

精神障礙者之功能評量工具之心理計量特性

王怡婷¹ 潘瓊琬^{1,2} 鍾麗英³ 陳詞章⁴ 劉麗婷¹ 陳韻玲¹

摘要：本研究的目的是同時採用驗證性因素分析以及羅序模式以驗證職能治療綜合評量表於台灣精神障礙者之心理計量特性。本研究之個案來自於北部某醫學中心的精神科病房，共 321 位。包含 142 位男性 (44.2%)，平均年齡為 30.2 歲 (標準差=9.43)。本研究採用的職能治療綜合評量表為全國精神科職能治療師經常使用之功能評估。研究資料以 WINSTEPS 3.56、AMOS7.0 及 SPSS 17.0 軟體進行統計分析。結果顯示社交行為次量表符合單一建構效度，但是其精確度較差。項目的難易度排序，精神障礙者在 3 個次量表 (一般行為次量表、社會行為次量表、工作行為次量表) 中最難的題目分別為「活動量」、「團體中角色」及「細動作協調」。社會行為次量表於不同性別、教育及診斷族群間並無任何系統性誤差，一般行為及工作行為次量表於教育程度或性別上各有二題具有項目難度差異性。本研究結果顯示，未來需繼續修訂一般行為以及工作行為次量表，以期符合單一建構之要求。應增加題目以改善社會行為次量表之精確性。

關鍵詞：職能治療綜合評量表，羅序模式，驗證性因素分析

(台灣醫學 Formosan J Med 2012;16:121-8)

前言

內政部統計資料顯示，民國 97 年九月台灣慢性精神障礙者突破十萬人，占身心障礙者總人數比率 9.6%，為所有障別中第二大增加族群，顯示慢性精神障礙者人數比重大幅增加之趨勢[1]。當精神病患病程逐漸慢性化，自我照顧能力變差而需依賴家庭照顧，往往造成家庭成員及社會負擔[2]。職能治療可協助個案在症狀穩定之外，轉而提升能力與技巧，並增進職能活動的參與度，以提升生活品質[3]。

根據美國職能治療學會公布的職能治療實作架構 (Occupational Therapy Practice Framework, Domain and Process)，列出職能治療師所著重的要素及領域。包含：動作計畫技巧、感覺知覺技巧、情緒調節技巧、認知技巧、溝通及社交技巧等，以及自我照顧、工作、休閒、睡眠等職能表現領域。職能治療師往往透過協助個案參與各職能領域之活動，增進其表現之元素，藉以改善個案的職能參

與[2]。

職能治療綜合評量表 (Comprehensive Occupational Therapy Evaluation Scale, COTES) 是一個涵蓋上述元素 (認知知覺、社交互動以及感覺動作等)，且使用普遍的一個評量表[4-5]。治療師透過觀察個案的活動參與，根據項目的定義，評定個案在各元素上的分數，並據以擬訂治療計畫，改善個案之表現，並可做為個案功能改變之指標[4,6]。雖然 COTES 使用廣泛，但相關的信、效度證據相當少。英文版本發展時，曾檢驗施測者間信度，獲得 95% 的一致性。效度之驗證是以觀察 56 位出院病人參加職能治療團體之第一天與最後一天之原始總分差，平均分數從 31 分降至 17 分，表示其能力進步來支持其效度[4]。國內學者高麗芷曾於民國 70 年間，修訂 COTES 之版本，並以 55 位住院患者檢驗其施測者間信度，結果顯示相關值介於 0.61~0.94 之間[6]。效標關聯效度部分，以患者出院 3 個月後是否重返工作或復發為指標，與出院前之分數對照，結果顯示，差異皆

¹台灣大學醫學院職能治療學系，²台灣大學醫學院附設醫院精神部，³台北大學統計系，⁴龍華科技大學機械系
 受文日期 2011 年 6 月 15 日 接受日期 2011 年 10 月 25 日

通訊作者聯絡處：潘瓊琬，台灣大學醫學院職能治療學系，台北市徐州路 17 號 407 室。E-mail: aywoan.pan@gmail.com

達顯著水準($p < 0.001$) [6]。

本研究所採用的是高麗芷等人(1981) [6]修改後的版本，因為此版本普遍使用於臨床中[5]。由於近年來不同測量理論提供了進一步檢測量表品質的方式[7-11]，同時，本量表仍缺少建構效度之證據，因此本研究使用了不同測量理論以進行驗證，以期提供較完整的實證基礎，作為臨床使用之依據。研究者曾於 2002 年以羅序模式驗證 COTES 的建構效度，顯示所有題目均符合單一建構，但項目的涵蓋範圍窄[12]。近幾年羅序模式之發展迅速，提供研究者更完整的方式以分析量表的測量品質[9-11]，因此本研究目的即同時應用羅序模式以及驗證性因素分析[13]以檢驗 COTES 之心理計量特性。

研究方法

一. 個案

本研究的個案來自於北部某醫學中心的精神科病房，共 321 位。年齡介於 14 至 61 歲間，平均為 30.2 歲(標準差=9.4)。其中包含 142 位男性(44.2%)，184 位診斷為精神分裂症(57.3%)，88 位為情感性疾患個案(27.4%)。詳細人口變項分布請見表一。

二. 研究工具

本研究所採用的 COTES 為高等人修訂後的 20 題版本[6]。本測驗之評分為五分等級量表，分數越高代表個案的能力越好[6]。有些題目依行為嚴重度予以評分，例如「責任感」(第 15 題)，五分為「認真獨立」；四分為「尚可」；三分為「依賴他人」；二分為「敷衍塞責」；一分為「逃避責任」。依出現頻率方式來評分以「持續時間」(第 14 題)為例，五分為「90 分鐘以上」；四分為「60~89 分鐘」；三分為「30~59 分鐘」；二分為「10~29 分鐘」；一分為「10 分鐘以下」。COTES 需經由治療師觀察個案表現後進行評分，共分為三個次量表，包含一般行為(七題)、社會行為(五題)、以及工作行為(八題)。詳細題目描述請見表二。

三. 資料分析

本研究以 WINSTEPS 3.56[14]，AMOS 7.0[15]

表一：臨床樣本人口學變項分布情形

		人數(位)	百分比 (%)
性別	男	142	44.2
	女	179	55.8
教育程度	小學	45	14.0
	國中	112	34.9
	高中	134	41.7
	大學	27	8.4
	遺漏值	3	0.9
職業	待業	227	70.7
	學生	10	3.1
	公務人員/教師	8	2.5
	軍人	0	0
	勞工	51	15.9
	商人	15	4.7
	遺漏值	9	2.8
診斷別	精神分裂症	184	57.3
	情感性疾患	88	27.4
	人格疾患	3	0.9
	藥物濫用	3	0.9
	情感性精神分裂症	15	4.7
	精神分裂症+智能障礙	3	0.9
遺漏值	22	6.9	
年齡	年齡範圍 14- 61	平均年齡	標準差
		30.2	9.34

及 SPSS 17.0[16]統計軟體進行資料分析，以平均值、標準差、百分比，進行描述性統計。使用題目－總分相關(item-total correlation)分析每個題目與各次量表總分的相關性。若是高於 0.3，則會進一步進行羅序及驗證性因素分析。

本研究採用的其中一種測量理論是羅序模式(Rasch measurement model)來分析資料[7-8,17]。羅序模式為 1960 年由丹麥數學家喬治-羅序所提出的對數機率理論[7]。可將量表的原始分數透過對數轉換，成為等距分數(logit)，以進行母數統計分析。羅序模式的假設中，個案的能力分數不受所測之題目難易度之影響(測驗獨立，test-free)，且題目之難易度不受樣本之能力影響(樣本獨立，sample-free)[17-18]。本研究將採用以下羅序模式指標作為判定標準：以近合適度(infit mean square)及其標準化合適度(ZSTD)大小來決定是否符合模式

表二：職能治療綜合評量表分量表項目合適度及難易度

項目	MEASURE	ERROR	IN.MSQ	IN.ZSTD
一般行爲				
3. 活動量	1.06	.10	.88	-1.5
1. 出席	.44	.10	1.29	3.3
7. 言語思考障礙	.29	.10	.88	-1.5
4. 情緒障礙	.27	.10	.64	-5.1
2. 儀容	-.45	.10	1.23	2.8
5. 無意義動作	-.73	.10	.96	-.4
6. 生理需求	-.89	.10	.98	-.2
社會行爲				
11. 團體中角色	.70	.12	.82	-2.1
12. 自我肯定	.64	.12	.98	-.2
10. 接受性	-.27	.12	.84	-1.9
9. 予人印象	-.45	.12	.75	-3.0
8. 社交性	-.63	.12	<u>1.55</u>	<u>5.5</u>
工作行爲				
20. 細動作協調	1.60	.12	1.15	1.7
19. 技巧	.67	.12	.88	-1.4
16. 忍受挫折	.30	.12	.75	-3.2
14. 持續時間	-.36	.12	.90	-1.3
17. 自我期許	-.44	.12	.88	-1.5
15. 責任感	-.45	.12	.78	-3.0
13. 動機	-.52	.12	1.30	3.5
18. 理解力	-.80	.12	1.23	2.7

註：三個分量表為獨立單一建構內容，分次進行羅序分析。

要求。近合適度對接近個案能力範圍的題目反應較為敏感，越接近 1，表示試題越符合假設。若近合適度超過 1.4，且標準化合適度大於+2 或小於-2 時，表示該題目與假設模式不符，可能屬於不同建構(construct)[18]。另一指標為殘差值之未轉軸之主成分分析(principal components factor analysis of residuals, PCA)，用以驗證是否為單一向度，若特徵值(eigenvalues)小於 3，表示屬於單一建構向度[9,11]。本研究另一採用的分析方式是驗證性因素分析(confirmatory factor analysis, CFA)[13]。採用的指標包含適合度指標(goodness-of-fit index, GFI)及調整適合度指標(adjusted goodness-of-fit index, AGFI)>0.9。[19-21]。本研究訂定的標準是需同時符合羅序模式以及驗證性因素分析之適合度檢驗的要求。

另外，本研究將探討題目之難易度分數是否符合理論預期，並進一步了解題目難易度之分佈是否

涵蓋精神障礙者之能力範圍[17-18]。另將以訊息函數(information function, IF)了解此量表之項目準確性，是否能準確測得所欲施測之個案的功能[19]。我們也將檢驗項目的分離值指標(separation index)，以了解此量表是否能將個案分類於不同層級之功能程度，研究者預期應具有分類為 2 個等級以上之品質[10-11]。本研究將檢驗項目分離信度，類似傳統測驗理論中的 Cronbach's α ，代表測驗的內在一致性，研究者期待其數值為 .90 以上[19-20]。研究者並將以項目函數差異(differential item functioning, DIF)檢視項目參數的穩定度[14]，WINSTEPS 軟體是以 Mantel-Haenszel 方法，在固定(anchored)了個案能力值之後，分析該題目是否因不同性別、教育程度或診斷，造成題目反應(困難度)之顯著差異。本研究將三個次量表的 alpha 值標準分別訂為 0.05/7=0.007(一般行爲次量表)、0.05/4=0.013(社會行爲次量表)、

0.05/8=0.0063(工作行為次量表)以調整其項目數。最後，將檢視題目之評分等級使用的恰當性，是否有錯置的現象[10-11]。

研究結果

一. 建構效度

以各次量表分別進行題目-總分(item-total)之相關分析，結果顯示所有次量表之相關皆高於0.4。進一步進行各次量表之羅序分析結果顯示，一般行為及工作行為次量表可符合模式預期，為單一建構效度。僅社會行為次量表之「社交性」(第八題)無法符合模式預期(表二)。刪除該題後，社會行為次量表可符合模式預期，因此研究者決定後續分析中刪除第八題「社交性」。殘差值之主成分分析結果顯示，一般行為次量表可解釋模式中76.5%之變異量，特徵值為1.9；社會行為次量表可解釋模式中85.8%之變異量，特徵值為1.6；工作行為次量表可解釋模式中78.3%之變異量，特徵值為2.7；三個次量表之殘差值皆在可接受之範圍。

驗證性因素分析結果顯示，「一般行為」次量表符合GFI適配度標準、「社會行為」次量表符合AGFI和GFI適配度標準。「工作行為」次量表未符合任何適配度指標標準，此結果代表僅有社會行為次量表之建構效度符合要求(表三)。

二. 項目難易度

羅序分析結果顯示之題目困難度分別敘述如下。「一般行為」次量表得分難易度介於-0.89至1.06之間，最困難的題目為「活動量」，最簡單的題目為「生理需求」；「社會行為」次量表得分難易度介於0.7至-0.45之間，最困難的題目為「團體中角色」，最簡單的題目為「予人印象」；「工作行為」次量表得分難易度介於1.60至-0.80之間，最困難的題目為「細動作協調」，最簡單的題目為「理解力」(表二)。此排序符合臨床觀察。

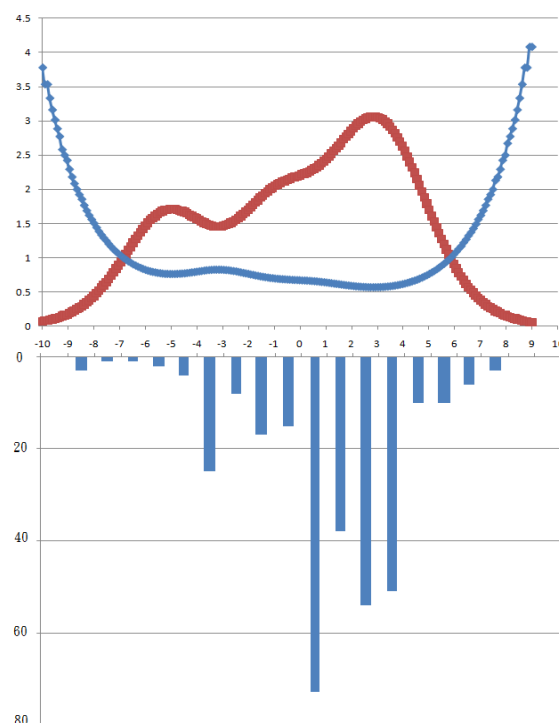
三. 項目精確度

測驗訊息函數(TIF)與人數分佈表之對照圖顯示，一般行為次量表可精確測量的範圍介於-6.87~5.87之間，超出範圍外的受試者共13人(4%)(圖一)。社會行為次量表可精確測量的範圍有

表三：職能治療綜合評量表整體適配度結果

次量表/適配度指標	AGFI>.90	GFI>.90
一般行為	.768	.927
社會行為	.936	.987
工作行為	.559	.755

註：統計分析人數為293人，排除遺漏值。

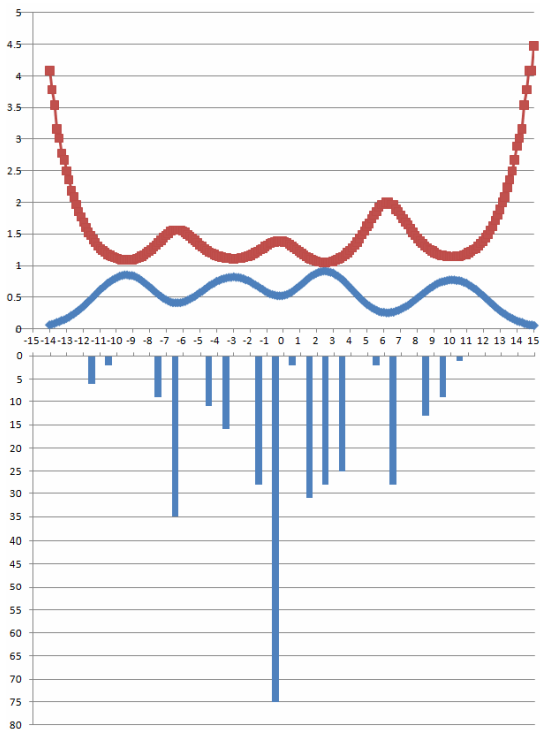


圖一：一般行為測驗訊息函數與人數分佈

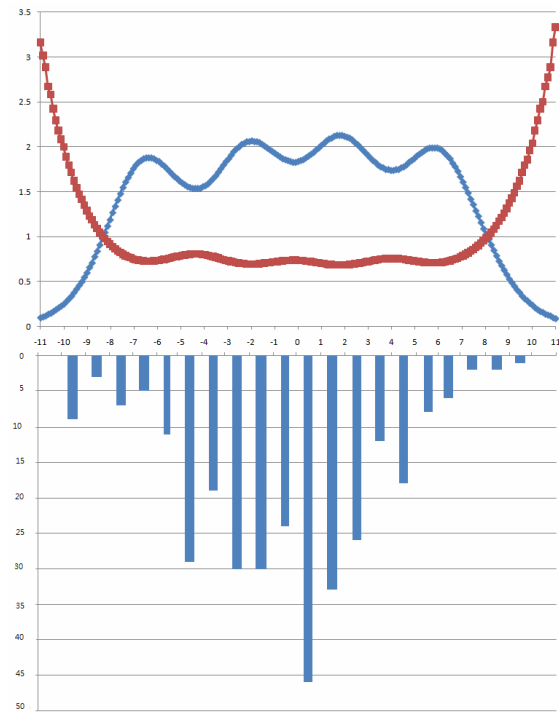
註：上圖中：粗線為測驗訊息函數(TIF)值，細線為相對應之標準誤(SE)，Y軸為TIF值，X軸為能力分佈，單位:logits。圖下為樣本的能力分佈，Y軸單位:人數。圖中可精準測得範圍介於-6.87 logits至5.87 logits，在此能力範圍外的受試者共13人(4%)。

4個，分別是介於-10~-8, -4~-2, 2~4, 9~11之間。可準確測出約36%個案的社會行為功能(圖二)。可能是因為社會行為次量表僅包含四題，造成誤差偏大。工作行為次量表可精確測量的範圍介於-8.36~8.14之間。超出範圍外的個案共15人(4.6%)(圖三)。

此外，四個次量表個案分離值分別為，一般行為為4.347、社會行為為3.853；工作行為為5.506。相等於職能治療綜合評量表的所有次量表能將個案的



圖二：社會行為測驗訊息函數與人數分布
 註：上圖中：粗線為測驗訊息函數(TIF)值，細線為相對應之標準誤(SE)。Y軸為 TIF 值，X軸為能力分佈，單位:logits。圖下為樣本的能力分佈，Y軸單位:人數。



圖三：工作行為測驗訊息函數與人數分布
 註：上圖中：粗線為測驗訊息函數(TIF)值，細線為相對應之標準誤(SE)，Y軸為 TIF 值，X軸為能力分佈，單位:logits。圖下為樣本人數於能力分佈，Y軸單位:人數。

一般行為、社會行為以及工作行為能力分為 3 到 5 個等級。

從項目 IF 中得知，尖峰越高之題目，誤差越小。亦即該項目為此次量表中變異量最小、最精確之題目。一般行為次量表中，「無意義動作」(第五題)最可反映個案之功能;社會行為及工作行為次量表中所有題目精確性相仿，並無特別顯著之題目。

四. 信度

項目分離係數之信度，一般行為次量表為 0.98、社會行為次量表為 0.95、工作行為次量表為 0.97，顯示 COTES 具有極佳的內在一致性。

五. 項目參數穩定度

透過 DIF 檢驗項目參數於性別、教育及診斷的穩定度。一般行為次量表中，高中教育背景者較國中教育程度者有情緒障礙問題，男性生理需求問

題較嚴重(第 4 題「情緒障礙」及第 6 題「生理需求。」) ($p < 0.007$)。在工作行為次量表中，高中畢業者較國中畢業者挫折忍受度差(第 16 題「挫折忍受。」);國中畢業者較高中畢業者理解力差(第 18 題「理解力。」) ($p < 0.0063$)。社會行為次量表之題目於性別、教育及診斷族群差異中並沒有出現任何系統性誤差($p < 0.0013$)。

六. 項目評分等級適當性

利用部份計分模式(partial credit model)檢驗各題之評分等級適當性。結果顯示，各題目的五分等級評分結構適當，無錯置現象。

討論

本研究應用羅序測量模式以及驗證性因素分析來檢驗職能治療綜合評量表之心理計量特性，羅

序分析結果顯示，職能治療綜合評量表可分成 3 個單獨的次量表。在社交行為次量表中之第 8 題「社交性。」不符合羅序模式預期，而予以刪除。從殘差的主成分分析結果顯示，三個次量表的特徵值大小，符合分別具有單一建構內容之要求。驗證性因素分析結果顯示，一般行為以及工作行為次量表的適配指標不符合要求。因此，最後研究結果顯示，職能治療綜合評量表的社會行為次量表符合單一建構效度。其他兩個次量表則需要重新修訂與設計。譬如：一般行為次量表中包含了精神症狀題目(如：言語思考障礙)以及行為題目(如：活動量)，可以修訂成可觀察的行為表現(如：答非所問)。而工作行為次量表中包含了細動作協調(動作技巧)以及動機、自我期許(意志層次)，建議將之改成可從工作行為表現而觀察到的行為題目(如：持續工作直到完成)。

職能治療綜合評量表具有良好的信度、一般行為及工作行為次量表具有準確的測量範圍，社會行為次量表因題目較少影響了估計的準確性，未來可增加精神疾病患者相關之社會行為題目。另外，透過項目訊息函數(item information function)得知「無意義動作」(第 5 題)為最精確的題目，可做為臨床篩選個案功能的指標項目。本研究也發現共有 4 題具有 DIF，國中教育程度以下的個案理解力較差；高中教育程度以上者，挫折忍受較差且易有情緒障礙問題；治療師需特別關注男性患者之生理需求議題等，建議研究者於臨床推理時可特別留意。

本研究的限制包括：由於個案來源僅限於接受職能治療服務的急性住院個案，未能包含具代表性的精神障礙者樣本，因此在研究結果的推論上有限。同時，本研究混合了不同診斷之個案資料進行驗證性因素分析，可能造成誤差。未來的研究方向包括：修訂一般行為以及工作行為次量表之題目以改善建構效度；增加社會行為次量表之題目，增進量表之精確度。收集更多單獨診斷之樣本資料以減少誤差；持續進行修訂後之心理計量品質之驗證，例如：施測者間信度、再測信度以及預測效度等等。

致謝

本研究感謝行政院國科會專題研究計畫

(NSC97-2628-S-002-001-MY3)之經費補助。

聲明

本研究，利益衝突：無。知情同意：無。受試者權益：無人體或動物實驗。

參考文獻

1. 內政部統計處：九十七年第五十一週內政統計通報-身心障礙之慢性精神病患者人數。台北市：內政部，2008。
2. American Occupational Therapy Association: Occupational therapy practice framework: Domain and process, 2nd ed. *Am J Occup Ther* 2008;62:625-83.
3. 蕭淑貞、陳賢佳、葉櫻媛等：病患居家自我照顧現象探討。中華衛誌 1997;16:478-87。
4. Brayman SJ, Kirby TF, Misenheimer AM, et al: Comprehensive occupational therapy evaluation scale. *Am J Occup Ther* 1976;30: 94-100.
5. 蕭小菁、潘璦琬、鍾麗英等：台灣精神科職能治療評估工具的現況調查。職能治療學會雜誌 2000;18:19-32。
6. 高麗芷、葉英堃、林民裕等：職能治療綜合評量表之設計。中華心理學刊 1981;1:1-7。
7. Rasch G: Probabilistic models for some intelligence and attainment tests. Copenhagen, Danmarks Paedagogiske Institute, 1960.
8. Wright BD, Linacre JM: Observational are always ordinal: measurements, however, must be interval. *Arch Phys Med Rehabil* 1989;70: 857-60.
9. Hwang JL, Davies PL: Brief Report-Rasch analysis of the School Function Assessment provides additional evidence for the internal validity of the Activity Performance Scales. *Am J Occup Ther* 2009; 63: 369-73.
10. Árnadóttir G, Fisher AG: Rasch analysis of the ADL Scale of the A-ONE. *Am J Occup Ther* 2008;62:51-60.
11. Clemson L, Bundy A, Unsworth C, et al:

- Validation of the modified assessment of living skills and resources, an IADL measure for older people. *Disability Rehabil* 2009;31:359-69.
12. Han CC, Pan AW: Validity study of a comprehensive functional scale for patients with psychiatric disorders in Taiwan. Poster presented at the 13th World Congress of Occupational Therapists, Stockholm, Sweden, 2002, June 23-28.
 13. Mulaik S, Nesselroade JR: Confirmatory factor analysis. In: Nesselroade JR, Cattell RB, eds. *Handbook of Multivariate Experimental Psychology*, 2nd ed. New York, Plenum Press, 1988, xxviii, 966 pp.
 14. Lincare JM: A User's Guide to WINSTEPS, MINISTEP: Rasch Model Computer Programs 2003 Unpublished Manuscript, Chicago.
 15. Arbuckle JL: Amos (Version 7.0) [Computer Program]. Chicago: SPSS, 2006.
 16. SPSS : Analysis without anguish: version 17.0 for windows 2010. Sheridan J. Coakes, Lyndall Steed, Clara Ong.
 17. Trevor GB, Cook J: Applying the Rasch Model. *Fundamental Measurement in the Human Sciences*, 2nd Edition. University & Christine M. Fox, University of Toledo, 2007.
 18. 潘瓊琬, 許瑋玲: 貝氏憂鬱量表第二版建構效度之探討: 羅序模式之應用。 *台灣醫學* 2008;16:627-34。
 19. Lai JS, Cook K, Stone AA, et al: Classical testing theory and item response theory/Rasch model to assess difference between patient-reported fatigue using seven-day and four-week recall periods. *J Clin Epidemiol* 2009;62: 991-7.
 20. Hays RD, Liu H, Spritzer K, et al: Item response theory analyses of physical functioning items in the Medical Outcomes Study. *Med Care* 2007;45(Suppl 1):S32-8.
 21. Chung L, Wang YH, Chen TJ, et al: The predictive factors for length of stay for stroke patients in Taiwan using the path model. *Int J Rehabil Res* 2006;29:37-143.

Psychometric Properties of the Comprehensive Occupational Therapy Evaluation Scale for Persons with Mental Illness

I-Ting Wang¹, Ay-Woan Pan^{1,2}, Lyinn Chung³, Tsyr-Jang Chen⁴, Li-Ting Liu¹, Yun-Ling Chen¹

Abstract: The aim of this study was to examine the construct validity, accuracy, and differential item functioning (DIF) of the Comprehensive Occupational Therapy Evaluation (COTES) with Rasch measurement model and confirmatory factor analysis (CFA). A total of 321 persons with mental illness were recruited from an inpatient unit of a teaching hospital in northern Taiwan. There were 142(44.2%) males and 179(55.8%) females. The mean age of the subjects was 30.2 years (SD=9.43). The WINSTEPS 3.56, AMOS 7.0 and SPSS 17.0 were used for data analysis. The results showed that only the social behavior subscale fitted with the model requirement for both CFA and Rasch model. But the accuracy of the subscale was not good. The most difficult endorsed items for three subscales were the activity level (general behavior subscale), group roles (social behavior subscale) and fine motor coordination (task behavior subscale). There was no DIF item in the social behavior subscale; two DIF items were found for the general behavior and task behavior subscales. Therefore, there are needs to revise the general behavior and task behavior subscales to fit the requirement of unidimensionality. More items are needed in the social behavior subscale to increase its accuracy.

Key Words: Comprehensive Occupational Therapy Evaluation Scale, Rasch analysis, confirmatory factor analysis

(Full text in Chinese: Formosan J Med 2012;16:121-8)

¹School of Occupational Therapy, College of Medicine, National Taiwan University; ²Department of Psychiatry, National Taiwan University Hospital; ³Department of Statistics, National Taipei University; ⁴Department of Mechanical Engineering, LungHwa University of Science and Technology, Taipei, Taiwan.

Received: June 15, 2011

Accepted: Oct 25, 2011

Address correspondence to: Ay-Woan Pan, School of Occupational Therapy, College of Medicine, National Taiwan University, No.17, Xu-Zhou Rd., Taipei, Taiwan. E-mail: aywoan.pan@gmail.com