開發與初步測試。待 VR 版本初步測試完畢後，本研究再進入「正式研究」階段，透過公開招募研究參與者，參與實徵研究，檢驗 VR 技術在衡鑑能力評量的可行性。圖 3 為本研究的研究架構。

圖 3
研究架構



測試階段 1：真人版測驗通用性評估

(一) 研究參與者與研究程序

本研究聘任培訓一位工作人員擔任真人 SP，扮演「初次晤談 OSCE」的模擬個案，協助進行本測驗的通用性評估；並聘任六位工作人員，擔任「模擬考生」工作，協助測試真人版測驗的通用性評估。這六位模擬考生中，來自臨床心理與諮商心理訓練者各半，包含四位碩士生（含全職實習生），一位碩士畢業生與一位新進心理師；共有四位女性，兩位男性。所有模擬考生都先接受「初次晤談 OSCE」，其中包含 SP 站實作測驗與紙筆站簡答題分測驗，再填寫一份「試後問卷 A」。

模擬考生在 SP 站實作測驗的表現，由負責試務的工作人員全程錄音錄影，再交由評分者評分。他們的試後問卷結果擬用來評估此測驗的施測規劃是否需要調整，以及測驗內容用於臨床與諮商心理的通用性。模擬考生在「初次晤談 OSCE」的測驗結果，由三位評分者獨立評分，以檢驗此測驗在臨床與諮商心理的通用性。三位評分者由本文第二、三、四作者擔任，其中兩位具諮商心理師證照，一位具臨床心理師證照。三位評分者已事先參與本文第一作者提供之「評分者共識訓練工作坊」，完成評分共識訓練後，再開始評分。

「評分者共識訓練工作坊」包含兩階段「評分實作演練」與「評分共識討論」。在第一階段評分實作演練，三位評分者使用本文第一作者所提供的一組 SP 站學生錄影檔，進行第一次評分實作練習。待完成評分實作練習後，再依據評分結果，展開第一輪評分共識討論，確認 SP 站評分表在臨床與諮商心理的通用性，並完成評分尺規（scoring rubrics）的修訂。第二階段的評分實作演練，三位評分者使用另一組 SP 站學生錄影檔，進行第二次評分實作演練與共識訓練討論，再次確認三位評分者的評分標準已達到共識。

(二) 測量工具

1. 初次晤談 OSCE

藍玉玲等人（2015）將衡鑑任務設定為「初次晤談」，並參考 Maloney 與 Ward（1976）的衡鑑資料整合模式，模擬「心理衡鑑——初次晤談」的部份歷程，將此歷程概分為「初步資料蒐集」、

「形成推論」、「推論拒絕／修正／接受」、「發展並整合假設」和「形成初步治療／處遇方案」五個階段，並完成「初次晤談 OSCE」的測驗發展。「初次晤談 OSCE」包含「SP 站」與「紙筆站」，SP 站用來模擬前兩個階段，紙筆站則用來模擬後三個階段，以評估受試者在心理衡鑑的專業知識、技術與態度。

SP 站由已完成訓練的 SP 扮演憂鬱症個案角色，並提供一個衡鑑任務設定為「初次晤談」的臨床情境，讓受試者擔任「心理師」角色，與 SP 互動，衡量受試者在初次晤談過程中，所展現的衡鑑專業知識、溝通技巧與專業態度。受試者在 SP 站的行為表現，會透過「SP 站評分表」，由受過評分者訓練且具有心理師證照的評分者評估。「SP 站評分表」包含「病史詢問」與「溝通技巧與專業態度」兩面向。「病史詢問」面向包含 18 個評估題項（如，第一題「壓力的可能來源」、第 16 題「過去心理治療經驗」），由評分者以三點量尺（即，「未做」、「已做，待改進」和「已做，表現良好」），衡量受試者能否從壓力源、過去病史、社會支持等方向，蒐集個案就診原因的相關資料，以形成對於個案的推論／假設。「溝通技巧與專業態度」面向包含 10 個評估題項，也採用上述三點量尺，由評分者衡量受試者在晤談過程所展現的溝通技巧（如，第五題「技巧性地詢問敏感問題」）與專業態度（如，第八題「適時展現同理心」）。

「紙筆站」包含「選擇題」與「簡答題」兩個分測驗，前者包含 30 題選擇題，用來評估心理測驗與心理衡鑑的基礎知識，後者為「SP 站」的連貫測驗，包含五題簡答題、SP 站個案的模擬測驗結果及測驗手冊，用來評估受試者能否有效地整合不同來源的衡鑑資料、形成個案概念化，以提供初步治療／處遇方案。透過「SP 站＋紙筆站」的連貫測驗設計，「初次晤談 OSCE」先以 SP 站，衡量受試者與真人 SP 互動時，是否展現初步鑑別診斷個案現況與憂鬱症狀評估的專業知識、技巧和態度；再透過紙筆站簡答題分測驗，檢驗受試者先透過晤談過程，形成個案的假設／推論後，能否進一步整合分析 SP 個案的「晤談資料」與「模擬測驗結果」，以評估個案推論／假設是否接受或需要修正、拒絕，並提供初步治療／處遇方案；最後再以紙筆站選擇題分測驗，評估受試者在測驗、衡鑑的基礎知識。

「初次晤談 OSCE」已透過多特質——多方法分析 (multitrait-multimethod, MTMM) (Campbell & Fiske, 1959) 檢驗，並證實具有良好的構念效度，其中「SP 站病史詢問面向」與「紙筆站簡答題分測驗」間，呈現中度正相關 ($r = .44$)，顯示有良好的聚斂效度；且「紙筆站選擇題分測驗」和 SP 站「病史詢問」($r = .14$)、「溝通技巧與專業態度」面向 ($r = .03$)，皆呈現低度正相關，顯示有良好的區辨效度 (藍玉玲等人, 2015)。在內部一致性信度方面，SP 站評分表的整體信度 ($\alpha = .89$) 和兩面向的信度 (病史詢問： $\alpha = .80$ ，溝通技巧與專業態度： $\alpha = .88$) 都達到 DeVellis (1991) 建議之新測驗信度下限 ($\alpha = .70$)，證實 SP 站評分表具有良好的內部一致性信度 (藍玉玲等人, 2015)。本研究僅使用 SP 站與其連貫測驗「紙筆站簡答題分測驗」，進行模擬考生在真人與虛擬 SP 之比較。

2. 試後問卷 A

本研究修改藍玉玲 (2013) 研究計畫的自編試後問卷，衡量「初次晤談 OSCE」的施測規劃品質與測驗內容的適切性。試後問卷 A 的作答時間約 20 分鐘，於受試者完成「初次晤談 OSCE」後施測。問卷內容包含「參與者背景」、「課程與實習經驗」、「施測規劃品質評估」和「測驗內容適切性」。在「參與者背景」方面，包含性別、就讀學校、年級、組別 (臨床／諮商) 等背景資料。在「課程與實習經驗」方面，包含衡鑑相關課程訓練 (如，列出修過之心理衡鑑／心理評估課程)，以及全職與兼職實習經驗調查，例如「已進入／完成一年全職實習」，「是否有其他兼職實習經驗？」。在「施測規劃品質評估」方面，包含「測驗整體規劃合宜」、「測驗內容貼近實際臨床個案」、以及對於「SP 站」與「紙筆站」的建議事項等題項。在「測驗內容適切性」方面，參與者評估測驗內容用於「展現溝通技巧」、「鑑別診斷心理疾病」等衡鑑能力評估的有效性。

(三) 研究結果

本研究聘任六位工作人員 (臨床與諮商心理訓練背景者各半)，擔任「模擬考生」工作，協助測試「初次晤談 OSCE」在諮商與臨床心理的通用性。模擬考生在真人 SP 站的測驗結果，由三位

具心理師證照且完成評分共識訓練的評分者評分。在評分者信度方面，本研究透過 Kendall's W 和諧係數、顯著水準設為 .05 評估，研究結果顯示不論在「病史詢問」($W = 0.89, \chi^2(5) = 13.28, p < .05$) 或「溝通技巧與專業態度」($W = 0.85, \chi^2(5) = 12.67, p < .05$) 面向，三位評分者對六位模擬考生評分都呈現顯著關聯性，具有一定的評分者信度。因此，六位模擬考生的測驗分數將使用三位評分者評分結果的平均數，作為「測驗通用性」分析之用。

在「測驗通用性」方面，透過 Mann-Whitney U 檢定、顯著水準設為 .05，檢驗臨床與諮商組模擬考生在「初次晤談 OSCE」真人 SP 站的得分，是否呈現顯著差異。研究結果顯示，諮商與臨床組模擬考生在「病史詢問」($U = 3.00, p = .51$) 或「溝通技巧與專業態度」($U = 4.00, p = .83$) 面向，皆未呈現顯著差異。上述結果初步證實，「初次晤談 OSCE」在諮商與臨床心理的通用性，可使用此版本進行「VR 版初次晤談 OSCE」的開發與測試。

模擬考生在試後問卷 A 的回覆，大多認為「初次晤談 OSCE」的施測規劃良好，測驗內容不但貼近實際臨床個案，也能有效地評估鑑別診斷、溝通技巧、專業態度等衡鑑所需知能。三位評分者也從測驗內容與模擬考生表現評估，認為「初次晤談 OSCE」不僅適用於臨床心理，亦可用於諮商心理，成為衡鑑專業能力評量的通用測驗。因此，本研究擬以「初次晤談 OSCE」的測驗內容為基礎，進行「VR 版初次晤談 OSCE」的測驗發展與 VSP 系統建置。

測試階段 2：VR 版測驗初步測試

(一) 研究參與者與研究程序

測試階段 2 同樣聘任上述六位工作人員，擔任「模擬考生」工作，協助「VR 版初次晤談 OSCE」的初步測試，此次測試與測試階段 1 的測試時間，間隔約 2~3 個月。所有模擬考生都先接受「VR 版初次晤談 OSCE」(包含 VSP 站實作測驗與紙筆站簡答題分測驗)，再填寫「試後問卷 B」。模擬考生在「VR 版初次晤談 OSCE」的測驗結果也由上述三位評分者獨立評分，並透過評分者共識會議討論，完成 VSP 站評分表與評分尺規建置。

(二) 測量工具

1. 「VR 版初次晤談 OSCE」之測驗發展

「VR 版初次晤談 OSCE」包含一個 VSP 站實作測驗與一個紙筆站簡答題分測驗，其測驗發展擬分成「VSP 站測驗內容」、「VSP 系統設計」與「紙筆站測驗內容」三項，加以說明。

(1) VSP 站測驗內容。本研究以「初次晤談 OSCE」的「SP 站」模擬個案內容為基礎，修改成適用於「VSP 站」的個案內容與對白。在「VSP 個案對白設計」方面，本研究依據「初次晤談 OSCE」的「SP 站」個案內容設計對白，並參考 Gibbons 等人 (2012) 發展之「憂鬱症電腦化適性測驗 (Computerized Adaptive Test for Depression, CAT-DI)」，將 VSP 個案對白分成「憂鬱症狀評估」、「病史及治療評估」、「人際關係與社會支持評估」和「壓力源評估」四個向度，各向度之下可再細分成一到四個不等的次向度。以「憂鬱症狀評估」向度為例，可進一步細分為「情緒」、「認知」、「行為、習慣與生理狀況」和「自殺評估」四個次向度。

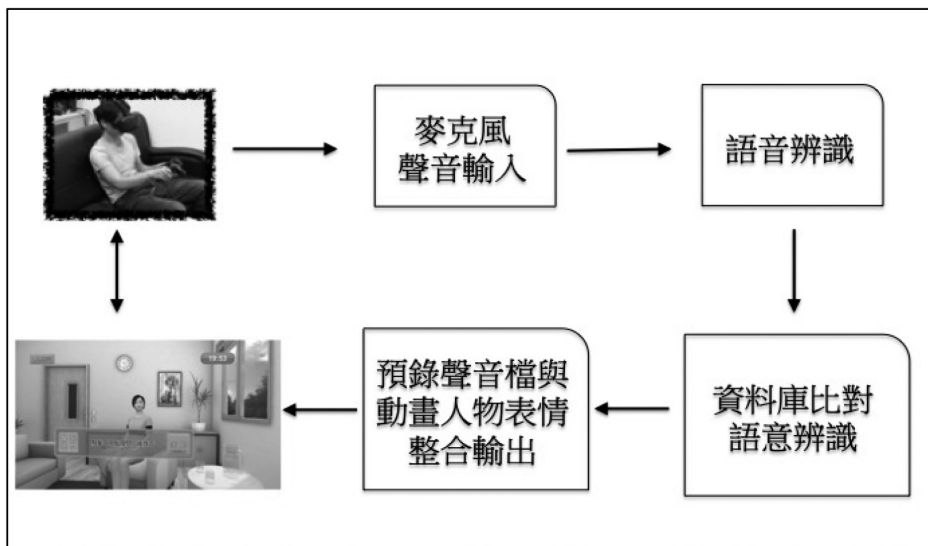
VSP 個案對白設計包含「VSP」與「考生」兩種，前者依據 VSP 個案腳本所寫，後者對照 VSP 個案腳本，並參考多個憂鬱症評估工具與實徵研究結果所寫。考生對白將匯入「VSP 站」的「求助功能」中，讓考生使用「求助功能」時，立即獲得 VSP 個案的可提問選項，輔助考生與 VSP 順利展開初次晤談工作，以蒐集 VSP 的就診原因、心理病史、社會支持等相關資訊，形成對於 VSP 個案病況的初步瞭解。

(2) VSP 系統設計。本研究在科技公司協助下，完成「VR 版初次晤談 OSCE」的 VR 應用程式之初期開發，並將此 VR 程式重新命名為「VR 心理衡鑑室」。依循表 1 對於 VSP 系統開發的建議，「VR 心理衡鑑室」在「VSP 系統設計」方面，採用 ASR 技術，讓考生透過「自由詢問問句」型式，與 VSP 展開初次晤談，蒐集 VSP 的就診原因、個人病史、壓力源、社會支持等資訊，以形成個案

病況的鑑別診斷和後續治療方案的提供。

在語音辨識方面，「VR 心理衡鑑室」使用麥克風收音，將考生的口述問句，透過網路連至「Microsoft Azure」雲端運算平臺，進行「語音辨識」工作；再將語音辨識結果，透過網路回傳至「VR 心理衡鑑室」，並使用系統內建資料庫，進行「語意識別」分析工作。待「VR 心理衡鑑室」將考生提問的語意分析結果，與資料庫內建的 VSP 回覆比對完畢後，將提供最適切的 VSP 回覆內容，再與預先錄製匯入的 VSP 真人語音檔，搭配已繪製的 3D 人物表情，合併輸出，形成 VSP 對於考生提問的答覆。圖 4 為「VR 心理衡鑑室」的「語音辨識」流程圖。

圖 4
「VR 心理衡鑑室」的「語音辨識」流程圖



(3) 紙筆站測驗內容。本研究在「VR 版初次晤談 OSCE」的紙筆站施測內容，僅包含「初次晤談 OSCE」的簡答題分測驗。此分測驗以 VSP 個案內容為基礎，提供 VSP 個案在指定測量工具的模擬測驗結果與測驗手冊，用來檢視考生能否整合 VSP 站的晤談資訊與紙筆站 VSP 個案的模擬測驗結果，對於 VSP 個案的心理疾病狀態，提供初步的鑑別診斷與後續的治療方案。

2. 試後問卷 B

試後問卷 B 用來評估「VR 版初次晤談 OSCE」的施測規劃品質與測驗內容適切性，作答時間約 20 分鐘，於受試者完成「VR 版初次晤談 OSCE」後施測。試後問卷 B 與試後問卷 A 大致相同，皆包含：(1) 參與者背景；(2) 課程與實習經驗；(3) 施測規劃品質評估；和(4) 測驗內容適切性，僅微調(3)與(4)的指導語，將評估對象從「初次晤談 OSCE」更改成「VR 版初次晤談 OSCE」。

(三) 研究結果

1. VR 版測驗初步測試

模擬考生在 VSP 站錄影資料顯示，本研究開發之「VR 心理衡鑑室」在「偵測問句——對應回覆形成」過程(見圖 4)所預設的「語意判定」標準過於嚴苛，以致考生的問話內容，不易達到系統預期的判定標準、難以與 VSP 形成互動，進而影響他們在 VSP 站的測驗表現。因此，在測試階

段，本研究僅依據模擬考生的錄影資料，完成 VSP 站評分表與評分尺規建置，而不提供評分結果。待本文第一作者重新修改語意判定標準，並再次修訂 VR 程式後，再以正式研究的受試者資料進行「VR 版初次晤談 OSCE」的評分工作。有關 VSP 站評分表的題項內容，將於「正式研究」的「測量工具」章節說明。

2. VR 版本施測規劃評估

有關 VR 版本的施測規劃與測驗內容評估，模擬考生提供的意見回饋如下：（1）與真人版相比，模擬考生認為「VR 版初次晤談 OSCE」，除了包含衡鑑能力評量之外，還包含對於 VR 設備與 VSP 系統使用操作的熟悉。對於沒有 VR 使用經驗的受試者而言，可能因 VR 設備與 VSP 系統使用操作不熟悉而影響測驗結果。因此，模擬考生建議在進入正式測驗階段，宜提供充足的練習時間，讓受試者先熟悉 VR 設備與 VSP 系統的使用操作，再進入正式測驗，以免影響測驗表現；（2）有模擬考生反映 VSP 站施測完畢後，有類似暈車的暈眩不適感，導致後續紙筆站的應答狀態不佳；（3）模擬考生大多認為 VR 版發展成熟後，可替代真人版，成為衡鑑能力的評量工具，也可成為教學輔助工具，讓學生在課後自主練習，強化自身的衡鑑能力。

依據模擬考生的意見回饋，本研究調整正式研究的施測程序。調整內容包含：（1）增加 VSP 站的練習測驗，使考生熟悉 VR 設備與 VSP 系統的使用、操作方式後，再開始 VSP 站的正式測驗；（2）調整「VSP 站與紙筆站」之間的休息時間，從 5 分鐘延長至 10 分鐘，並提供瓶裝水、口香糖、梅子、白花油等用品，協助考生舒緩可能的暈眩、不適感後，再接受紙筆測驗。

正式研究：VR 技術在衡鑑能力評量的可行性評估

（一）研究參與者與研究程序

正式研究共招募 37 位研究參與者，包含 28 位學生與 9 位心理師。學生群體包含臨床與諮商心理系所碩士生與全職實習生，心理師群體皆來自諮商心理訓練背景的心理師。學生群體的參與者以女性為主，佔 85.70%；來自諮商與臨床訓練者約為 6：4，有半數尚未進入一年全職實習，有半數已進入或已完成一年全職實習。心理師群體的參與者也以女性為主，佔 88.90%，皆具有諮商心理師證照，平均工作年資為 7.22 年。表 2 為正式研究的參與者基本資料摘要表。

表 2
正式研究樣本的基本資料摘要表

| 背景變項 | 學生 | | 心理師 |
|--------|------------|------------|-----------|
| | | N (%) | N (%) |
| 性別 | 男性 | 4 (14.30) | 1 (11.10) |
| | 女性 | 24 (85.70) | 8 (88.90) |
| 主修／證照別 | 諮商 | 17 (60.70) | 9 (100) |
| | 臨床 | 11 (39.30) | 0 (0) |
| 年級 | 碩士一、二年級 | 15 (53.60) | |
| | 碩士三、四年級及以上 | 13 (46.40) | |
| 一年全職實習 | 尚未進入 | 14 (50.00) | |
| | 已進入／已完成 | 14 (50.00) | |
| 工作年資 | 平均數 | | 7.22 |
| | 標準差 | | 8.90 |

所有參與者都先接受「VR 版初次晤談 OSCE」，再填寫「試後問卷」，其中「試後問卷」包含「學生版」與「心理師版」，將於測量工具章節說明。正式研究的施測程序有七項：（1）進行 10 分鐘

SP 站練習測驗；(2) 休息 5 分鐘；(3) 進行 20 分鐘 SP 站正式測驗；(4) 休息 10 分鐘；(5) 進行 35 分鐘紙筆站，完成紙筆測驗；(6) 休息 5 分鐘；(7) 填寫 20 分鐘試後問卷。這些資訊有助於評估 VR 技術用於衡鑑能力評量的可行性，也可檢視 VSP 系統設計的優、缺點。此外，諮商心理師們對於本測驗的適用性評估，可協助優化測驗內容，促進本測驗在諮商與臨床心理的通用性。

(二) 測量工具

正式研究的測量工具有二：VR 版初次晤談 OSCE 與試後問卷 C。「VR 版初次晤談 OSCE」已於「VR 版初次晤談 OSCE 之測驗發展」章節說明，以下僅簡述測驗內容與 VSP 站評分表的題項內容；並描述「試後問卷 C」的問卷內容。

1. VR 版初次晤談 OSCE

「VR 版初次晤談 OSCE」包含一個 VSP 站和一個紙筆站。紙筆站使用「初次晤談 OSCE」的「簡答題分測驗」，用來評估受試者能否有效地整合 VSP 站的晤談內容與紙筆站提供的 VSP 個案模擬測驗結果和測驗手冊，提出對於個案的推論／假設，並提供適當的初步治療／處遇。VSP 站透過虛擬 3D 動畫人物扮演標準化病人（見圖 2），提供考生一個初次晤談的臨床情境，讓考生與 VSP 互動，完成初次晤談的衡鑑工作。「VSP 站評分表」參考測試階段 2 的模擬考生 VSP 站錄影資料，將「初次晤談 OSCE」的「SP 站評分表」修改成「VSP 站評分表」，以供 VSP 站評分使用。

「VSP 站評分表」包含「病史詢問」與「溝通技巧與專業態度」兩面向。「病史詢問」面向包含原「SP 站評分表」的 18 個評估題項，由評分者以三點量尺（即，「未做」、「已做，待改進」和「已做，表現良好」），衡量受試者在初次晤談的病史資料蒐集能力。考量 VSP 系統仍在初期開發階段，VSP 在專業態度面向的模擬尚未成熟。因此，本研究在「VSP 站評分表」的「溝通技巧與專業態度」面向，僅包含原「SP 站評分表」的 7 個溝通能力評量題項，並簡化計分方式為「未做」與「已做」兩種，以評估受試者的溝通技巧。待「VR 版初次晤談 OSCE」完成後續的程式優化工作後，再放入「專業態度」的評估題項。附錄 1 包含 VSP 站評分表的題項內容。

2. 試後問卷 C

正式研究時，本文第一作者依據施測對象不同，自編「心理師版」與「學生版」兩種試後問卷 C，兩個版本的作答時間皆為 20 分鐘。「學生版」試後問卷 C 包含四個部份：(1) 研究參與者的背景資料、心理衡鑑相關課程訓練與實習經驗；(2) 自評衡鑑專業能力；(3) 虛擬實境使用性量表，(4) 心理衡鑑能力評量表——簡明版。第一部份涵蓋「試後問卷 A」在「參與者背景」和「課程與實習經驗」部份的所有題項。第二部份由受試學生評估自己與其他同組同年級學生在「VSP 站」與「紙筆站」的測驗分數，並以五點量表（1 = 極待加強～5 = 非常傑出）評估自己在「與個案建立專業關係」、「展現溝通技巧」、「鑑別診斷心理疾病」、「整合個案資訊形成推理假設」和「選擇適當的測驗工具與分數解釋」五項衡鑑能力指標的專業性。第三、四部份，將於下兩段逐一說明。「心理師版」與「學生版」相似，僅依據心理師身份，調整基本資料的題項內容，並將自評衡鑑專業能力調整為此測驗對臨床／諮商心理學生衡鑑能力評量的適用性。

3. 虛擬實境使用性量表

「虛擬實境使用性量表」（Usability Scale for the Virtual Reality Implementation, USVRI）包含「VSP 系統設計」、「VR 穿戴設備」與「VSP 替代性」三面向，共有 17 題，其中有 13 題採用李克特式五點量尺評估，以 1 代表非常不同意，到 5 代表非常同意，分數越高代表使用經驗越佳。USVRI 的測量結果可做為 VSP 系統設計與 VSP 替代性的評估指標，也可為 VR 版測驗的施測規劃，提供參考資訊。以下將簡述 USVRI 的測驗發展與題項內容。

(1) 「VSP 系統設計」面向。依據 Burdea 與 Coiffet (2003) 提出虛擬實境設計需具備的互動性與沉浸性特質，本文第一作者設計五種 VSP 系統設計的品質評估指標，分別是「使用者介面」、「語

音辨識」、「VSP 互動性」、「求助功能」與「VSP 擬真性」，並針對上述指標來發展題項。「VSP 系統設計」面向包含八個評估題項，其中有兩題評估「使用者介面」，題項內容為「施測說明清楚易懂」與「操作介面清楚易懂」；兩題為「語音辨識」功能評估，題項內容為「練習測驗：VR 系統能辨識我的提問內容」與「正式測驗：VR 系統能辨識我的提問內容」；「VSP 互動性」指標也有兩個評估題項，分別是「VR-SP 很配合、有問必答」與「VR-SP 經常答非所問」（為反向題）；「求助功能」與「VSP 擬真性」指標都各有一題，分別是「求助設計功能佳」與「表情生動、擬真」。

(2) 「VR 穿戴設備」面向。參考測試階段 2 模擬考生的意見回饋，本文第一作者設計此面向的三個評估題項，分別是：a. 「VR 設備配戴舒適」；b. 「VR-SP 站施測後，產生暈眩不適感」；和 c. 「建議休息時間為 ___ 分鐘，再接續紙筆測驗」。第一題以五點量尺評估 VR 穿戴設備的舒適度；第二題採「是／否」二元計分方式，評估使用 VR 設備後是否出現不適情況；第三題用來調查受試者使用 VR 後，需要的休息間隔時間。

(3) 「VSP 替代性」面向。Kenny 等人 (2007) 認為用於臨床治療技巧訓練的 VSP 系統，應以「VSP 互動性」與「系統預設特殊功能」為主，評估 VSP 系統的設計品質。參考上述建議，本研究以「VSP 替代真人 SP 在衡鑑能力評量」（簡稱為「VSP 替代性」）為主軸，完成「VSP 替代性」面向的題項發展，共有六題。第一至四題採上述五點量尺評估，題項內容包含「展現溝通技巧」、「展現專業態度」、「鑑別診斷心理疾病」與「整合個案資訊形成推理假設」，第五與六題則採開放式問句，題目內容為「VR-SP 替代真人 SP 的優缺點」與「VR-SP 替代真人 SP 的建議事項」，蒐集受試者的意見回饋。

4. 心理衡鑑能力評量表——簡明版

Lan 與 Chang (2016) 參考 PAWG 的八大衡鑑能力評量面向與兩階段專家訪談結果，完成「心理衡鑑能力評量表」(PACS) 的初稿，共包含 84 題與 13 個面向，用於國內碩士級心理師衡鑑能力評量。透過因素分析，PACS 從 84 題精簡至 45 題，並萃取出九個評量面向：(1) 科學精神（如，邏輯推理能力）；(2) 人際關係（如，關係建立）；(3) 情感能力（如，情緒穩定性）；(4) 心理學的科學基礎（如，心理藥物學知識）；(5) 專業形象（如，建立專業形象）；(6) 心理測驗知識（如，測驗分數解釋）；(7) 跨領域合作（如，跨領域合作能力）；(8) 衡鑑技巧（如，個案概念化）；以及 (9) 心理執業的科學基礎（如，人格理論與人格衡鑑方式），其累積總變異量為 65.82%，顯示 PACS 具有良好的構念效度。在內部一致性信度方面，除了「專業形象」與「心理測驗知識」面向的 Cronbach's α 值為 .76 與 .74 外，其他面向的 Cronbach's α 值界於 .83 到 .89，且整體測驗的 α 值為 .95，顯示 PACS 具有良好的內部一致性信度。

本研究以 Lan 與 Chang (2016) 發展的 PACS 為基礎，並參考林家興等人 (2015) 的「心理師能力評量表」之「衡鑑診斷與概念化能力」的測量內容，修訂編製成「心理衡鑑能力評量表-簡明版」(the short version of the PACS, PACS-S)。PACS-S 包含六種衡鑑能力評量指標，採用李克特式五點量尺評估，以 1 代表極待加強，到 5 代表非常傑出，分數越高代表衡鑑專業能力越好。PACS-S 六個題項的題目內容分別是「與個案建立專業關係」、「展現溝通技巧」、「展現專業態度」、「鑑別診斷心理疾病」、「整合個案資訊，形成推論假設」和「選擇適當的測驗工具與分數解釋」。

本研究因應研究需求，調整 PACS-S 指導語，從原本對研究參與者個人衡鑑能力的評估，更改為研究參與者對於「VR 版初次晤談 OSCE」測驗內容的評估，以瞭解參與者對本測驗用於衡鑑能力評估的可行性（以下簡稱為「測驗可行性」）；並同步調整李克特式五點量尺的評量方式，為 1 代表非常不同意，到 5 代表非常同意，分數越高代表使用者認為此測驗用於衡鑑能力評估的可行性越高。

透過上述調整後，PACS-S 將透過「使用者觀點」，評估「VR 版初次晤談 OSCE」的測驗內容（包含 VSP 站與紙筆站簡答題分測驗），用於衡鑑能力評量的可行性，而非自己評估個人在衡鑑能力的專業性。研究參與者在衡鑑能力的專業性，則是透過「試後問卷 C」第二部份「自評衡鑑專業能力」題項，加以評估。而 USVRI 的「VSP 替代性」面向則是依循 Kenny 等人 (2007) 的建議，針對 VSP 系統預設特殊功能評估，著重於 VSP 系統（即，軟體效能的評估），用來瞭解本研究開

發的 VSP 系統中，VSP 替代真人 SP 用於衡鑑能力評量的可行性。

(三) 研究結果

本研究擬透過研究參與者的「使用者觀點」，評估 VR 技術在衡鑑專業能力評量的可行性，以及 VSP 系統的設計品質／使用性。在進行主要分析之前，本研究擬先對於上述分析所使用的兩項測量工具「USVRI」與「PACS-S」，進行信效度評估來檢驗測量品質。因此，研究結果呈現將分成下列三部分：「USVRI 與 PACS-S 的信效度分析」、「VSP 系統使用性評估」與「VR 版初次晤談 OSCE 在衡鑑能力評估的可行性」。

1. USVRI 與 PACS-S 的信效度分析

(1) USVRI 的信效度評估。本研究使用「虛擬實境使用性量表」(USVRI)，瞭解研究參與者對本研究開發之 VSP 系統的使用性評估。USVRI 包含「VSP 系統設計」、「VR 穿戴設備」與「VSP 替代性」三個面向，其中「VR 穿戴設備」面向僅一題採用李克特式五點量尺，故不計算此面向的信度係數。以 DeVellis (1991) 建議之新發展測驗的信度係數下限 ($\alpha = .70$) 為標準，「VSP 系統設計」與「VSP 替代性」的內部一致性信度係數 (Cronbach's α) 分別是 .71 與 .75，代表達到信度係數下限，有一定的內部一致性信度。「使用者介面」與「語音辨識」評估指標的 Cronbach's α 值，分別是 .89 與 .79，具良好的內部一致性信度。「VSP 互動性」指標受到正向題與反向題間關聯性接近零相關 ($r = .05$) 所影響，不計算信度係數，後續分析僅使用互動性正向題資料，進行分析。「求助功能」與「VSP 擬真性」指標都只有一題，不計算信度係數。未來可增加題項數量，來提高 USVRI 的內部一致性信度。

透過 Pearson 相關分析、顯著水準設為 .05，檢驗學生群體在 VSP 替代性、VR 版初次晤談 OSCE 用於衡鑑能力評量的可行性、自評衡鑑專業能力與一年全職實習經驗的關聯性 (見表 3)。研究結果顯示，認為 VSP 替代真人 SP 可行性越高者，傾向認為「VR 版初次晤談 OSCE」在衡鑑能力評量應用的可行性越高 ($r = .75$)，反映測量工具的聚斂效度。但 VSP 替代性與自評專業能力 ($r = -.18$)、一年全職實習經驗 ($r = .24$) 之間，皆未呈現顯著的關聯性。

表 3
VSP 替代性、測驗可行性、專業性與實習經驗之相關矩陣 ($N = 28$)

| | 1 | 2 | 3 | 4 |
|-----------------------|-------|-------|-------|---|
| 1. VSP 替代性 | — | | | |
| 2. 測驗可行性 ^a | .75** | — | | |
| 3. 自評衡鑑專業能力 | -.18 | -.37* | — | |
| 4. 一年全職實習經驗 | .24 | .09 | .51** | — |

^a 測驗可行性為 VR 版初次晤談 OSCE 用於衡鑑能力評量的可行性，以 PACS-S 評估。

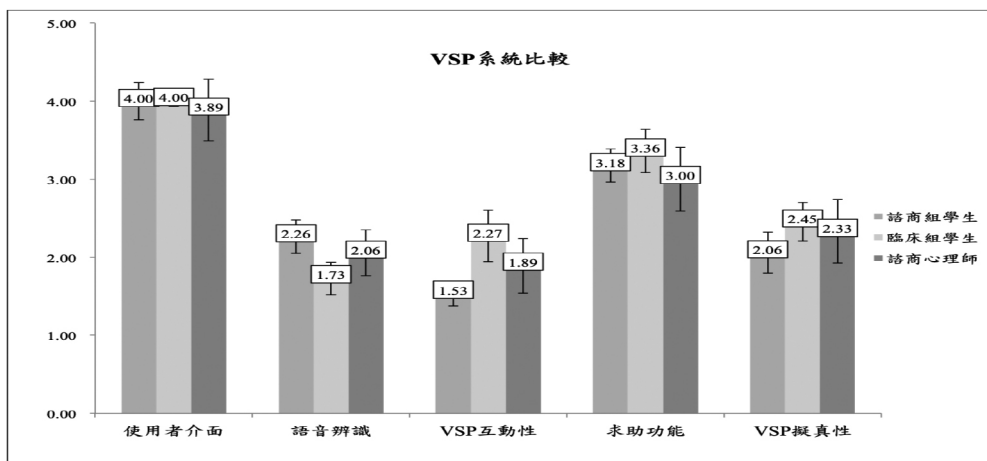
* $p < .05$. ** $p < .01$.

(2) PACS-S 的信效度評估。本研究以「心理衡鑑能力評量表——簡明版」(PACS-S)，測量研究參與者對於「VR 版初次晤談 OSCE」在衡鑑能力評量的可行性評估。此量表的內部一致性信度係數 (Cronbach's α) 為 .75，高於 DeVellis (1991) 建議的信度係數下限 ($\alpha = .70$)，顯示 PACS-S 達到信度係數下限，有一定的內部一致性信度。表 3 的研究結果顯示，認為「VR 版初次晤談 OSCE」可行性越高者，傾向認為 VSP 替代真人 SP 的可行性越高 ($r = .75$)，代表 PACS-S 具有一定的聚斂效度。令人好奇的是，認為「VR 版初次晤談 OSCE」在衡鑑能力評估可行性越高者，在自評專業能力時，卻傾向給自己較低的評價 ($r = -.37$)。

2. VSP 系統使用性評估

(1) **VSP 系統設計**。透過單因子變異數分析、顯著水準設為 .05，檢驗不同組研究參與者在五個「VSP 系統設計」評估指標的差異性。圖 5 為三組研究參與者在「VSP 系統設計」五個指標分數摘要圖。三組參與者在「使用者介面」($F(2, 34) = 0.05, p = .95$)、「語音辨識」($F(2, 34) = 1.40, p = .26$)、「VSP 互動性」($F(2, 34) = 2.32, p = .11$)、「求助功能」($F(2, 34) = 0.34, p = .71$)與「VSP 擬真性」($F(2, 34) = 0.52, p = .60$)的平均分數，都未呈現顯著差異。由圖 5 及檢定結果可知，來自不同訓練背景、不同專業發展階段的參與者在「VSP 系統設計」五個指標的使用性評估，並未呈現顯著差異。從使用者觀點而言，他們對於「使用者介面」與「求助功能」的評價最高，認為本系統在「語音辨識」與「VSP 互動性」設計最需要改善。

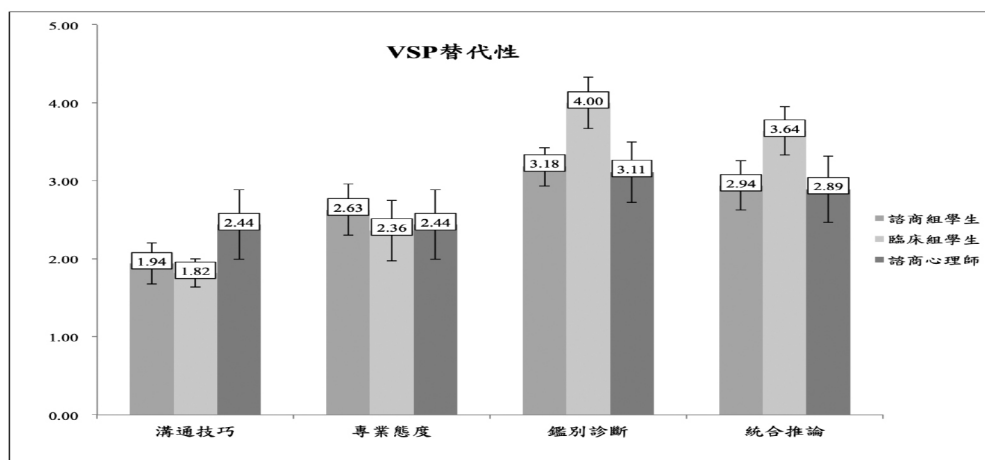
圖 5
三組研究參與者在「VSP 系統設計」五個次面向分數摘要圖



(2) **VR 穿戴設備**。透過單因子變異數分析、顯著水準為 .05，檢驗不同組研究參與者對於「VR 穿戴設備」的舒適度評分，是否呈現顯著差異。三組參與者在「VR 穿戴設備」的舒適度評分 ($F(2, 34) = 1.67, p = .20$)，並未呈現顯著差異；其平均分數為 3.08 ($SD = 1.01$)。在 VSP 站施測後的暈眩不適感評估方面，大部份研究參與者 (91.90%) 並未感到暈眩不適；有不適感者僅 3 位，佔研究參與者總數的 8.10%。在「VSP 站與紙筆站 1」間的休息時間安排，參與者建議的平均休息時間為 5.64 分鐘 ($SD = 3.71$)。

(3) **VSP 替代性**。透過單因子變異數分析、顯著水準設 .05，檢驗不同組研究參與者在「VSP 替代性」評估，是否呈現顯著差異。三組參與者在「溝通技巧」($F(2, 34) = 1.00, p = .38$)、「專業態度」($F(2, 33) = 0.14, p = .87$)、「鑑別診斷」($F(2, 34) = 2.40, p = .11$)、「統合推論」($F(2, 34) = 1.33, p = .28$) 分數，皆未呈現顯著差異。他們對於 VSP 替代真人 SP 在衡鑑能力評量的可行性，由高至低依序為：鑑別診斷、統合推論、專業態度與溝通技巧。圖 6 為三組研究參與者在「VSP 替代性」各題項分數摘要圖。

圖 6
三組研究參與者在「VSP 替代性」各題項分數摘要圖



2. 「VR 版初次晤談 OSCE」在衡鑑能力評估的可行性

透過單因子變異數分析、顯著水準設為 .05，檢驗不同組研究參與者對於「VR 版初次晤談 OSCE」在衡鑑能力評估的可行性，是否呈現顯著差異。研究結果顯示，三組參與者在「VR 版初次晤談 OSCE」用於「鑑別診斷」($F(2, 34) = 5.29, p < .05, \eta^2 = .24$)與「統合推論」($F(2, 34) = 5.20, p < .05, \eta^2 = .23$)的可行性評估，呈現顯著差異，但在其他四種衡鑑能力的可行性評估，則未呈現顯著差異。

透過 Scheffé 法、顯著水準設為 .05 進行事後檢定，研究結果顯示，臨床組學生認為「VR 版初次晤談 OSCE」在「鑑別診斷」($M = 4.09$)與「統合推論」($M = 3.91$)能力評估的可行性，明顯高於諮商組學生(鑑別診斷： $M = 3.00$ ，統合推論： $M = 2.76$)，但與諮商心理師(鑑別診斷： $M = 3.22$ ，統合推論： $M = 3.11$)比較時，兩組的平均數並未呈現顯著差異。表 4 為三組研究參與者對於「VR 版初次晤談 OSCE」在衡鑑能力評估的可行性分析摘要表。

表 4
VR 版初次晤談 OSCE 在衡鑑能力評估的可行性分析摘要表

| PACS-S 題項內容 | 1. 諮商組學生 | | 2. 臨床組學生 | | 3. 諮商心理師 | | F | η^2 | Scheffé 法 |
|----------------|----------|------|----------|------|----------|------|-------|----------|-----------|
| | M | SD | M | SD | M | SD | | | |
| 專業關係 | 2.06 | 1.03 | 1.82 | 0.41 | 2.11 | 0.78 | 0.39 | | |
| 溝通技巧 | 2.29 | 0.99 | 2.27 | 1.01 | 2.44 | 1.01 | 0.09 | | |
| 專業態度 | 2.59 | 1.06 | 2.82 | 0.87 | 2.56 | 1.01 | 0.23 | | |
| 鑑別診斷 | 3.00 | 1.06 | 4.09 | 0.30 | 3.22 | 0.97 | 5.29* | 0.24 | 2 > 1* |
| 統合推論 | 2.76 | 1.03 | 3.91 | 0.70 | 3.11 | 0.93 | 5.20* | 0.23 | 2 > 1* |
| 測驗解釋 | 3.41 | 1.06 | 3.82 | 0.75 | 3.33 | 0.71 | 0.91 | | |

* $p < .05$.

綜合討論

本研究已完成「VR 版初次晤談 OSCE」的初期開發；並透過正式研究，招募研究參與者，由他們以「使用者觀點」，評估 VR 技術在衡鑑能力評量的可行性，以及本研究開發的 VSP 系統的設計品質。

（一）「VR 版初次晤談 OSCE」的初期開發

就 VSP 系統（即，VR 程式的軟體設計）而言，本研究已達到 Matsumura 等人（2018）對於未來 VSP 系統開發的建議，使用 ASR 技術，讓使用者透過「自由問句」詢問模式，與 VSP 進行互動，來模擬真實臨床情境的初次晤談過程。由於本研究所開發的 VSP 系統仍在初期開發階段，不論「語音辨識」的正確率，或是「語意辨識」的正確率都有待提昇，這也影響使用者對本系統在「語音辨識」與「VSP 互動性」功能的評估。

如同 Burdea 與 Coiffet（2003）所言，虛擬實境設計具有良好的「沉浸性」與「互動性」後，將可促發使用者的「想像力」，讓他們產生「身歷其境、信以為真」的知覺感受。因此，作者希望透過下一階段的程式優化工作，強化本系統的「語音辨識」與「VSP 互動性」功能，以提昇本系統在「互動性」的設計品質，讓使用者更容易產生「身歷其境」的知覺感受，並達到替代真人 SP 的功能。

（二）從「使用者觀點」評估 VR 技術在衡鑑能力評量的可行性

就 VSP 替代性而言，受試學生與心理師認為 VSP 在「鑑別診斷」（ $M=3.41$ ）與「統合推論」（ $M=3.14$ ）面向，替代真人 SP，進行衡鑑能力評估的可行性較高；但其在「溝通技巧」（ $M=2.03$ ）與「專業態度」（ $M=2.50$ ）面向的可行性較低。雖然本研究在 VSP 個案腳本設計，已依循 Shah 等人（2012）建議，加入同理心與敏感議題對白，以測量考生在初次晤談的專業態度與溝通技巧。但，或許受到本研究開發之 VSP 系統在「語音辨識」與「VSP 互動性」功能，尚未發展成熟所影響，受試學生與心理師對於 VSP 替代真人 SP，用於「溝通技巧」與「專業態度」面向的衡鑑能力評估，持保留態度。與「溝通技巧」和「專業態度」面向相比，受試學生與心理師較贊同使用 VSP 替代真人 SP，進行「鑑別診斷」與「統合推論」面向的衡鑑能力評量。換言之，受試學生與心理師較認同以 VSP 替代真人 SP，進行「鑑別診斷」與「統合推論」的衡鑑能力評估，這兩個面向需要較多專業知識的統合，比較偏重於認知層面的能力評估；但他們對於 VSP 替代真人 SP，用於「溝通技巧」與「專業態度」的衡鑑能力評估，持保留態度，而這兩個面向需要展現個人和 VSP 互動的溝通能力與人機互動時的專業態度，比較屬於態度與技術層面的能力評估。

除了受到本系統尚未發展成熟所影響，受試學生與心理師對 VSP 系統在專業態度與溝通技巧面向衡鑑能力評估的疑慮，或許與個人接觸 VR 技術的經驗值有關。如同 Mucic 與 Hilty（2018）所言，遠距精神醫療已是時代趨勢。不論系所主管、教授或是醫學院學生對於遠距精神醫療的態度，往往與個人使用遠距科技的經驗值有關。Mucic 與 Hilty 建議，增加精神醫學相關從業人員在遠距精神醫學的使用經驗，可提高他們未來把遠距精神醫學納入臨床執業選項的意願，並同時降低他們認為遠距精神醫學無法有效地從事臨床照護的疑慮（不易判讀非語言訊息、難以因應緊急狀態）。後續研究可進一步探索個人 VR 使用經驗對於 VR 技術在衡鑑能力評估應用的影響。

（三）從使用者觀點評估 VR 版初次晤談 OSCE 在衡鑑能力評量之應用

就「測驗可行性」而言，研究參與者對於「VR 版初次晤談 OSCE」在六種衡鑑能力評量的可行性評估，依分數高低，由高至低的排序結果為：「測驗解釋」、「鑑別診斷」、「統合推論」、「專業態度」、「溝通技巧」與「專業關係」。此研究結果反映，受試學生與心理師認為「VR 版初次晤談 OSCE」用於前三種（較偏重認知層面）衡鑑能力評估的可行性較高，但對其用於後三種（較偏重態度與技術層面）衡鑑能力評估的可行性較低，有較多的疑慮。

令人好奇的是，認為「VR 版初次晤談 OSCE」在衡鑑能力評估可行性越高者，在自評衡鑑專業能力時，卻傾向給自己較低的評價（ $r = -.37$ ）。或是，換個角度來看，自評衡鑑能力越高者，認為「VR 版初次晤談 OSCE」在衡鑑能力評估的可行性越低。這些對自己的衡鑑能力較有信心的受試學生們，或許對於 VSP 的期待較高，認為 VSP 無法充份模擬真人 SP 所呈現的臨床情境，尤其是「溝通技巧」與「專業態度」面向，以致比較不認同「VR 版初次晤談 OSCE」用於衡鑑能力評估的可行性。

然而，如同林家興等人（2015）與許雅涵等人（2016）提到自陳量表在能力評估可能的潛在測量問題（如，仁慈偏誤／謙虛偏誤）一樣，受試學生對自己衡鑑能力的評估，以及他們對「VR 版初次晤談 OSCE」用於衡鑑能力評量的可行性評估，也可能出現上述的測量問題，這也是本研究的研究限制之一。透過使用者經驗所提出的評估結果，與測驗實際表現之間的一致性有多高，值得深思。後續研究將進一步分析研究參與者在「VR 版初次晤談 OSCE」的測驗結果，若心理師群體在此測驗的表現明顯優於學生群體，或許可說明此測驗從團體差異性而言，具有構念效度，用於衡鑑能力評量有其可行性。

（四）研究限制與未來發展

1. 本研究的研究限制有四

（1）USVRI 的內部一致性信度仍有待改善。在本研究樣本數較小（ $N = 37$ ）的限制下，USVRI 僅在「使用者介面」與「語音辨識」呈現信度係數 0.80 左右，其他面向／評估指標的內部一致性信度，有待增加題項內容，以及增加研究參與者的團體變異性來改善。

（2）研究參與者人數不多，且以女性為主。囿於正式研究的研究樣本較小（ $N = 37$ ），且女性佔八成以上。據此研究樣本對 VR 技術在衡鑑能力的可行性評估，僅提出初步驗證結果。待「VR 版初次晤談 OSCE」發展成熟後，將招募更多來自不同專業發展階段、不同背景的受試者，並加強男性參與者的招募，以提供更完整的驗證結果。

（3）僅以「使用者觀點」評估測驗可行性。本研究僅以「使用者觀點」評估 VR 技術用於衡鑑能力評量的可行性，待「VR 版初次晤談 OSCE」發展成熟後，擬進一步透過實徵研究檢驗此測驗在衡鑑能力評量的信效度，以再次檢視 VR 技術用於衡鑑能力評量的可行性。

（4）施測時間規劃安排與設備提昇。從模擬考生與研究參與者的回饋可知，未來使用「VR 版初次晤談 OSCE」進行衡鑑能力評量時，需注意施測時間的規劃安排（如，提供充足的休息時間）。除了施測時間的規劃安排之外，Saredakis 等人（2020）透過統合分析，發現影響暈眩感的主要因素是「VR 的呈現內容」，當呈現內容屬於靜態風景或不動物件時，使用者較不易感到暈眩不適；但是，當呈現內容屬於動態遊戲或互動性較高的情境時，使用者較容易產生暈眩不適感。未來進一步強化 VSP 系統的互動性時，也需同時提昇電腦硬體設備，以降低使用者的暈眩不適感。

2. 未來發展

本研究已完成「VR 版初次晤談 OSCE」的初期開發，也透過「使用者觀點」，瞭解受試學生與心理師認為本測驗用於衡鑑能力評量有一定的可行性。在未來發展方面，本文分別從「VR 技術在衡鑑能力評量應用」、「VR 技術用於衡鑑能力訓練」以及「VSP 系統的未來開發」三個方向，逐一說明。

（1）VR 技術在衡鑑能力評量應用。未來本研究團隊擬從「精神疾患」與「臨床情境」面向，進一步擴充「VR 版初次晤談 OSCE」在 VSP 個案的多樣性。在精神疾患面向，擬從已完成的憂鬱症 VSP 個案，拓展至焦慮症、思覺失調症等不同類型精神疾患的 VSP 個案，更全面地評估受試學生與心理師面對不同類精神疾患個案是否具有有良好的鑑別診斷能力、溝通技巧與專業態度。在臨床情境面向，本研究團隊擬針對較敏感、不易訓練真人扮演，或是真人扮演風險較高的臨床情境（如，性侵、兒虐、自殺），開發特殊臨床情境的 VSP 個案，評估考生面對這些特殊臨床情境的衡鑑專

業知能。

待測驗發展成熟後，作者擬透過實徵研究檢驗「VR 版初次晤談 OSCE」在衡鑑能力評量的信效度，並評估其在施測成本節約與施測效能提昇的效果。為招募較大規模的研究樣本來驗證上述議題，作者擬與多個專業心理團體、醫療院所、臨床／諮商心理相關系所，洽談聯合教育訓練或跨機構研究合作模式，拓展受試者招募管道，以招募來自不同機構、不同專業背景與不同專業發展階段的碩士班學生、全職實習生和心理師參與研究，並同時加強男性參與者的招募，來提昇研究樣本的代表性與推論性。

在測量工具的品質提昇方面，就 USVRI 而言，作者擬透過增加題數、增加研究樣本的人數與團體變異性，來改善信度水準。就 VR 版初次晤談 OSCE 而言，作者擬透過下一階段 VR 程式的優化工作，強化 VSP 系統在「語音辨識」與「VSP 互動性」的效能，進而改善其用於「溝通技巧與專業態度」的測量品質。在施測時間規劃與硬體設備方面，從模擬考生與研究參與者的回饋可知，未來使用 VR 設備進行衡鑑能力評量時，需注意施測時間的規劃安排。除了提供考生充足的休息時間之外，待 VSP 系統的互動性再次提昇時，需要同步搭配更先進的 VR 設備（如，具有更先進眼球追蹤技術的 VR 設備）與含高階顯示卡的電腦設備，來提高虛擬情境的解析度，以降低虛擬情境與真實世界的視覺差異，進而降低受試者的暈眩不適感。

(2) VR 技術用於衡鑑能力訓練。精神醫學領域已將 VSP 用於憂鬱症 (Shah et al., 2012)、自殺風險評估 (Foster et al., 2015)、行為規範障礙症 (Parsons et al., 2008) 與失智症 (Matsumura et al., 2018) 病患的臨床實務能力訓練，評估醫學院學生在鑑別診斷、溝通技巧與治療處遇的專業知能。心理相關系所與實習督導機構也可參考精神醫學模式，依據衡鑑課程的教學內容，使用本研究團隊開發之「VR 版初次晤談 OSCE」或是自行開發 VSP 系統，將不同病症、不同臨床情境的 VSP 個案融入教學現場，輔助課堂上衡鑑能力訓練的解說與實作演練活動。

同樣的，教師也可依照教學單元，在 VSP 系統中，建置形成性評量與總結性評量，或是提供課後實作練習作業，讓學生與 VSP 互動、反覆練習，來精進個人專業知能。未來有志投入此議題的學者專家與研究團隊，更可致力於心理治療、專業倫理等能力訓練為主的 VSP 系統開發工作，將 VR 技術更廣泛地應用於各種心理師能力養成的教育訓練中。

(3) VSP 系統的未來開發。本研究在科技公司協助下，成功開發出具語音辨識功能的 VSP 系統，讓使用者透過語音輸入方式，與 VSP 進行初次晤談工作。未來在 VSP 系統的功能開發方面，可參考 Matsumura 等人 (2018) 的 VSP 系統之「醫師回饋」功能，在本研究團隊所開發的 VSP 系統（即，「VR 心理衡鑑室」）中，增加「虛擬教師／督導」的功能設計。此功能擬將教師／督導的評分機制與意見回饋，內建於 VSP 系統中，讓學生與 VSP 每完成一次能力評量或實務能力訓練後，立即透過 VSP 系統內建的「虛擬教師／督導」功能，獲得即時、客觀的評分結果與意見回饋。透過這些即時評分與意見回饋，學生可以迅速調整、改善，再與 VSP 展開下一輪互動，並透過「反覆練習——意見回饋——調整修正」的學習歷程，強化學習成效、提昇個人專業知能。學校教師與機構督導也可藉由「虛擬教師／督導」功能，減輕臨床實務訓練的評分工作與一部份工作負擔。

參考文獻

- 心理師法 (2020 年 1 月 15 日) 修正公布。 <https://law.moj.gov.tw/LawClass/LawAll.aspx?pcode=L0020098>
[Psychologists Act. (2020, January 15). Amendment to Articles. <https://law.moj.gov.tw/LawClass/LawAll.aspx?pcode=L0020098>]
- 心理師法施行細則 (2011 年 6 月 28 日) 修正公布。 <https://law.moj.gov.tw/LawClass/LawAll.aspx?pcode=L0020101> [Enforcement Rules of the Psychologists' Act. (2011, June 28). Amendment to Articles. <https://law.moj.gov.tw/LawClass/LawAll.aspx?pcode=L0020101>]

- 吳英璋、許文耀、翁嘉英（計畫主持人）（2001）：《臨床心理師現況分析與未來需求研究》（計畫編號：DOH89-TD-1185）。行政院衛生署八十八年度委託研究計劃成果報告，衛生署。<https://www.cdc.gov.tw/uploads/files/88db6c59-335b-4dd3-9c04-59cc3b3e69ca.pdf> [Wu, E.-C., Hsu, W.-Y., & Weng, J.-Y. (Principal Investigator). (2001). *On the social and professional demand of clinical psychology manpower: Present and future* (No. DOH89-TD-1185) (Grant). Ministry of Health and Welfare. <https://www.cdc.gov.tw/uploads/files/88db6c59-335b-4dd3-9c04-59cc3b3e69ca.pdf>]
- 李喬芳、曾芳美、盧煜煬（2018）：〈主要領域虛擬實境學術論文之發展軌跡〉。《科技管理學刊》，23（2），61–96。[Lee, C.-F., Tseng, F.-M., & Lu, L. Y. Y. (2018). The development trajectory of virtual reality literature in main areas. *Journal of Technology Management*, 23(2), 61–96.]
- 李選、顏文娟、張婷、王郁琇（2015）：〈推動護理臨床技能測驗（OSCE）納入國家專技考試應考資格〉。《國家菁英》，11（2），119–137。[Lee, S., Yen, W.-J., Chang, T., & Wang, Y.-H. (2015). Promoting nursing OSCE as requisite qualification in national examination for professional technical personnel. *National Elite*, 11(2), 119–137.]
- 沈煜棠（2015）：〈客觀結構式臨床技能考試（OSCE）應用在諮商心理師訓練及評量之初探〉。《諮商與輔導》，349，51–54。[Shen, Y.-T. (2015). The initial investigation of using Objective Structured Clinical Examination in counseling psychologists' training and evaluation. *Counseling & Guidance*, 349, 51–54.]
- 林丞增、林家興（2019）：〈諮商心理師全職實習與專業督導的現況調查研究〉。《教育心理學報》，51（2），275–295。[Lin, C.-T., & Lin, C.-H. (2019). Survey on the current internship training and supervision in counseling psychology. *Bulletin of Educational Psychology*, 51(2), 275–295.] [https://doi.org/10.6251/BEP.201912_51\(2\).0005](https://doi.org/10.6251/BEP.201912_51(2).0005)
- 林家興（2013）：〈心理師的核心能力與人力培育〉。《人文與社會科學簡訊》，14（2），87–92。[Lin, C.-H. (2013). Core competencies and human development for psychologists in Taiwan. *Humanities and Social Sciences Newsletter Quarterly*, 14(2), 87–92.]
- 林家興、林旻沛、黃佩娟、胡薇瑄、江信男（2015）：〈心理師能力評量表之編製及其信效度分析〉。《教育心理學報》，47（1），63–86。[Lin, C.-H., Lin, M.-P., Huang, P.-C., Hu, W.-H., & Jiang, S.-N. (2015). The construction, reliability and validity study of psychologist competency rating scale. *Bulletin of Educational Psychology*, 47(1), 63–86.] <https://doi.org/10.6251/BEP.20141008>
- 林家興、黃佩娟（2013）：〈臺灣諮商心理師能力指標建構之共識研究〉。《教育心理學報》，44（3），735–750。[Lin, C.-H., & Huang, P.-C. (2013). Development of competency benchmarks for counseling psychologists in Taiwan. *Bulletin of Educational Psychology*, 44(3), 735–750.] <https://doi.org/10.6251/BEP.2012-1003>
- 高聖博（2013）：《精通標準化病人》。力大圖書有限公司。[Kao, S.-P. (2013). *Mastering standardized patients*. Leader Book.]
- 許文耀、張素凰（2008）：〈臨床心理師考試教考用之探討〉。《國家菁英》，4（4），51–82。[Hsu, W.-Y., & Chang, S.-W. (2008). About the teaching, examination and usage of clinical psychologist.

National Elite, 4(4), 51–82.]

- 許育光、刑志彬（2019）：〈臺灣學校心理師何去何從？—從現況評述、課程檢核到培育反思〉。《當代教育研究季刊》，27（3），35–64。[Hsu, Y. K., & Hsing, C. P. (2019). Phenomenon review, curriculum investigation, and reflections on the professional training of school psychologist in Taiwan. *Contemporary Educational Research Quarterly*, 27(3), 35–64.] [https://doi.org/10.6151/CERQ.201909_27\(3\).0002](https://doi.org/10.6151/CERQ.201909_27(3).0002)
- 許秋田、黃創華、鄭志強（2009）：〈臺灣臨床心理師執業與工作現況調查〉。《應用心理研究》，41，43–55。[Hsu, C.-T., Huang, C.-H., & Cheng, C.-C. (2009). Practice survey of clinical psychologists in Taiwan. *Research in Applied Psychology*, 41, 43–55.]
- 許雅涵、藍玉玲、陳曉蘭（2016）：〈從實習心理師與督導觀點看臺灣實習臨床心理師衡鑑專業能力〉。《教育心理學報》，47（3），355–372。[Hsu, Y.-H., Lan, Y.-L., & Chen, W.-L. (2016). An investigation of core competencies in psychological assessment: Comparing the evaluation between interns and supervisors. *Bulletin of Educational Psychology*, 47(3), 355–372.] <https://doi.org/10.6251/BEP.20150310>
- 專門職業及技術人員高等考試心理師考試規則（2018年12月11日）修正公布。<https://law.moj.gov.tw/LawClass/LawAll.aspx?pcode=R0040037> [Senior Professional and Technical Examinations Regulations for Psychologists. (2018, December 11). Amendment to Articles. <https://law.moj.gov.tw/LawClass/LawAll.aspx?pcode=R0040037>]
- 陳皎眉、孫旻暉、楊芷頤（2015）：〈客觀結構式臨床試驗納入心理師養成教育與國家考試之可行性〉。《國家菁英》，44（11），3–20。[Chen, J.-M., Suen, M.-W., & Yang, T.-Y. (2015). The feasibility of using Objective Structured Clinical Examination in clinical and counseling psychologists' education and national license examination. *National Elite*, 44(11), 3–20.]
- 黃政昌（2008）：〈諮商中需要的心理評估（Psychological Assessment）：談臨床心理衡鑑 VS. 諮商心理評估〉。《諮商與輔導》，265，38–41。[Huang, C.-C. (2008). Psychological assessment: What's the difference between clinical psychology and counseling psychology? *Counseling and Guidance*, 265, 38–41.]
- 劉克明、曾慧敏（2010）：〈高階客觀結構式臨床測驗（High-Stakes OSCE）在國家考試的應用與展望〉。《國家菁英》，6（3），111–129。[Liu, K.-M., & Tseng, H.-M. (2010). The application of a high-stakes OSCE in a national medical licensure examination. *National Elite*, 6(3), 111–129.]
- 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（2），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(2), 118–132.]
- 盧怡蓓、姚珮華、呂奕熹（2017）：〈應用 OSCE 於諮商心理師訓練評量之實踐與挑戰〉。《醫療品質雜誌》，11（4），67–77。[Lu, Y.-C., Yao, P.-H., & Leu, Y.-S. (2017). Training and evaluation of counseling psychologist: The practice and challenge in applying the OSCE. *Journal of*

- Healthcare Quality*, 11(4), 67–77.]
- 賴崇閔、黃秀美、廖述盛、黃雯雯（2009）：〈3D 虛擬實境應用於醫學教育接受度之研究〉。《教育心理學報》，40(3)，341–361。[Lai, C.-M., Huang, H.-M., Liaw, S.-S., & Huang, W.-W. (2009). A study of user's acceptance on three-dimensional virtual reality applied in medical education. *Bulletin of Educational Psychology*, 40(3), 341–361.] <https://doi.org/10.6251/BEP.20080422>
- 藍玉玲（計畫主持人）（2013）：《初次晤談客觀結構式臨床測驗在臨床心理碩士生之信效度研究》（計畫編號：NSC101-2410-H-259-027）。行政院國家科學委員會專題研究成果報告，行政院國家科學委員會。<https://www.grb.gov.tw/search/planDetail?id=2638090> [Lan, Y.-L. (2013). *Psychometric properties of Objective Structured Clinical Examination (OSCE) to assess intake interview skills among master students in clinical psychology* (No. NSC 101-2410-H-259-027) (Grant). National Science Council in Taiwan. <https://www.grb.gov.tw/search/planDetail?id=2638090>]
- 藍玉玲（計畫主持人）（2021）：《虛擬實境技術在心理衡鑑能力評估之應用》（計畫編號：MOST108-2410-H-259-015）。科技部補助研究計畫成果報告，科技部。<https://www.grb.gov.tw/search/planDetail?id=13139868> [Lan, Y.-L. (Principal Investigator). (2021). *The use of virtual reality in measuring psychological assessment* (Report No. MOST 107-2511-H003-003-MY3) (Grant). Ministry of Science and Technology. <https://www.grb.gov.tw/search/planDetail?id=13139868>]
- 藍玉玲、張玉萱、陳曉蘭（2015）：〈「心理衡鑑初次晤談」客觀結構式臨床測驗之初探研究〉。《中華輔導與諮商學報》，44，133–158。[Lan, Y.-L., Chang, Y.-H., & Chen, W.-L. (2015). Development and initial validation of the intake Objective Structured Clinical Examination. *Chinese Journal of Guidance and Counseling*, 44, 133–158.]
- 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(2)，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(2), 209–245.] [https://doi.org/10.30074/FJMH.201106_24\(2\).0003](https://doi.org/10.30074/FJMH.201106_24(2).0003)
- American Psychological Association. (2009). *Guidelines and principles for accreditation of programs in professional psychology*.
- Burdea, G. C., & Coiffet, P. (2003). *Virtual reality technology*. John Wiley & Sons.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046.016>
- Day, C., Barker, C., Bell, E., Sefcik, E., & Flournoy, D. (2018). Flipping the Objective Structured Clinical Examination: A teaching innovation in graduate nursing education. *Nurse Educator*, 43(2), 83–86. <https://doi.org/10.1097/nne.0.0000000000000421>
- DeVellis, R. F. (1991). *Scale development: Theory and applications*. SAGE Publications.
- Fertleman, C., Aubugeau-Williams, P., Sher, C., Lim, A.-N., Lumley, S., Delacroix, S., & Pan, X. (2018). A

- discussion of virtual reality as a new tool for training healthcare professionals. *Frontiers in Public Health*, 6, Article 44. <https://doi.org/10.3389/fpubh.2018.00044>
- Fodor, L. A., Cotet, C. D., Cuijpers, P., Szamoskozi, S., David, D., & Cristea, I. A. (2018). The effectiveness of virtual reality based interventions for symptoms of anxiety and depression: A meta-analysis. *Scientific Reports*, 8, Article 10323. <https://doi.org/10.1038/s41598-018-28113-6>
- Foster, A., Chaudhary, N., Murphy, J., Lok, B., Waller, J., & Buckley, P. F. (2015). The use of simulation to teach suicide risk assessment to health profession trainees-rationale, methodology, and a proof of concept demonstration with a virtual patient. *Academic Psychiatry*, 39(6), 620–629. <https://doi.org/10.1007/s40596-014-0185-9>
- Gibbons, R. D., Weiss, D. J., Pilkonis, P. A., Frank, E., Moore, T., Kim, J. B., & Kupfer, D. J. (2012). Development of a computerized adaptive test for depression. *Archives of General Psychiatry*, 69(11), 1104–1112. <https://doi.org/10.1001/archgenpsychiatry.2012.14>
- Hodges, B. D. (2006). The Objective Structured Clinical Examination: Three decades of development. *Journal of Veterinary Medical Education*, 33(4), 571–577. <https://doi.org/10.3138/jvme.33.4.571>
- Kaslow, N. J. (2004). Competencies in professional psychology. *American Psychologist*, 59(8), 774–781. <https://doi.org/10.1037/0003-066X.59.8.774>
- Kaslow, N. J., Rubin, N. J., Bebeau, M. J., Leigh, I. W., Lichtenberg, J. W., Nelson, P. D., Portnoy, S. M., & Smith, I. L. (2007). Guiding principles and recommendations for the assessment of competence. *Professional Psychology: Research and Practice*, 38(5), 441–451. <https://doi.org/10.1037/0735-7028.38.5.441>
- Kenny, P., Parsons, T. D., Gratch, J., Leuski, A., & Rizzo, A. A. (2007). Virtual patients for clinical therapist skills training. In C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, & D. Pelé (Eds.), *Intelligent virtual agents: 7th International Working Conference, IVA 2007, proceedings* (Vol. 4722, pp. 197–210). Springer Publishing Company. https://doi.org/10.1007/978-3-540-74997-4_19
- Kononowicz, A. A., Woodham, L. A., Edelbring, S., Stathakarou, N., Davies, D., Saxena, N., Car, L. T., Carlstedt-Duke, J., Car, J., & Zary, N. (2019). Virtual patient simulations in health professions education: Systematic review and meta-analysis by the digital health education collaboration. *Journal of Medical Internet Research*, 21(7), Article e14676. <https://doi.org/10.2196/14676>
- Krishnamurthy, R., VandeCreek, L., Kaslow, N. J., Tazeau, Y. N., Miville, M. L., Kerns, R., & Benton, S. A. (2004). Achieving competency in psychological assessment: Directions for education and training. *Journal of Clinical Psychology*, 60(7), 725–739. <https://doi.org/10.1002/jclp.20010>
- Lan, Y.-L., & Chang, Y. (2016). Development and initial psychometrics of the Psychological Assessment Competency Scale. *Training and Education in Professional Psychology*, 10(2), 93–101. <https://doi.org/10.1037/tep0000108>
- Leigh, I. W., Smith, I. L., Bebeau, M. J., Lichtenberg, J. W., Nelson, P. D., Portnoy, S., Rubin, N. J., & Kaslow, N. J. (2007). Competency assessment models. *Professional Psychology: Research and Practice*, 38(5), 463–473. <https://doi.org/10.1037/0735-7028.38.5.463>

- Lu, J., Pan, Z., Lin, H., Zhang, M., & Shi, J. (2005). Virtual learning environment for medical education based on VRML and VTK. *Computers & Graphics, 29*(2), 283–288. <https://doi.org/10.1016/j.cag.2004.12.015>
- Maloney, M. P., & Ward, M. P. (1976). *Psychological assessment: A conceptual approach*. Oxford University Press.
- Matsumura, Y., Shinno, H., Mori, T., & Nakamura, Y. (2018). Simulating clinical psychiatry for medical students: a comprehensive clinic simulator with virtual patients and an electronic medical record system. *Academic Psychiatry, 42*(5), 613–621. <https://doi.org/10.1007/s40596-017-0860-8>
- Meghani, D. T., & Ferm, B. R. (2021). Development of a standardized patient evaluation exam: An innovative model for health service psychology programs. *Training and Education in Professional Psychology, Training and Education in Professional Psychology, 15*(1), 37–44. <https://doi.org/10.1037/tep0000291>
- Melluish, S., Crossley, J., & Tweed, A. (2007). An evaluation of the use of simulated patient role-plays in the teaching and assessment of clinical consultation skills in clinical psychologists' training. *Psychology Learning and Teaching, 6*(2), 104–113. <https://doi.org/10.2304/plat.2007.6.2.104>
- Mucic, D., & Hilty, D. (2018). Training in telepsychiatry. In M. H. M. Hermans, T. C. Hoon, & E. Pi (Eds.), *Education about mental health and illness* (pp. 1–31). Springer Publishing Company. https://doi.org/10.1007/978-981-10-0866-5_26-1
- Navas-Ferrer, C., Urcola-Pardo, F., Subiron-Valera, A. B., & German-Bes, C. (2017). Validity and reliability of Objective Structured Clinical Evaluation in nursing. *Clinical Simulation in Nursing, 13*(11), 531–543. <https://doi.org/10.1016/j.ecns.2017.07.003>
- Norcross J. C., & Karpiak, C. P. (2012). Clinical psychologists in the 2010s: 50 years of the APA division of clinical psychology. *Clinical Psychology: Science and Practice, 19*(1), 1–12. <https://doi.org/10.1111/j.1468-2850.2012.01269.x>
- Parsons, T. D., Kenny, P., Ntuen, C. A., Pataki, C. S., Pato, M. T., Rizzo, A. A., St-George, C., & Sugar, J. (2008). Objective structured clinical interview training using a virtual human patient. *Studies in Health Technology and Informatics, 132*, 357–362.
- Powers, M. B., & Emmelkamp, P. M. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders, 22*(3), 561–569. <https://doi.org/10.1016/j.janxdis.2007.04.006>
- Rodolfa, E., Bent, R. J., Eisman, E., Nelson, P. D., Rehm, L., & Ritchie, P. (2005). A cube model for competency development: Implications for psychology educators and regulators. *Professional Psychology: Research and Practice, 36*(4), 347–354. <https://doi.org/10.1037/0735-7028.36.4.347>
- Rothbaum, B. O., Anderson, P., Zimand, E., Hodges, L., Lang, D., & Wilson, J. (2006). Virtual reality exposure therapy and standard (in vivo) exposure therapy in the treatment of fear of flying. *Behavior Therapy, 37*(1), 80–90. <https://doi.org/10.1016/j.beth.2005.04.004>
- Rubin, N. J., Bebeau, M., Leigh, I. W., Lichtenberg, J. W., Nelson, P. D., Portnoy, S., Smith, I. L., &

- Kaslow, N. J. (2007). The competency movement within psychology: An historical perspective. *Professional Psychology: Research and Practice, 38*(5), 452–462. <https://doi.org/10.1037/0735-7028.38.5.452>
- Rus-Calafell, M., Garety, P., Sason, E., Craig, T. J. K., & Valmaggia, L. R. (2018). Virtual reality in the assessment and treatment of psychosis: A systematic review of its utility, acceptability and effectiveness. *Psychological Medicine, 48*(3), 362–391. <https://doi.org/10.1017/s0033291717001945>
- Rushforth, H. E. (2007). Objective Structured Clinical Examination (OSCE): Review of literature and implications for nursing education. *Nurse Education Today, 27*, 481–490. <https://doi.org/10.1016/j.nedt.2006.08009>
- Saredakis, D., Szpak, A., Birkhead, B., Keage, H., Rizzo, A., & Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis. *Frontiers in Human Neuroscience, 14*, Article 96. <https://doi.org/10.3389/fnhum.2020.00096>
- Segal, R., Bhatia, M., & Drapeau, M. (2011). Therapists' perception of benefits and costs of using virtual reality treatments. *Cyberpsychology Behavior and Social Networking, 14*(1–2), 29–34. <https://doi.org/10.1089/cyber.2009.0398>
- Shah, H., Rossen, B., Lok, B., Londino, D., Lind, S. D., & Foster, A. (2012). Interactive virtual-patient scenarios: An evolving tool in psychiatric education. *Academic Psychiatry, 36*(2), 146–150. <https://doi.org/10.1176/appi.ap.100300.49>
- Tsang, M. M. Y., & Man, D. W. K. (2013). A virtual reality-based vocational training system (VRVTS) for people with schizophrenia in vocational rehabilitation. *Schizophrenia Research, 144*(1–3), 51–62. <https://doi.org/10.1016/j.schres.2012.12.024>
- 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. <https://doi.org/10.1111/j.1365-2648.2009.05054.x>
- Yedidia, M. J., Gillespie, C. C., Kachur, E., Schwartz, M. D., Ockene, J., Chepaitis, A. E., Snyder, C. W., Lazare, A., & Lipkin, M. (2003). Effect of communications training on medical student performance. *Journal of the American Medical Association, 290*(9), 1157–1165. <https://doi.org/10.1001/jama.290.9.1157>

收稿日期：2020年10月11日

一稿修訂日期：2020年10月12日

二稿修訂日期：2021年02月04日

三稿修訂日期：2021年03月29日

四稿修訂日期：2021年04月28日

接受刊登日期：2021年04月28日

附錄 VSP 站評分表題項內容

面向一「病史詢問」

1. 壓力的可能來源
2. 目前的情緒狀況（型態、期間）
3. 平日興趣與活動參與狀況的改變
4. 生理狀況評估（睡眠習慣、食慾體重變化）
5. 認知功能影響評估（記憶力、注意力）
6. 個案面對憂鬱症狀影響之因應方式
7. 目前用藥狀況（藥名、服藥次數、成效）
8. 自殺危險評估（意念／行動、頻率、強度）
9. 學業／工作狀況
10. 與家人的關係（父母、手足、姻親）
11. 親密關係或婚姻關係
12. 社會支持系統（教友、社團、朋友、人際關係）
13. 與主訴有關的過去病史
14. 與主訴有關的家族病史
15. 創傷經驗或受虐經驗
16. 過去心理治療經驗【求助專業助人工作者，如，精神科醫師、心理師、諮商中心】
17. 過去其他求助經驗（宗教、算命等非正式心理治療方式）【求助非專業助人工作者，如，老師、導師、牧師、神父】
18. 個案對於治療效果的期待

面向二「溝通技巧與專業態度」

1. 招呼並與向個案自我介紹（姓名、職稱）
2. 提出適當問題來詢問／確認個案的就診目的
3. 適當地回應個案的談話內容
4. 用個案能理解的語言與溝通方式
5. 技巧性地詢問敏感問題（如，自殺與創傷經驗詢問）
6. 計畫性地提問來蒐集個案的重要資訊
7. 簡單摘要本次晤談內容並安排後續晤談事宜

Bulletin of Educational Psychology, 2021, 53(2), 481–510
National Taiwan Normal University, Taipei, Taiwan, R. O. C.

The Feasibility of Virtual Reality in Measuring Psychological Assessment: Student and Psychologist Views

Yu-Ling Lan

Department of Counseling and Clinical Psychology,
National Dong Hwa University

Wan-Lan Chen

Department of Human Development and Psychology,
Tzu Chi University

Yuhuan Chang

College of Management Yuan,
Ze University

Ying-Fen Wang

Department of Educational Psychology and
Counseling, National Taiwan Normal University

Psychological assessment has been regarded an essential competency domain in professional psychology. Influenced by the competency-based movement, researchers and educators start to redesign new measurement tools for assessing psychological assessment competencies (PACs) of psychologists from entry to practice. Compared with physicians and nurses, license examinations of professional psychologists still heavily rely on the paper-pencil tests to measure professional competencies of psychologists. In the past five decades, an objective structured clinical examination (OSCE) has become one of the most commonly used measurement tools to assess professional knowledge, clinical skills and professional attitudes of healthcare professionals at the same time. Several meta-analysis studies have confirmed that an OSCE is not only a reliable and valid measurement tool for assessing clinical competencies of healthcare professionals, but also a good teaching aid to facilitate a wide range of clinical skill training courses.

In Taiwan, OSCEs have been widely used to assess professional competencies of physicians and nursing practitioners, however, its application in professional psychologists is still very limited. Lan et al. (2011) proposed four possible factors affecting the OSCE application in professional psychology. First of all, competencies in professional psychology (e.g., rapport building skills, expressions of empathy) are more abstract and difficult to measure than those in medicine and nursing. Second, it requires longer time and more effort to develop and administer a psychology OSCE than a medical OSCE or a nursing OSCE. Third, faculty members and practicing psychologists need to receive further training (e.g., taking an OSCE rater training workshop) to ensure that they reach inter-rater agreement before scoring student or participant performance on a psychology OSCE. Last but not least, the major obstacle to adopting OSCEs in professional psychology is high cost of running an OSCE on a regular basis. Compared with the resource input of a medical OSCE, the resource input of a psychology OSCE is relatively insufficient. If the cost of running a psychology OSCE could be effectively reduced, it would help to promote the OSCE application in professional psychology.

To overcome the aforementioned problem, the virtual reality (VR) technology was applied in this study to develop a VR version of the Intake OSCE (VR-intake OSCE) to serve this need. The purpose of the present study was to examine the feasibility of the VR technology in assess competencies in psychological assessment through the viewpoints of students and psychologists. The present study comprises three phases: the testing phase 1, the testing phase 2 and the formal study phase.

In the testing phase 1, six psychologists from master education to practice were hired as mock test takers to examine the versatility of the intake OSCE in assessing PACs of psychologists from counseling and clinical psychology. The intake OSCE was developed by Lan et al. (2015) to examine psychologists' PACs when conducting an intake interview. This test consists of a standardized patient test (SP test), a follow-up short answer test (S test) and a multiple-choice test. In the testing phase 1, only the SP test and the S test were selected to examine the versatility of the intake OSCE in assessing PACs of psychologists from counseling and clinical psychology. No significant score differences were found between clinical psychologists and counseling psychologists. This result confirmed that the intake OSCE could be used to assess psychologists in both fields.

In the testing phase 2, the same mock test takers (from the testing phase 1) were hired to evaluate the initial version of the VR-intake OSCE, which consists of a virtual standardized patient test (VSP test) and a follow-up S test. To administer a VSP test on the VR equipment, a new VSP system (called VR psychological assessment room) was developed in the present study to serve this need. By adopting the automatic speech recognition technology, the design of the VSP system allows participants to conduct an intake interview with the VSP through a free questioning format, and collect assessment related information (e.g., personal medical history, stressors, and social support) in order to generate a differential diagnosis of the VSP and also provide an initial treatment plan. All mock test takers first took the VR-intake OSCE, and then filled out the post-test questionnaire. This post-test questionnaire was designed to evaluate the quality of the initial version of the VR-intake OSCE and its test administration procedure.

Responses provided by mock test takers were summarized as the followings. First of all, mock test takers considered the VR-intake OSCE did not only measure their PACs, but also their abilities to use the VR equipment and the VSP system. For psychologists with little or no VR experience, their test scores might be affected by unfamiliar use of the VR equipment and the VSP system. Second, a mock test taker reported that she felt dizzy and discomfort similar to motion sickness after taking the VSP test, and then affected her performance on the follow-up S test. Third, most of the mock test takers agreed that the VR-intake OSCE could be used to replace some assessment functions of the intake OSCE when its VSP system has been fully developed.

Based on responses provided by mock test takers, the test administration procedure of the VR-intake OSCE in the formal study phase was revised as the followings. First, an additional practice VSP test would be provided to allow participants to familiarize the VR equipment and the VSP system before taking the VR-intake OSCE. Second, an intermission session between the VSP test and the follow-up S test would be extended from 5 minutes to 10 minutes to help participants to relieve possible dizziness and discomfort before taking the S test.

In the formal study phase, a sample of 37 participants was recruited on a voluntary basis. All participants were divided into three groups (i.e., clinical student group, counseling student group, and counseling psychologist group) to examine the feasibility of the VR technology in assessing PACs of psychologists and the design of the VSP system. All participants first took the VR-intake OSCE, and then filled out a post-test questionnaire, which was designed to examine the feasibility of the VR technology in assessing PACs and the design of the VSP system.

Results of the formal study revealed that the VR-intake OSCE was more feasible in assessing cognitive aspects of PACs (e.g., differential diagnosis and integrated inference) than behavioral aspects of PACs (e.g., communication skills and professional attitudes). In terms of the design of the VSP system, participants gave the "user interface" design and the "help function" design the highest evaluations. They also suggested that the designs of "voice recognition" and "interactivity" needed further improvement.

From the viewpoints of students and psychologists, the application of VR technology in assessing PACs of psychologists from entry to practice is feasible. The cost-effective effect of using the VR-intake OSCE to replace the intake OSCE can be further investigated after its VSP system has reached a mature state. In the near future, researchers and educators who are willing to devote themselves to the VR application in psychology OSCEs can develop different types of VSP systems in order to assess a wide range of professional competencies (e.g., intervention, consultation, and supervision/teaching). They can also apply the VR technology to different types of clinical training courses to enhance the training effects and enrich in-class activities.

Keywords: objective structured clinical examination, psychological assessment, virtual reality

