

國立臺灣大學醫學院職能治療學系



碩士論文

School of Occupational Therapy

College of Medicine

National Taiwan University

Master Thesis

使用羅序分析交叉檢驗中風復健動作評估量表

上肢／下肢動作次量表之心理計量特性

Cross-validation of the psychometric properties of
Upper-Limb / Lower-Limb Movements subscales of the
Stroke Rehabilitation Assessment of Movement
(STREAM) with Rasch Analysis

林信豪

Hsin-Hao Lin

指導教授：謝清麟 博士

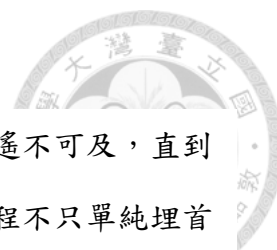
Advisor: Ching-Lin Hsieh, Ph.D.

中華民國 107 年 5 月

May 2018



誌謝



終於，我的碩士生活要落幕了。過去總是認為畢業這件事遙不可及，直到口試結束的那一刻起，才驀然產生真實感。對我來說，這段歷程不只單純埋首於學術，更多的是自我覺察許多不足並加強，是生長至今重要的蛻變之一。

由衷感謝我的指導教授謝清麟老師，當我汲汲營營於論文的撰寫時，老師除了研究上的指導以外，也時常與我分享待人處世的智慧、以及身為一位研究生該有的自覺與能力養成。十分感謝吳建德老師與黃小玲老師，在我陷入掙扎時給予的教導與溫暖，使我的研究之路上擁有更強大的支持。研究過程中著實受到許多人的協助，謝謝恭宏學長及怡靜學姊不厭其煩地為我解惑統計分析相關的問題，以及研究夥伴揚揚、姿廷彼此間的督促鼓勵，使我面對研究困境時不再是獨自一人的困獸之鬥，而是互相砥礪切磋。

也感謝這段時間內不斷關心並鼓勵我的朋友們，即便是社群軟體上的問候也總讓我獲得不少能量、提振士氣。感謝我的家人們，永遠是我最有力的靠山，由於你們全心全意的支持與包容，使我得以自由選擇想做的事情。

回首來時路，這份歷練帶來的酸甜苦辣難以言喻，但如果再讓我選擇一次，我仍會深吸一口氣，進入學術世界闖蕩。現在，新的旅程即將拉開序幕，期許自己能帶著氣定神閒的笑容迎接各種挑戰，在名為信念的旗幟下！

中文摘要

背景：中風復健動作評估量表 (Stroke Rehabilitation Assessment of Movement, STREAM) 目前已有許多古典測驗理論相關的心理計量特性研究。但仍有三個問題待解決：(1) STREAM 上、下肢次量表題目是否符合單向度 (unidimensionality)? (2) 將順序量尺 (ordinal) 直接當作等距量尺 (interval) 加總計分，有數學運算不當的問題。(3) 題目難度之估計具有樣本依賴性。羅序分析 (Rasch analysis) 以數學模式為基礎，依據個案作答反應以等距量尺同時估計個案能力及題目難度，且當評量數據符合羅序模式 (Rasch model) 預期時，該題目符合單向度。然而至今僅有一研究使用羅序分析驗證 STREAM，無法呈現充分且穩定的證據。

研究目的：本研究目的為交叉檢驗 STREAM 上、下肢動作次量表之心理計量特性，包含建構效度之單向度、羅序信度，以及題目難度排序相關性。並且另外檢驗，以本研究及昔日研究之題目難度參數估計個案分數的相關性。

研究方法：採用次級資料分析法。先自原始資料中篩選出符合收案標準的 302 位亞急性期中風患者，再以羅序分析之評等量尺模式 (rating scale model) 驗證：單向度以各題目之羅序模式適配度為指標，適配度良好表示該題目具有單向度；羅序信度則假設測量標準誤會隨著個案能力不同而有所差異。最後將分析結果與過去研究進行交叉檢驗，比較心理計量特性的一致性及其相關性。

結果：(一) 上肢次量表僅有 6 題目具單向度，包含：舉手碰頭頂、手向後摸對側屁股、手臂舉高過頭、前臂旋前／後、握拳、手打開。下肢次量表有 8 題目具單向度，包含：大腿抬高、膝關節伸直、膝關節彎曲、於坐姿下腳板翹高且腳跟著地、腳跟翹高且腳尖著地、膝關節伸直且腳趾翹起、髖關節伸直且膝關節彎曲、於站姿下腳板翹高且腳跟著地。(二) 上、下肢次量表羅序信度分別為

0.92、0.93，均略高於 Hsueh 等研究的 0.86、0.91。上、下肢次量表之個案信度皆為 243 人大於 0.90、59 人介於 0.80~0.90 之間。(三) 二篇研究之上、下肢次量表題目難度排序分別為低度及中度相關 (Spearman's rho = -0.31、0.48)。(四) 二篇研究之上、下肢次量表題目難度參數估計個案分數之相關性分別為高度及中度相關 (Pearson's r = 0.96、0.62)。

結論：交叉檢驗 STREAM 上、下肢動作次量表之心理計量特性，發現上肢次量表題目單向度結果不一致、其題目難度排序僅有低度相關，但其題目難度參數估計之個案分數具高度相關。下肢次量表單向度結果相對較一致、題目難度排序有中度相關，然而其難度參數估計之個案分數卻僅有中度相關。分析結果與預期之間的落差也許跟樣本特性有關，但僅憑二篇研究之比較，較難推論二數據之優劣。未來宜再驗證並釐清之。

關鍵字：中風復健動作評估量表、羅序分析、心理計量特性、動作功能評估量表

ABSTRACT



Background: There have been many studies based on classical test theory (CTT) which validated the psychometric properties of the Stroke Rehabilitation Assessment of Movement (STREAM). But there were at least 3 psychometric issues that needed to be investigated, including (1) the unidimensionality of upper extremity (UE) / lower extremity (LE) movements subscales of the STREAM; (2) the direct transformation of ordinal raw STREAM total scores into interval scores, which was not mathematically valid; and (3) the estimation of the item difficulty parameters was sample dependent. The Rasch analysis enables examining unidimensionality and Rasch reliability. Moreover, Rasch analysis could transform the STREAM from an ordinal-level measure into an interval-level measure, and estimate both person's ability and item difficulty. However, so far there was only one study using the Rasch analysis to validate the STREAM, which couldn't provide sufficient and solid evidence.

Purpose: To cross-validate the psychometric properties of UE / LE movements subscales of the STREAM with the Rasch Analysis, including (1) the unidimensionality of construct validity; (2) the Rasch reliability; and (3) the correlation of the order of item difficulty parameters between previous study and this study. Besides, researchers validated the correlation of latent trait scores based on previous parameters and this study's.

Methods: Researchers employed secondary data analysis. First, 302 stroke patients during subacute period were derived by screening secondary data source. Then all items were examined with rating-scale model of Rasch analysis. For

example, the unidimensionality was examined by each item's goodness-of-fit index; and the Rasch reliability was estimated by standard error of measurement, which was assumed that it varied with different latent trait level. Last but not least, we cross-validated the outcomes between previous study and this study, and compared the consistency and the correlation of the psychometric properties.

Results: (1) Only 6 items from UE subscale possessed the unidimensionality, including "Raises hand to touch top of the head", "Places hand on sacrum", "Raises arm overhead to fullest elevation", "Supinates and pronates forearm", "Closes hand from fully opened position", and "Opens hand from fully closed position." While 8 items from LE subscale possessed the unidimensionality, including "Flexes hip in sitting", "Extends knee in sitting", "Flexes knee in sitting", "Dorsiflexes ankle in sitting", "Plantar flexes ankle in sitting", "Extends knee and dorsiflexes ankle in sitting", "Flexes affected knee with hip extended", and "Dorsiflexes affected ankle with knee extended." (2) The Rasch reliabilities of UE / LE subscale of this study were 0.92 and 0.93 respectively, which were both slightly higher than Hsueh's study (0.86 and 0.91 respectively). Besides, 243 patients' person reliabilities of UE / LE subscale of this study were higher than 0.90, and remaining 59 patients' person reliabilities were between 0.80 ~ 0.90. (3) The order of item difficulty parameters from UE / LE subscale between previous study and this study was modestly / moderately correlated, with Spearman's rho equaled to -0.31 and 0.48 respectively. (4) The latent trait scores from UE / LE subscale estimated by previous parameters and this study's was highly / moderately correlated, with Pearson's r equaled to 0.96 and 0.62 respectively.

Conclusions: With regard to UE subscale, although the outcome of the

unidimensionality was inconsistent, and the order of item difficulty parameters was modestly correlated, the latent trait scores estimated by previous parameters and this study's was highly correlated. On the contrary, LE subscale had more consistent outcome of the unidimensionality as well as moderately correlated order of item difficulty parameters, but the latent trait scores estimated by previous parameters and this study's was just moderately correlated. The difference between analysis results and expectation was probably related to sample properties. But the limitation was that it's still difficult to identify which parameter was better since the comparison was only between two studies. It's suggested to revalidate the psychometric properties of STREAM with Rasch Analysis and clarify the problems in the future.

Keywords: Stroke Rehabilitation Assessment of Movement (STREAM), Rasch analysis, psychometric properties, movement assessment

目 錄



| | |
|-----------------------------------|------|
| 誌謝..... | i |
| 中文摘要..... | ii |
| 英文摘要..... | iv |
| 圖目錄..... | viii |
| 表目錄..... | ix |
| 第一章 前言..... | 1 |
| 第一節 腦中風與其動作功能損傷..... | 1 |
| 第二節 中風患者動作功能評量之重要性..... | 4 |
| 第三節 簡介羅序分析..... | 12 |
| 第四節 常用於中風患者之動作功能評估量表優缺點分析..... | 20 |
| 第五節 中風復健動作評估量表動作次量表羅序分析結果之不足..... | 26 |
| 第二章 研究方法..... | 28 |
| 第一節 研究對象..... | 28 |
| 第二節 研究工具..... | 29 |
| 第三節 研究步驟..... | 29 |
| 第四節 資料分析..... | 30 |
| 第三章 結果..... | 34 |
| 第四章 討論..... | 37 |
| 第五章 結論..... | 44 |
| 參考文獻..... | 45 |

圖目錄



| | | |
|------|---|----|
| 圖 1 | 試題特徵曲線 (ICC)..... | 49 |
| 圖 2 | 上肢次量表一個案能力與個案信度散佈圖..... | 50 |
| 圖 3 | 下肢次量表一個案能力與個案信度散佈圖..... | 51 |
| 圖 4 | 上肢次量表—本研究及 Hsueh 等研究參數各別估計個案分 數之散佈圖..... | 52 |
| 圖 5 | 下肢次量表—本研究及 Hsueh 等研究參數各別估計個案分 數之散佈圖..... | 53 |
| 圖 6 | 上肢次量表—本研究題目難度參數及個案能力分布圖..... | 54 |
| 圖 7 | 上肢次量表—Hsueh 等研究題目難度參數及個案能力分布 圖..... | 55 |
| 圖 8 | 上肢次量表—本研究及 Hsueh 等研究題目難度散佈圖..... | 56 |
| 圖 9 | 下肢次量表—本研究題目難度參數及個案能力分布圖..... | 57 |
| 圖 10 | 下肢次量表—Hsueh 等研究題目難度參數及個案能力分布 圖..... | 58 |
| 圖 11 | 下肢次量表—本研究及 Hsueh 等研究題目難度散佈圖..... | 59 |

表目錄



| | | |
|------|--|----|
| 表 1 | STREAM 之應用文獻分類表..... | 60 |
| 表 2 | STREAM 之心理計量驗證文獻分類表..... | 63 |
| 表 3 | STREAM 之上下肢動作次量表..... | 65 |
| 表 4 | 本研究對象之基本人口學資料..... | 66 |
| 表 5 | 上下肢次量表之羅序信度..... | 67 |
| 表 6 | 上肢次量表之個案信度分布情形..... | 68 |
| 表 7 | 下肢次量表之個案信度分布情形..... | 69 |
| 表 8 | 本研究及 Hsueh 等研究之上肢次量表題目難度參數與排序.. | 70 |
| 表 9 | 本研究及 Hsueh 等研究之下肢次量表題目難度參數與排序.. | 71 |
| 表 10 | 上下肢次量表—本研究及 Hsueh 等研究題目難度排序之相關性..... | 72 |
| 表 11 | 上下肢次量表—本研究及 Hsueh 等研究參數各別估計個案分數之相關性..... | 73 |

第一章 前言

第一節 腦中風與其動作功能損傷

腦血管疾病，俗稱腦中風，主因為腦部血流受阻導致無法供應氧氣及營養，使腦組織缺氧受到損傷，進而造成該腦區所支配部位產生障礙。根據病因不同，主要分為缺血性腦中風 (ischemic stroke) 和出血性腦中風 (hemorrhagic stroke)，前者約占台灣急性腦中風患者的 70% (胡漢華, 2008)。

根據行政院衛生福利部統計，民國 104 年國內十大死因的前三名分別是惡性腫瘤 (癌症)、心臟疾病、腦血管疾病 (cerebrovascular accident, CVA)，分別佔總死亡人數的 28.6%、11.7%、6.8% (行政院衛生福利部, 2015)。依據台灣地區中風發生率的研究指出，36 歲以上人口的發生率約為千分之三，若以台灣地區 36 歲以上人口數接近一千萬，每年的中風新發生數約為三萬人 (邱弘毅, 2008)。在台灣每年約有 17,000 人會因為腦中風而導致日常生活失能，腦血管疾病的預防及治療一直是國內醫學界非常重視的領域，近年來雖然腦血管疾病的死亡比率持續以些微幅度下降，但是存活下來的患者卻仍因為殘留著不同程度的損傷、癱瘓或失能，其所耗費的醫療資源長期占據健保使用的前三名 (邱弘毅, 2008)。

腦中風的後遺症相當廣泛，包含感覺功能、認知理解、溝通能力、動作功能及精神心理等不同層面。例如：感覺消失或異常、言語不清或溝通困難、暈眩、意識混亂、劇烈頭痛、失智、臉部表情不對稱、吞嚥困難、大小便失禁或智力下降等多重功能障礙。其中在動作功能損傷方面，單側手腳的麻痺或無力是最常見的一種腦中風典型症狀，且根據腦血管病變的位置和大小差異，會讓

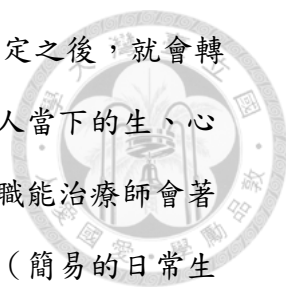
患者產生不同症狀、嚴重程度的動作障礙，例如：姿勢控制能力變差、平衡及協調性不佳、肌肉張力異常、肌力下降等症狀，使得患者日常生活能力受限，需要照顧者投入大量心力及時間，連帶影響患者個人與其家庭之生活品質 (Kwok et al., 2006)。因此，腦中風患者的動作功能復原程度與其未來日常生活功能及生活品質之間皆有重大關聯。

職能治療為中風復健醫療團隊中重要的一員，患者會先經過相關專科醫師診斷並開立處方後轉介至職能治療部門，接受「以患者為中心」的照護。職能治療的主要目標為協助患者恢復動作功能與自我照顧能力，盡可能讓患者重新回到發病前日常生活功能之獨立自主，以降低其對他人在生活上的依賴，並減輕家庭與社會的負擔，提升患者與家屬的生活品質，進而重建患者原來的職能角色。職能治療師會先評估患者的生、心理功能損傷程度以便訂定後續的治療介入，並且設計分級的治療活動協助患者逐步達成短至長期的復健目標。

依據職能治療於中風患者介入的歷程可略分為三期：

一、急性期：指中風發病後最初的幾天，患者病情不穩定。有些患者的患肢仍呈現軟癱無力的狀態，無法自行活動及翻身，容易造成壓瘡、肌力衰退、關節攣縮及肢體水腫等症狀。以出血性中風患者而言，大約 1/4 患者在中風發作後一小時恢復些微自主動作、而大約 1/2 會在發作後六小時恢復些微自主動作(Biller J, 1990)。然而由於患者能做出的動作有限，能夠執行的動作相關復健活動也就不多，通常職能治療師會教導適當的擺位、翻身技巧及被動關節運動，或藉由副木幫助損傷較嚴重患者達到正確的擺位，在保護患者為前提下進行治療，也為後續的介入作準備。

二、亞急性期：中風發病約一週後，病情逐漸穩定。輕、中、重度動作損傷患者的比例皆差不多；患者於發作一個月後，不論性別、患側，動作損傷程



度皆有顯著的改善 (Bonita R, 1988)。當患者身體狀況相對較穩定之後，就會轉入一般病房，進行較密集的復健治療。醫療人員通常會依據病人當下的生、心理狀況，決定做床邊治療或另行移至復健部門做治療。此時期職能治療師會著重於上下肢動作的誘發、姿勢平衡及日常生活自理能力等訓練（簡易的日常生活活動包括：床上移動、翻身、起床及進食等，而洗澡、穿脫衣物及上下樓梯則相對難度較高）。患者動作功能的恢復情形與其日常生活獨立程度息息相關，職能治療師也常常將動作相關的復健訓練融入患者的日常活動之中，例如穿脫衣物的訓練中便包含上肢及肩關節活動度的復健；此外也可以結合患者的興趣嗜好或未來職能角色於復健活動之中，以增進個別化的日常生活功能獨立。

三、慢性期：大約為中風發病後六個月起始，此時患者腦部損傷及整體生理狀況已相當穩定，而用復健門診的形式繼續接受治療。依據台灣中風登錄系統資料顯示，患者在中風出院後、一個月及三個月之所在地分布顯示，出院後大部份的中風患者皆是返回住家，比例高達 92.9%，轉往一般安養機構與呼吸照護機構的比例分別為 6.4%與 0.7%(邱弘毅, 2008a)。此時期職能治療師會更著重於工具性日常生活活動的訓練，包括：家務處理、社區行動、準備餐點、使用電話、服藥等，並讓患者將治療活動中習得的動作模式實際應用於日常生活中，增進患者居家生活的能力。

此時期患者的動作治療潛力，理論上，已達到高原期 (plateau)。不過國內許多治療師仍會將治療重點放在持續加強患肢的動作功能、增進動作品質。這現象可能由於個案要求、家屬期待或是國內治療界習慣等緣故，進而形成一種長時間著重於患肢動作恢復的特別復健景象。然而，除了動作恢復，復健的最終目的應為盡可能地達到生活獨立自主程度，而非僅專注於動作功能上的恢復。

第二節 中風患者動作功能評量之重要性

動作功能損傷是中風常見的後遺症 (G. Broeks, Lankhorst, Rumping, & Prevo, 1999)，為了制訂治療計畫及監測病程進展，動作功能的評量是關鍵。具有良好信效度的動作功能評量能客觀、科學性地呈現患者的問題及恢復情形，其重要性包含：

一、對於臨床人員而言（以臨床職能治療為例）：

1. 能在不同臨床階段提供臨床推理及決策之參考。

(1) 剛接案的初評－「掌握個案動作損傷程度，並訂定對患者而言最適切的治療計畫及長、短期目標」：

此時期患者的患側肢體大多癱軟無力，治療師在掌握所需之動作功能相關資訊後，便能在確保患者安全的前提下，教導病人如何使用僅存的動作進行復健活動，以預防臥床導致的肌力衰退、關節攣縮、水腫及褥瘡等症狀

另外，根據初評的結果，治療師便能初步制定患者的長、短期治療目標，並且提供患者此時期能夠執行的動作訓練計畫，依據不同動作功能損傷程度，調整介入的方案，以達到理想中的復健目標。

(2) 療程中的再評及結案時的療效評量－「呈現患者動作功能治療後的變化及進步，以調整治療目標及計畫」：

經過一系列治療活動及自身身體機能恢復，多數患者的動作功能應會逐漸進步，因此治療師必須在每次定期的再次評量後，根據患者各自的恢復情況調整最適合的治療計畫。而考慮到患者在中風發病後一個月內的復健潛能為最高

時期 (P W Duncan, Goldstein, Matchar, Divine, & Feussner, 1992)，如何妥善利用再評結果、讓患者的治療潛力發揮到最大就是治療師的責任。反之，若經評量後發現患者的動作功能沒有達到預期的變化，治療師也必須立即採取對應措施，並與復健團隊討論可行的目標及療法。

舉例來說，當治療師以布朗斯壯動作功能恢復量表 (Brunnstrom Recovery Scale, BRS) 評量患者後得知患肢在協同模式 (synergy pattern) 下開始出現自主性動作 (voluntary movement)、且張力增加，此時便可開始利用治療手法誘發出更多的自主動作，而接下來的治療目標則可能修改為由協同模式中慢慢分離出獨立動作 (isolated movement)、降低張力等。

而最終的療效評量，除了可以了解患者在接受完整療程之後動作功能的恢復情形，也能根據評量結果制定患者未來的治療計畫。假若患者仍具有一定的復健潛能，則可安排患者持續接受門診治療；而當動作功能的療效評量結果顯示患者的治療成效已停留在高原期，可是其動作功能仍不足以達成日常功能獨立的話，治療師也會教導患者其他可行的、可被接受的代償方式，盡可能讓患者達到日常功能上的獨立。反之，如果患者已經恢復差不多，則可建議患者出院返家休養或轉介至一般安養機構等。

2. 預知患者在動作功能的恢復情形，亦能推測其它功能層面的預後 (prognosis)。

經過動作評量獲得患者動作損傷資訊後，便可根據既有的研究實證及經驗推估出患者的治療潛力，例如發病後 30 天的傅格—梅爾評估量表 (FMA) 分數能預測超過 86% 患者發病後六個月的動作功能變異量 (P W Duncan et al., 1992)，而掌握患者的預後除了能在治療過程中反覆檢驗治療成效是否理想之外，也能提供患者及家屬更為實際的衛教資訊。

另外，動作功能的預後也能推測出其他層面的預後。例如在接受治療後患者仍是膝關節肌力不足的話，則勢必會影響必須具備站立平衡能力才能執行的活動，如掃地、拖地、曬衣服等家務。根據動作功能評量的結果，治療師就可以提前為患者作準備，提供更加全面的照護訓練，提升患者的生活品質。

二、對於研究人員而言：

以中風復健動作評估量表 (STREAM) 相關的文獻為例，大致上可分為二類：(1) STREAM 之應用（在信任 STREAM 的基礎下應用 STREAM）、(2) STREAM 之心理計量驗證（驗證 STREAM 本身的評量能力與價值）。詳細文獻分類於表 1、表 2。

根據表 1、表 2 可知，如 STREAM 之類的中風復健動作評量於研究中，通常有二大類的重要性：

第一類為動作功能評量的應用，功能包含量化動作功能及其變化、預測動作功能的復原程度，或是當作校標來驗證其他工具的效度。

第二類為動作功能評量本身的心理計量驗證，驗證包含信度、效度、內在一致性及反應性等特性。然而驗證方式不盡相同，至少包含傳統的驗證法（將順序量尺逕自當作等距量尺加總計分）及以羅序分析（使用等距 logit 量尺）為主的二種方式。

結合上述 STREAM 之文獻回顧，以功能性來看，中風患者動作功能評量對於研究人員而言，具有以下 3 大重要性：

1. 以數據量化中風患者的動作損傷程度及其變化，達成不同研究目的。

(1) 探索性研究：對先前較少或沒有特定相關資訊的問題進行探索，以釐清或對欲探究問題之本質有初步認識，而進行的初始研究。

以” Factors determining functional ability of individuals with stroke in community ” (Vongsirinavarat & Hiengkaew, 2014) 研究為例，

使用 STREAM 作為評量工具之一，並將 STREAM 分數與 BI (Barthel Index) 分數比較以探討預測程度。也就是以 STREAM 當作預測因子，預測患者的日常生活功能 (Activities of Daily Living, ADL)。

結果顯示，STREAM、PASS-maintain (Postural Assessment Scale for Stroke patients, PASS)、PASS-change、Glasgow coma score 及患者年齡能解釋 85.6% 的 BI 變異量。其中又以 PASS-maintain 為最強的預測因子，可解釋 81.5% 的變異量，可知維持坐、站姿是 ADL 的關鍵。藉由探索性研究幫助研究者蒐集更多實證資料之後，未來便可統整出更有脈絡的研究推論。

由上述文獻可知，除了研究目的即為探討「動作功能」之外，也會有其它研究主題需要動作功能評量，像是探討中風患者發病後的生活品質 (Quality of Life) 及 ADL 等，均可發現患者的殘餘動作功能亦是重要的決定因子。

(2) 心理計量特性驗證研究：心理計量的定義為應用數學模式的方式計量出欲評估變項之特質，而心理計量特性驗證即為檢驗評量工具的心理計量特性，以確保評量所得數據是否可靠。

常見的心理計量驗證包含信度、效度、內在一致性及反應性等特性。此處依據計分方式可略分為二類：i) 將順序量尺直接當作等距量尺加總計分、ii) 使用羅序分析 (等距 logit 量尺)。

i) 將順序量尺直接當作等距量尺加總計分：

過去依照傳統測驗理論設計的評量工具常依照李克特量表 (Likert scale) 的形式，直接將本來僅有順序排列性質的順序量尺變項當作等距量尺，將回答正確或能達成任務的題數加總計分並做後續運算（如平均等）。

以”Reliability of scores on the Stroke Rehabilitation Assessment of Movement (STREAM) measure” (Daley K, 1999) 研究為例，

研究中 STREAM 的計分方式，即為將原本應為順序量尺 (ordinal) 之變項直接當作等距量尺加總計分。STREAM 的分數評量於操作上將動作完成度分為：0=無法執行動作、1=可完成部份動作、2=可完成所有動作，全部評量完再將各題目原始分數加總以代表該受試者的動作功能程度。量尺中的 0、1、2 僅代表程度上的「順序」，並非等距量尺。

然而，將順序量尺當作等距量尺具有三項缺陷：數學運算上的不當、題目依賴性 (item dependent) 及樣本依賴性 (sample dependent)。

數學運算上的不當：只有真正的等距量尺變項能執行合理的數學運算。而順序量尺上的數字僅能代表不同類別的標籤，雖然有順序排列，但無法確認相同單位差距是否代表相同的變化（例如 0 分至 1 分、1 分至 2 分之間皆為一個單位，但不見得代表相同變化）。如此一來，不僅是單一個案的分數加總會出現問題，後續如個案間的分數差異、同一個案不同時間的分數差異、同一組別個案平均分數等統計操作皆會因為數學運算問題而失準。

題目依賴性：意即對分數的詮釋 (interpretation) 會依賴於所執行的題目。由於順序量尺上的單位僅有順序性質、未必代表相同的距離或改變，因此即便

評量出相同的加總分數也無法確認個案的程度是否真的相同。例如，達成簡單的題目得到 2 分，達成較困難的題目也是得到 2 分，但事實上獲得這相同的 2 分應代表不同的意義。像這樣對分數的詮釋會依賴於所執行的題目，可稱之為題目依賴性。

樣本依賴性：以往直接將評量中獲得的原始觀察分數 (observed score) 用來判定個案的能力程度，造成對同一份測驗的難度、鑑別度等評量指標會因為受試者樣本不同而產生差異。舉例來說，假設在某測驗中達到等級 90% 代表高度能力程度。然而，同一份測驗中，同樣達到等級 90%、但來自不同群組的二位個案，可推測來自群組平均能力較高的個案，其能力自然會高於另一個群組平均能力低的個案，並且容易回答正確（或覺得簡單）。像這樣對於題目難度的估計會取決於樣本特性，稱為樣本依賴性。

整體而言，此類計分方式除了有數學操作上的缺陷外，還有個案能力及題目難度二者之估計會彼此互相影響、並不客觀。

ii) 使用羅序分析（等距 logit 量尺）：

有鑑於將順序量尺直接當作等距量尺會產生數學運算上的不當、題目依賴性及樣本依賴性等問題，近來有學者提倡使用真正的等距量尺計分。其中一種方式，即藉由羅序分析 (Rasch analysis) 改善上述三項缺陷。

數學運算變得合理：唯有當測驗分數真正以等距量尺計分時，數學運算才合理。當評量數據符合羅序分析之預期時，便可將具有順序性質的原始分數轉換為具有等距性質的 logit 分數。

解決題目依賴性：此處指的是題目分數配置固定、無法反映實際題目難易度之分數詮釋問題。羅序分析能依據每個題目的難易度配予對應的等距 logit 分數，因此答對每個題目都能獲得應有的配分，可解決分數詮釋會依賴於題目的問題。

解決樣本依賴性：以往判定題目難易度的方式，會因為個案能力程度不同而影響。以答對率低的題目為例，有二個可能原因，「題目難度較高」、或「個案能力較低」。然而，羅序分析能在計分方式同樣使用 logit 等距量尺的情況下，以羅序模式 (Rasch model) 估計每個題目之「題目難度」參數。如此一來，題目難度之估計就不會因為個案能力因素的加入而混淆，可解決樣本依賴的問題。

羅序分析近年廣被採用來驗證評量工具，並藉以取得等距 logit 分數。上述提及的「估計題目難度參數」等內容，其相關理論及使用方式會在後面的章節另行詳細介紹。

(3) 動作功能療效驗證研究：為中風患者經過特定復健治療後，檢驗該療法對於患者動作功能恢復的效果。

以「鏡像治療對慢性腦中風患者上肢動作功能之療效」(林嘉皇, 2013) 研究為例，

i) 證實療效並探究療效差異背後可能的原因。

研究結果顯示，每天三十分鐘的傳統治療加上一小時的鏡像治療介入，二週後即可改善慢性腦中風患者的上肢功能表現。

使用相同動作評量工具（此處為 FMA）所得到的數據，研究人員可交叉比對並推論進步幅度差異的原因，可能是人口學資料、中風發病後時間長短、治療人員、治療方法、治療劑量、執行場所甚至是評量人員等不同因素影響。

ii) 確立更加實際可行的研究目標。

以 FMA 的上肢動作功能細項來分析，可發現鏡像治療主要成效是改善遠端精細（腕及手部）動作，改善幅度達 31%；近端（肩及手臂）動作從結果來看也有改善，但和控制組沒有顯著差異。若經比較相關研究後發現皆為相似結果的話，研究人員便可預期之後鏡像治療的療效目標，於中風患者上肢遠端的動作功能進步幅度會比較大。

2. 篩選受試者或區分不同實驗組別，達成研究需求。

根據想探討的「研究目的」所需，而將「受試者動作能力」列入收案或排案條件，否則可能出現受試者無法順利參與實驗的情形。例如探討踩車運動訓練對於中風患者下肢動作恢復之影響，則必須於研究開始前以 STREAM 之下肢次量表篩選出可執行該訓練的患者；又或者想要探討機器輔助治療 (robot-assisted therapy) 於不同程度動作損傷之中風患者的療效，勢必要先執行受試者的動作功能評量之後才能分組。

3. 呈現受試者的動作功能變化，以執行後續分析探討。

針對患者進行長時間的定期評量蒐集資料，紀錄同一群患者在不同時期的動作功能演變，屬於組內變化。例如中風研究中常見的縱貫性研究 (longitudinal study)，於研究期間會執行連續性的動作功能評估，由於研究資料

涵蓋了不同時間點，研究人員便可依據結果分析動作功能隨著時間演進上的變化，以及可能的影響因子、患者的復健成效與持續情形等等。

以 "Multivariate analysis of stroke rehabilitation outcomes : A pilot study in a Chinese population" (Chen, Liu, Li, & Quiben, 2006) 研究為例，

研究方法為從 1999 年六月至 2002 年六月之間，以 STREAM 施測 92 位中風患者的動作恢復程度，再以迴歸分析去探究患者年齡、性別、患側（左或右）、中風類型（出血性或缺血性）及中風發病後至開始復健之時間間距等因子與動作功能恢復之間的關係。結果顯示中風發病後至開始復健之時間間距為最強的預測因子。

整體而言，不論是臨床治療或是學術研究皆需仰賴良好信效度的動作評量，以呈現客觀的動作功能，並可進一步用來探索最適切的訓練參數、研擬研究或治療計畫，以及預測未來的動作功能。因此，對於目前常用於中風後動作功能治療或研究之評量工具，皆需要確認其是否能確實量化損傷並呈現恢復的歷程。

第三節 簡介羅序分析

一、羅序模式及項目反應理論 (IRT)：

羅序分析，意指使用羅序模式 (Rasch model) 驗證測量工具。羅序模式為丹麥學者 Georg Rasch 於 1960 年所創，可視作項目反應理論 (item response theory, IRT) 的分支、或者數學上等同於 IRT 中的單參數模式 (one-parameter logistic model, 1PLM) (Wohlk Olsen, 2003)。

IRT 由美國學者 Lord 及 Novick 於 1968 年所創，相較於古典測驗理論 (classical test theory) 而言屬於比較新的測驗理論，常應用於教育、心理及醫學等相關領域中。

IRT 具有二個重要的假設：「單向度 (unidimensionality)」及「局部獨立性 (local independence)」。

1. 單向度是 IRT 各種模式中最常用的基本假設，意指測驗中的所有題目都評量到同一種潛在特質 (latent trait) (指的是評量者想評量的建構，可以是個案的能力、態度、特徵或特質)。此外，僅有同一種潛在特質也代表該評量工具僅有一個建構，所有題目都在測量該建構。

2. 局部獨立性指的是當影響測驗表現的能力 (潛在特質) 固定不變時，個案對測驗中某一題目之反應與其他任一題目之反應為獨立或不相關的。此意謂著個案能力才是唯一影響個案在測驗試題上做反應的因素。

此外 IRT 又被稱作「潛在特質理論」(latent trait theory) (Lord, 1968)，原因來自於其具有二個核心理念：

1. 個案在某一測驗題目的作答反應，可由其潛在特質加以預測或解釋。
2. 個案之作答反應及該潛在特質間的關係，可透過一條連續性遞增的函數加以詮釋，此函數稱作試題特徵曲線 (item characteristic curve, ICC)。(ICC 便是該題答對機率對個案潛在特質的迴歸線，後續會詳細說明)。

綜上所述，IRT 重視個案的潛在特質，透過描述個案對於每個題目的答對機率，可推論潛在特質及題目特性 (難度、鑑別度) 如何影響作答反應。

後來的研究學者按照 IRT 之假設及核心理念發展出多種適用於不同測驗情境的模式：依據參數個數的不同，可分為單參數模式、及二參數模式 (two-parameter logistic model, 2PLM)；依據計分量尺的不同，可分為二分題 (dichotomous items) 模式、及多元計分 (polytomous) 模式；而依據作答方式的不同，可分為評等量尺模式 (rating scale model)、及部份得分模式 (partial credit model) 等。

羅序模式的設計其實就等同於 IRT 之中的單參數模式，也擁有前述提及 IRT 之核心理念及假設，雖然二者之間仍有部分原理及方法學上的差異，但於基本應用上是類似的。

之後有一派學者承襲了 Rasch 精神發展了因應不同情境的多種模式，通稱為 Rasch 家族 (王文中, 1997)。為求容易理解，接下來先介紹 Rasch 家族之中最簡單的二分題模式，其適用的作答反應必須為二元計分 (對或錯、是或否、成功或失敗、同意或不同意等)。

二、羅序模式之二分題模式：

以下會先介紹二分題模式之數學式，並且說明依據羅序模式畫出之 ICC。

1. 數學式 (Rasch, G., 1960)

$$P_{ni1} = \frac{\exp(\theta_n - b_i)}{1 + \exp(\theta_n - b_i)}$$

也能轉換成另一種形式： $\log\left(\frac{P_{ni1}}{P_{nio}}\right) = \theta_n - b_i$

(1) P_{ni1} ：第 n 個人在第 i 題答對的機率。

此處以最基本的二分題為例，以 1 代表正確、0 代表錯誤。故 P_{ni0} 則為第 n 個人在第 i 題答錯的機率，且 $P_{ni1} + P_{ni0} = 1$ 。

(2) θ_n ：第 n 個人的能力或稱為潛在特質。

(3) b_i ：第 i 題的難度。

(4) $\frac{P_{ni1}}{P_{ni0}}$ ：同一個案、答對題目機率除以答錯相同題目機率，又稱為勝算

比。

若個案能力 (θ_n) 大於題目難度 (b_i)，則答對機率大於 50%；反之，若個案能力 (θ_n) 小於或等於題目難度 (b_i)，則答對機率變為小於或等於 50% (Wright, 2000)。

由上述公式可知，羅序模式認為個案回答某一題目是否正確之機率取決於個案能力和題目難度之間的比較。也可以說，只要了解個案能力及題目難度，便能知道該個案在某題目的答對機率為多少。

2. ICC

依據羅序模式之假設，個案的作答反應及其潛在特質的關係都能由 ICC 詮釋，ICC 便是該題目之答對機率對個案潛在特質的迴歸線，如圖 1。因此，每一個題目都能畫出自己的 ICC。

ICC 具有以下四個特性：

(1) 對同一個題目而言，當個案的能力值越高，其答對該題目的機率也會隨之變高，代表 ICC 為單調遞增 (monotonic increasing) 曲線。



(2) 因為個案答對機率取決於個案能力和題目難度之間的比較，故 ICC 形狀亦由個案能力及題目難度決定。

(3) ICC 具有預測作答反應的功能。憑藉著 ICC 可以預測某能力程度的個案，在回答某種難度的題目時，其答對的機率是多少。

(4) 於應用層面上，ICC 有助於評量人員針對不同能力程度的個案，挑選適合難度的題目，此即為適性測驗 (adaptive testing) 的精神。只要挑選對個案而言最適當的題目就能獲得最多的資訊，也能避免讓個案執行對其太過困難、或者太過容易的題目，同時減輕個案、治療人員及研究人員的負擔。

三、優勢：

相較於傳統的測驗理論，羅序模式具有三個優勢：「計分轉換為等距量尺」、「解決題目依賴性及樣本依賴性問題」、及「個案能力及題目難度可立即、直接比較」。

1. 計分轉換為等距量尺：

透過羅序分析，只要測驗數據符合羅序模式的預期，即可將原始分數轉換為等距量尺之羅序分數 (logit) (Rasch, G., 1960; Wright, 2000)。

轉換為等距量尺後，能夠進行合理的數學運算，分數間的比較、後續統計分析才有意義。

2. 解決題目依賴性及樣本依賴性問題：

前一章節提過，羅序分析能同時估計個案能力及題目難度參數，且二者之估計彼此不會互相影響，解決題目依賴性及樣本依賴性問題。

3. 個案能力及題目難度可立即、直接比較：

以往的測驗理論中，個案能力及題目難度無法立即、直接比較。依據以往的測驗理論，個案能力之判定是來自答對題目分數的加總，加總分數高者被視為能力較高；而題目難度之判定是來自答對該題目的百分比，答對率低者被視為難度較高。然而，即使發現某題目答對機率較低，也無法直接判斷是因為題目本身難度較高、或是答題者的能力太低造成的。

羅序模式能夠同時估計個案能力及題目難度，並將二者置於共同的等距 logit 量尺上。因此，評量者能夠直接、立即地比較個案能力及題目難度，也能清楚指出二者間的差距。

個案能力及題目難度可立即、直接比較有二個好處：

(1) 就題目的角度而言，協助評量者了解哪些題目對於個案是可以答對的、或比較困難的，也能更清楚辨別個案能力及題目難度之間的差距有多大。

(2) 就測驗的角度而言，可幫助評量者檢視整份測驗的難度是否適合該群個案。舉例來說，如果分析結果發現多數個案的能力值皆高於測驗中所有題目的難度值，則可視為該測驗無法有效地評量能力較高的個案，因為測驗中的題目難度對能力高者而言偏易。

四、應用：

羅序模式於分析、驗證評量工具上擁有三種應用方式：「可同時估計個案能力及題目難度參數」、「可檢驗評量工具是否為單向度」、「可檢驗評量工具之羅序信度 (Rasch reliability)」。

1. 可同時估計個案能力及題目難度參數：

羅序模式能將原始作答分數轉換為 logit 分數，估計出個案能力；也能同時以 logit 量尺估計出每一題目之難度參數。羅序模式估計個案能力或題目難度的參數估計法有很多種，常用的包含最大近似估計法 (maximum likelihood estimation)、邊際最大近似估計法 (marginal maximum likelihood estimation) 等。不過通常在實際的估計情境中，往往無法事先得知能力及題目的參數，因此必須同時進行估計。而其中常用的同時估計能力及題目參數的最大近似估計法，叫作聯合最大近似估計法 (joint maximum likelihood estimation, JMLE)。

2. 可檢驗評量工具是否為單向度 (適配度檢定, goodness-of-fit test)：

單向度是羅序模式的基本假設，若蒐集而來的樣本資料符合羅序模式的預期，則該評量工具便具有單向度的性質。換句話說，唯有選用的羅序模式適用於測驗資料時，單向度的假設才會成立。因此，評量者可以透過檢定資料及模式間是否具有滿意的適配度，以確認評量工具是否為單向度。

羅序模式使用「加權均方適配統計值 (infit mean square, infit MNSQ)」及「未加權之均方適配統計值 (outfit MNSQ)」當作適配度指標。當 infit MNSQ 及 outfit MNSQ 皆落於 0.60~1.40 區間代表符合羅序模式之預期 (Linacre, 1994)，也就是該題目具有單向度。

此外補充說明單向度的意義，當檢驗出評量工具為單向度時也代表其具有良好的建構效度 (construct validity)，因為當評量工具只評量一種潛在特質也代表本身僅具有一個建構，所有的題目都在評量此一建構。

3. 可檢驗評量工具之羅序信度：

羅序模式也能計算出評量工具的信度，不過其估計理念跟以往的信度不盡相同。相較於以往測驗理論假設所有個案都具有相同的測量誤差，在羅序模式中，測量標準誤會隨著個案能力程度不同而有所差異。

首先，不論是古典測驗理論或是羅序模式，信度計算方式皆為真實變異量 (true variance) 除以觀察變異量 (observed variance) 之比值，其中，分母的觀察變異量等於真實變異量及誤差變異量 (error variance) 之和。因此當測量誤差越小時，信度會越大。

以此計算出之信度在古典測驗理論中便是該測驗信度之核心指標。然而在羅序模式中更強調估計每位個案的精準度，當個案能力越接近中等程度時，會有較小之測量誤差、羅序信度則較大；反之，當個案能力程度較極端時（偏大或偏小），測量誤差則較大、羅序信度較小 (Boone, Staver, & Yale, 2014)。

此外，古典測驗理論計算之測驗信度僅屬於「個案信度 (person reliability)」；羅序信度則可估計「個案信度」及「題目信度 (item reliability)」，二者具有不同的涵義及應用方式。個案信度代表測驗能區辨個案並分為不同群體的程度，會因為個案能力範圍越廣、測驗題目數量越多而越大；題目信度則代表題目影響總分數變異量的程度，會因為題目難度範圍越廣、受測個案數量越多而越大。

有時候研究人員會將個案信度轉換為另一種估計指標，稱作「個案區辨指標 (person separation index, PSI)」，PSI 的數值代表該測驗能區辨不同能力個案群之群體數。轉換標準為：個案信度為 0.90 時，PSI 等於 3~4 (可依個案能力程度分成三到四群)；個案信度為 0.80 時，PSI 等於 2~3；個案信度為 0.50 時，PSI 等於 1~2。一般將個案信度大於 0.90 視為具有良好信度 (WJ, 1992)。

第四節 常用於中風患者之動作功能評估量表優缺點分析

目前較常被提及或使用的中風患者動作評估量表包括：

1. 運動功能評估量表 (Motor Assessment Scale, MAS) (Carr, Shepherd, Nordholm, & Lynne, 1985)：為澳洲雪梨大學 Carr 和 Shepherd 二位物理治療師依據中風患者動作恢復的理論架構發展的評量工具，也是 Post-Stroke Rehabilitation Clinical Practice Guideline 唯一推薦的上肢評量工具。

MAS 包含九大項目：由仰躺側翻至健側 (supine to side-lying onto intact side)、由仰躺轉身坐在床緣等 (supine to sitting over side of bed)、坐姿平衡 (balance sitting)、坐到站 (sitting to standing)、行走 (walking)、上肢功能 (upper arm function)、手部動作 (hand movements)、進階手部活動 (advanced hand activities)、及肌肉張力 (general tonus)。每個項目評分均採 0~6 分的七等級制，滿分 54 分，得分越高代表動作功能表現越佳。總施測時間大約為 15~30 分鐘。

目前 MAS 以古典測驗理論驗證其心理計量特性之研究結果如下：



(1) 信度：具有極佳的施測—再測信度 (test-retest reliability)、施測者內／間信度 (interrater / intrarater reliability) (Carr, Shepherd, Nordholm, & Lynne, 1985)。

(2) 效度：具備良好的內容效度 (content validity) (Carr et al., 1985)、建構效度 (construct validity) (Lannin, 2004)，與傅格—梅爾評估量表 (Fugl-Meyer Assessment, FMA) 比較下之共同效度 (concurrent validity) 亦有不錯結果 (Malouin, Pichard, Bonneau, Durand, & Corriveau, 1994)。

(3) 反應性：MAS 上肢次量表 (UL-MAS) 及動作研究上肢測驗 (Action Research Arm test, ARAT) 之總分反應性指標為適中且相似，研究結果支持二者均適合評量中風個案上肢功能之復原程度 (Hsueh & Hsieh, 2002)。

另外，MAS 以羅序分析之研究結果：

以急性／亞急性患者在 UL-MAS 之羅序分析結果顯示，呈現單向度性質、良好的適配度 ($\chi^2 = 20.451, p = 0.252$)、極佳的 PSI (0.96)、扣除手腕橈側偏移 (radial deviation of the wrist) 項目後沒有出現差異試題功能 (differential item functioning, DIF)，且上肢功能及手部動作二次量表皆有題目難度階層 (item difficulty hierarchy)，進階手部活動次量表則不符合此特性 (Miller, Slade, Pallant, & Galea, 2010)。整體而言，UL-MAS 於羅序分析驗證後，顯示其具備良好的心理計量特性。

MAS 優點除了前述提及已驗證過之心理計量特性多為良好外，其量化分數也能應用在研究分析或療效探討，有助於學術與臨床上的使用。

然而 MAS 仍具有以下缺點，包含：



(1) 存在天花板及地板效應 (floor and ceiling effects) (I-Ping Hsueh, 2002; Lannin, 2004; Williams, Galea, & Winter, 2001)。

(2) 部分題目有 DIF、及進階手部活動次量表不符合題目難度階層。

(3) 預測效度及反應性等特性仍未被完整驗證。

(4) 實際應用上的考量，由於其評估方式需要不少道具及設備的轉換，因此較不受到研究者與臨床者的青睞。

綜上所述，MAS 之心理計量驗證不完整、部分題目存在爭議（例如有 DIF 或不符合題目難度階層等問題），以及實際應用時道具及設備轉換較繁瑣等因素，影響 MAS 之使用。

2. 布朗斯壯動作功能恢復量表 (Brunnstrom Recovery Scale, BRS)

(Signe Brunnstrom, 1983)：承襲 1951 年 Twitchell 提出中風患者的肢體動作恢復具有某種規律性順序之想法，由美籍瑞典物理治療師 Brunnstrom 提出的一種主要適用於中風患者復健的方法。Brunnstrom 認為中風患者在中風時會依序經過六個階段，分別為：

Stage I-沒有任何動作產生，肌肉張力也很低；

Stage II-開始出現痙攣 (spasticity)；

Stage III-開始有自主性動作 (voluntary movement)，但只能遵循協同模式 (synergy pattern)，又分為 flexor synergy 跟 extensor synergy；

Stage IV-開始有獨立動作 (isolated movement)，肌肉張力也開始下降；

Stage V-幾乎可以不受 synergy pattern 的影響，肌肉張力逐漸接近正常，可以做獨立動作；

Stage VI-coordination 已接近正常，肌肉張力也接近正常，只有在速度快或做交替動作時才會出現異樣。

上述 BRS 六階段，每個人在各階段之時間可能不同，也有些人停留在某個階段而未再進行下去。

BRS 施測項目分成上肢、下肢及手部動作三大類，每一類評分均採 Stage I~ VI 之六等級制，依據患者動作恢復情形評為各種不同階段。

目前 BRS 以古典測驗理論驗證其心理計量特性之結果，不論國內外研究均不多。目前知道此量表具有良好的同時效度，但施測者內及施測者間信度之一致性偏低(祝旭東、王淳厚、謝清麟、陳美香、陳瓊玲, 1996)。

另外，BRS 之羅序分析驗證結果：在 2016 年 Huang 等以羅序分析檢驗 BRS，結果呈現出單向度、上肢次量表及整體量表符合羅序模式 (infit / outfit MNSQ = 0.57~1.40)、以及上肢次量表與整體量表皆具有高度的羅序信度 (0.91~0.92) 及足夠之反應性 (effect size, ES=0.35~0.41；standardized response mean, SRM=0.85~0.99) (Huang et al., 2016)。

BRS 優點包含施測時間短、施測方式簡單，再加上本身依據中風患者動作恢復歷程而設計，能直接呈現患者的動作功能恢復到何種階段，是目前國內復健醫學界最常用於評估中風患者動作功能的評估工具。

而 BRS 缺點除了前述提到心理計量驗證較缺乏之外，由於設計理念較著重在將患者動作能力分期及協同模式，並未全面評量整體動作功能的恢復，如肌力、動作協調、速度等，因此即便被評為同樣階段的患者，實際動作品質或表現可能仍有差異，評量內容較不足。

3. 傅格—梅爾評估量表 (Fugl-Meyer Assessment, FMA) (Fugl-Meyer, Jaasko, Leyman, Olsson, & Steglind, 1975)：為瑞典學者 Fugl-Meyer 根據 Twitchell 與 Brunnstrom 的中風患者動作恢復進程理論設計而成，且除了將原有 BRS 的評估範圍轉換成更細緻的量化分數之外，也針對其他不同面向的症狀做評估。

依據 Leire Santisteban 等研究顯示，FMA 是研究學者最常用於評估中風患者上肢功能恢復情形的評量工具，在 Pubmed、CinHal 及 PeDRO 等搜尋引擎裡，佔了所有中風結果測量相關之研究的 36% (Santisteban et al., 2016)。

FMA 一共包含三大面向：動作功能及平衡能力、感覺、關節活動程度與疼痛，評分採 0、1、2 的三等級制，表現越佳，分數越高。評估項目共 113 項，滿分為 226 分。總施測時間大約為 30~60 分鐘。三個面向中以動作功能及平衡能力部分的比重最高（動作功能佔 44.2%、平衡能力佔 6.2%），關節活動程度與疼痛次之（各佔 19.5%），感覺部分最低（10.6%）。

此外，依據 FMA 動作功能分數可劃分不同程度之障礙標準：動作功能得分小於 50 分者訂為嚴重動作障礙、50~84 分為明顯動作障礙、85~95 分為中度動作障礙，而 96~99 分為輕微動作障礙。

目前 FMA 以古典測驗理論驗證其心理計量特性大多有不錯的結果：具備良好的再測信度、施測者內／間信度、內容效度、建構效度、同時效度、內在一致性及反應性 (Alt Murphy, Resteghini, Feys, & Lamers, 2015; Pamela W. Duncan, Propst, & Nelson, 1983; Fugl-Meyer & Jääskö, 1980)。

FMA 優點包含：操作上簡單、擁有良好的心理計量特性、敏感度佳、適用於各個恢復階段之中風患者、評估面向多元完整，有助於使用者針對不同範疇

的症狀構思應對的療法或計畫，而且 FMA 之量化分數也有助於學術研究上的數據分析比對。

然而於實際應用層面上 FMA 仍有缺點，由於其施測項目較多，若完整評量所耗費時間過長，因此臨床上醫療人員其實較少使用。

4. 中風復健動作評估量表 (Stroke Rehabilitation Assessment of Movement, STREAM) (Daley K, 1997)：為加拿大物理治療師 Daley 等人發展之量表，總共 30 個評分項目，其中又分為三個次量表，包含了上肢動作 (upper-limb movements)、下肢動作 (lower-limb movements) 及行動能力 (mobility)。

上、下肢動作次量表各有 10 個評分項目，評分採 0、1、2 的三等級制。行動能力亦有 10 個評分項目，不過評分採 0、1、2、3 的四等級制。三個次量表合計總分為 70 分。總施測時間約 20 分鐘。

STREAM 以古典測驗理論驗證心理計量特行結果：

(1) 信度：最初 Daley 等量表發展者便在 1999 年發表了關於 STREAM 信度的研究，結果顯示內在一致性良好(Daley K, 1999)。在 2002 年 Wang 等也證實了 STREAM 具有中度至極佳的施測者間一致性、高度的施測者間信度(Wang CH, 2002)。此外也具有高度的再測信度(李素, 2006)。

(2) 效度：與 FMA 相較下之同時效度及與巴氏量表 (Barthel Index, BI) 相較下之收斂效度也呈現良好結果 (Wang CH, 2002)。此外，STREAM 具有良好的預測效度—發作後一周之 STREAM 分數能預測急性期中風患者出院後能否回到家庭、及中風三個月後的走路速度及日常生活功能表現 (Ahmed et al., 2003)。

(3) 反應性：STREAM 具有令人滿意之反應性，上肢、下肢及動作次量表之 SRM 各為 0.57、0.72、1.00 (Huang YJ, 2015)。

另外，STREAM 之羅序分析驗證結果顯示，扣除上肢次量表中聳肩（兩側）及拇指尖碰食指尖、下肢次量表中的患腿外展，三個次量表其餘的項目皆為單向度 (Hsueh IP, 2006)。

STREAM 的優點除了前述提及良好且完整的古典心理計量驗證之外，也因操作簡單、施測時間短，相當適合臨床者及學術研究者使用，是極有發展潛力的評量工具。不過由於 STREAM 較晚發展，與其他評量工具相比下較未受到青睞，目前不論臨床上或相關研究上的使用仍然不多。

若想增進臨床及研究人員使用 STREAM 之頻率，或許可藉由更完善地驗證 STREAM 之心理計量特性，為其提供更加穩固的實證。然而，目前 STREAM 除了前述提及以古典測驗理論為基礎的驗證之外，以羅序分析驗證之研究結果似乎尚且不足，仍需要更多的實證支持。

第五節 中風復健動作評估量表動作次量表羅序分析結果之不足

承前述所提，STREAM 目前已具有許多以古典測驗理論為基礎之心理計量驗證，尤其是評量工具較為重視的信度、效度，及反應性等要素，代表 STREAM 可準確評量不同階段中風個案的動作功能。再加上操作簡便、施測時間短，是值得推薦使用的評量工具。因此，若有更加充分、完整的證據支持，將能提升其臨床及研究的使用價值。

以羅序分析驗證評量工具可提供不同於古典測驗理論之實證資訊。羅序分析的用途包含：(1) 可將各題目加總分數轉換成等距量尺的羅序分數，改善原本評量工具以順序量尺計算總分的限制。(2) 能同時估計各題目難度參數及各個受試者的能力，並置於共同的 logit 量尺上，因此二者可立即、直接比較。(3) 符合羅序模式預期之施測題目則具有單向度（屬於建構效度），且也能依據個案不同能力程度計算出各自的測量誤差，並估計出更精準之羅序信度等。

然而以目前文獻回顧所知，至今僅有一個研究使用羅序分析驗證 STREAM 之心理計量。其結果顯示上肢次量表一聳肩（兩側）及拇指尖碰食指尖、下肢次量表一患腿外展共三個項目並不符合羅序模式的預期，代表這些題目並非為單向度，題目測量到不只一個主要的潛在特質 (Hsueh IP, 2006)。而扣除上述三個項目後，上肢／下肢動作、行動能力三個次量表均具有良好的羅序信度，分別是 0.86、0.91、0.97。

此外，該研究亦存在另一限制，其徵召樣本多為中風後超過一年的慢性期患者，因此對於 STREAM 用於其他恢復階段患者之情形仍不清楚。

整體而言，目前 STREAM 之羅序分析驗證仍不足。因為單一研究成果仍無法呈現充分且穩定的證明，再加上先前研究結果具有明確的樣本限制，無法提供亞急性期患者的資訊；若欲增進 STREAM 於臨床及學術研究之使用，需要更多的證據支持。因此，本研究目的為交叉驗證 STREAM 上／下肢動作次量表於中風患者之心理計量特性，包含建構效度（單向度）、羅序信度、本研究及 Hsueh 等研究題目難度排序之相關性，以及依據本研究參數及 Hsueh 等研究參數各別估計個案分數之相關性。

第二章 研究方法



第一節 研究對象

本研究採用次級資料分析法，資料來源為已發表研究“Predicting Recovery of Voluntary Upper Extremity Movement in Subacute Stroke Patients with Severe Upper Extremity Paresis”所收集之資料 (Koh et al., 2015)。該研究於一家醫學中心徵召腦中風受試者，期間為從 2009 年 1 月至 2012 年 1 月。該研究計畫經國立台灣大學附設醫院之研究倫理委員會審核通過，且所有受試者均完成填寫受試者同意書。

本研究之收案標準包含：

(1) 診斷為第一次的缺血性腦中風或顱內出血性腦中風，且經過電腦斷層掃描 (computed tomography, CT) 或核磁共振造影 (magnetic resonance imaging, MRI) 確診；(2) 具有溝通能力且能遵循一步驟指令。

另外，排除標準包含：

(1) 具有其他中樞或周邊神經性疾病（如腦瘤或巴金森氏症等可能影響動作控制）；(2) 接受住院復健治療時間少於七天（考慮到患者可能尚未受到足夠的復健治療）。

第二節 研究工具

使用的研究工具包含 STREAM 中的二個次量表（上肢動作及下肢動作）以及美國國家衛生研究院腦中風量表（National Institute of Health Stroke Scale, NIHSS）。

STREAM 的上、下肢動作次量表各有 10 個評分項目，各項目之施測動作與姿勢詳見表 3。二者評分皆採用 0、1、2 的三等級制：0 分表示無法執行動作（看不到任何動作，包括顫動或輕微的移動）；1 分表示可完成部分動作，或可完成所有動作，但其動作模式與正常人之間仍有顯著差異；2 分表示可完成所有動作，且其動作模式與健側相似。二者合計總分為 40 分。

美國國家衛生研究院腦中風量表（NIHSS）是中風領域最著名的嚴重程度指標，適用於最初評估及後急性期的照護，也被廣泛用於臨床試驗的預後評估。評估內容包含：意識程度、眼球運動、視野、面部肌力、上下肢運動、肢體運動失調、感覺功能、語言功能、構音困難及忽略。總分為 42 分，分數越高代表中風嚴重度越高。NIHSS 具備良好的信度與效度（Goldstein, Bertels, & Davis, 1989; Lyden, Lu, Levine, Brott, & Broderick, 2001）。

第三節 研究步驟

由於本研究採用次級資料分析法，故將研究步驟分為二部分說明：(1) 原始資料來源之收案步驟、(2) 本研究之研究步驟。

1. 原始資料來源之收案步驟 (Koh et al., 2015)：

該研究於一家醫學中心徵召腦中風受試者。收案過程中，由一位受過訓練的研究助理負責篩檢每位將入住復健病房的個案。篩檢後符合收案標準之個案，會在入住後七天之內接受 STREAM 評量 (baseline assessment)，並於出院前三天之內再次接受同一位研究助理之評量 (final assessment)。

此外，個案之人口學資料、NIHSS 分數及其他共病症等病歷皆由該位研究助理記錄。依據記錄，共病症包含高血壓、糖尿病、高脂血症及心臟疾病。

除了二次 STREAM 評量之外，研究助理不會再與個案接觸，以免產生任何評量結果偏差。

2. 本研究之研究步驟：

首先，本研究篩選原始資料中符合收案標準之個案。接著進行羅序分析之評等量尺模式驗證，最後再與目前僅有一篇的 STREAM 羅序分析研究 (Hsueh IP, 2006) 進行交叉驗證，比較建構效度之單向度、羅序信度、本研究及 Hsueh 等研究題目難度排序之相關性，以及本研究與 Hsueh 等研究參數所估計個案分數之相關性。

第四節 資料分析

本研究之資料分析分為二部分：1. 人口學資料分析。2. 羅序分析。

1. 人口學資料分析：



本研究以描述性統計分析（平均值、標準差與百分比等）呈現樣本之特性。以 SPSS 17.0 進行統計分析。

2. 羅序分析：

本研究採用羅序家族中的評等量尺模式各別驗證 STREAM 上／下肢動作二次量表之模式適配度，因為評等量尺模式適用於多元計分（如二次量表皆為 0、1、2 分）、且所有題目的各量尺之評分標準均相同（0 分表示無法執行動作；1 分表示可完成部分動作，或可完成所有動作但仍與正常人有顯著差異；2 分表示可完成所有動作）(Andrich, 1978)。

本研究使用 Conquest 軟體進行羅序分析，分析 STREAM 上／下肢動作次量表的四個統計指標，包含：(1) 各題目之羅序模式適配度（建構效度之單向度驗證）、(2) 羅序信度、(3) 本研究及 Hsueh 等研究題目難度排序之相關性，以及(4) 本研究參數及 Hsueh 等研究參數各別估計個案分數之相關性。

(1) 各題目之羅序模式適配度（建構效度之單向度驗證）：

題目適配度分析用於檢驗 STREAM 上下肢動作次量表是否為單向度，若樣本於各題目之得分資料符合羅序模式之預期，則這些題目必具有單向度的特性 (Hattie, 1985)。

本研究以「加權均方適配統計值 (infit mean square, infit MNSQ)」及「未加權之均方適配統計值 (outfit MNSQ)」當作適配度指標 (Linacre, 1994)。infit MNSQ 對於題目難度較接近個案能力的範圍較敏感，outfit MNSQ 則對題目難度遠大於或小於個案能力的範圍較敏感 (Bond TG, 2001)。



二者 (infit MNSQ 與 outfit MNSQ) 判斷標準值為：若介於 0.60~1.40 之間，為可接受的範圍，代表個案在該題目的反應結果符合羅序分析—評等量尺模式預期的範圍、該題目符合單向度 (Linacre, 1994)。

此外，適配度不佳 (misfit) 的情形有二：i) 二者 MNSQ 數值大於 1.40 時，代表個案在該題目的反應不符合羅序模式之預期，該題目可能不具有單向度。ii) 二者 MNSQ 數值小於 0.60 時，代表個案在該題目的反應過度符合羅序模式之預期，該題目所能提供的訊息有限 (Linacre, 1994)。

當出現任何適配度不佳的題目時，研究者先刪除 infit MNSQ 數值大於 1.40 且差距最大的題目，刪除後再次執行羅序分析。假若分析結果仍有題目的 infit MNSQ 數值大於 1.40，即刪除之，直到所有題目之 infit MNSQ 皆小於 1.40。接著刪除 outfit MNSQ 數值大於 1.40 的題目。再依據上述原則刪除 infit / outfit MNSQ 數值小於 0.60 的題目，直到最後所有題目均達到可接受之適配度範圍 (0.60~1.40)，再停止羅序分析。

(2) 羅序信度：

羅序信度可經由羅序分析檢驗量表所得分數之標準誤估計，羅序信度大於 0.70 及 0.90 各自代表此量表適合群組比較以及個人比較 (Mokkink et al.)。學術研究常以群組比較為主、臨床人員則常以個人比較為主。本研究分別估計整體與個別樣本於上肢、下肢動作次量表的羅序信度。

(3) 本研究及 Hsueh 等研究題目難度排序之相關性：

此部分包含二個步驟，

i) 估計每個題目之題目難度參數。可分為二種：評分等級難度、平均困難度。評分等級難度，意指每個題目中三個評分等級（0、1、2分）各自的難度（理論上評分等級難度應為單調遞增，等級越高，所對應的難度也隨之越高）。題目的平均困難度，則為各題目之三個評分等級難度取平均值。

ii) 比較本研究及 Hsueh 等研究之題目難度排序相關性。以斯皮爾曼等級相關係數 (Spearman's rank correlation coefficient, Spearman's rho) 驗證二研究估計題目難度參數（本研究驗證平均困難度）之相關程度。Spearman's rho 值範圍介於 -1~1 之間，rho 值為正負 1 代表完全相關；當介於正負 0.6~0.9 時稱為高度相關；介於正負 0.3~0.6 時稱為中度相關；介於正負 0.3 之間時稱為低度相關；等於 0 時則表示無相關。

(4) 本研究參數及 Hsueh 等研究參數各別估計個案分數之相關性：

以本研究參數估計本研究中所有個案之個案分數，另依據 Hsueh 等研究參數估計本研究之個案分數，再以皮爾遜積差相關係數 (Pearson product-moment correlation coefficient, Pearson's r) 驗證二研究估計個案之分數是否相似，確認其相關性。Pearson's r 值若高於 0.80 代表相關程度良好。

第三章 結果



一、樣本特性

表 4 為本研究樣本之基本人口學資料。本研究共收集 302 名中風患者之中風復健動作評估量表資料。其中包含 180 位男性、122 位女性，平均年齡為 65.3 歲（標準差為 13.7）。

二、羅序分析

1. 各題目之羅序模式適配度（建構效度之單向度驗證）：

本研究以 MNSQ 數值介於 0.6~1.4 為標準，分析結果顯示上肢次量表有三個題目（舉手朝向天花板並出拳、伸直手肘、聳肩）不符合模式預期；下肢次量表則有一題（沿床面彎曲下肢）不符合預期。

刪除上述適配度不佳之題目後，再各自執行羅序分析。結果顯示，剩餘 7 題之上肢次量表及剩餘 9 題之下肢次量表均符合羅序模式預期，且二量表之 MNSQ 數值範圍分別為 0.63~1.36 及 0.66~1.17，表示二量表均具有單向度特性。

2. 羅序信度：

表 5 列出上下肢次量表刪除適配度不佳之題目後，最終之羅序信度。二次量表之信度均大於 0.90，代表具有良好信度，能產生相當穩定之估計結果。

此外，羅序信度與以往信度另一不同點在於，依據個案能力不同而有不同

的測量標準誤及羅序信度。表 6、表 7 各自列出本研究樣本於上、下肢次量表的個案信度分布情形，可知本研究於上、下肢次量表之個案信度分布皆相當高（均為 243 人大於 0.90；59 人介於 0.80~0.90）。



如圖 2 所示，全體個案之個案信度均在 0.80~0.90 之間。其中，能力屬於中等的個案，其個別測量標準誤較低，因此個案信度較能力位於二極端者更高。

如圖 3 所示，全體個案之個案信度亦均在 0.80~0.90 之間。同樣地，能力屬於中等的個案，其個案信度較能力位於二極端者略高。

3. 本研究及 Hsueh 等研究題目難度排序之相關性：

此分析包含二個部分：(1) 估計每個題目之題目難度參數。(2) 比較本研究及 Hsueh 等研究題目難度排序之相關性。

由於本研究及 Hsueh 等研究之刪除題目不盡相同，故僅比較二者共同保留題目之難度參數。表 8、表 9 分別為二篇研究之上、下肢次量表難度參數與排序。

研究者預期二篇研究之題目難度排序應該相當接近，理由是中風患者之動作功能恢復有一定的恢復歷程，即便因為受試者特性或能力差異使得「題目難度參數的數值」不完全相同，但每個題目間的難度排序仍應有固定的趨勢，故預期難度排序之相關性為中度至高度相關。

二篇研究題目難度排序之相關性，採用 Spearman's rho 分析二篇研究共同保留題目之難度排序（上肢 6 題、下肢 8 題）。根據表 10 顯示，二研究之上肢次量表難度排序相關性為 -0.31、屬於低度相關，下肢次量表難度排序相關性為

0.48、屬於中度相關。



4. 本研究參數及 Hsueh 等研究參數各別估計個案分數之相關性：

研究者預期二篇研究參數估計分數之相關性應該介於中度至高度相關，原因是由於羅序分析的特性：個案作答反應來自於個案能力及題目難度之比較，故同一群個案、操作同樣的評量題目應有固定之作答反應，故依據作答反應結果計算之個案分數也應有一定相關性。且共同保留之題目越多，所估計之整體分數相關性也應越高。

此部分先以二篇研究參數估計同一群個案（本研究樣本）之分數，再以 Pearson's r 分析二者分數間之相關性。根據表 11 顯示，以二篇研究之上肢次量表參數估計之個案分數相關性為 0.96，為高度相關；而下肢次量表參數估計之個案分數相關性為 0.62，為中度相關。

圖 4、圖 5 分別為上、下肢次量表以二篇研究參數估計同一群個案之分數散佈圖。可知二研究以上肢次量表保留題目的難度參數，所估計之個案分數呈現特定的線性關係，意即具有高度相關性；然而以下肢次量表題目難度參數所估計之個案分數，其分布則相對較分散，二分數間相關性也相對較低。

第四章 討論



一、建構效度（單向度）：

建構效度之單向度驗證是以各題目之羅序模式適配度為指標，若題目之適配度良好（MNSQ 數值介於 0.6~1.4 間），表示該題目之作答情形符合羅序模式之預期、該題目具有單向度。反之，若分析時發現題目之適配度不佳，則會逐一刪除直到所有題目之適配度均在可接受的範圍內。

1. 上肢次量表：

本研究刪除三個題目（舉手朝向天花板並出拳、伸直手肘、聳肩）之後，其餘題目均具有單向度。而過去 Hsueh 等研究則刪除二個題目（聳肩、拇指尖碰食指間）後，其餘題目均具有單向度。其中，「聳肩」為二篇研究共同刪除之題目，表示其 MNSQ 統計值與羅序模式間之適配度在二篇研究中均不佳，推論其與量表中其他上肢動作測量到不同向度之可能性很高。而若量表題目沒有測量到相同向度，會使得該建構（能力）之評量變得不單純，可能導致評量者對於個案能力產生誤判。因此，未來使用 STREAM 時是否繼續將「聳肩」納入上肢次量表之中，建議可依照評量目的自行決定（如欲評量「聳肩」功能，可以納入；如想評量整體上肢功能，或可刪除）。

其餘刪除題目在二篇研究中都不一致，且以上肢動作功能分類而言差異大。本研究刪除之「舉手朝向天花板並出拳」、「伸直手肘」，均為上肢近端功能；而 Hsueh 等研究則是刪除上肢遠端功能的「拇指尖碰食指間」。交叉檢驗結果之不同，可能是由於個案特性不同所造成，例如個案動作功能恢復階段、或動作功能表現等差異。例如本研究個案多為中風發病後一個月內之亞急性期、而 Hsueh 等研究個案多為發病後一年之慢性期，二者動作功能之恢復階段相當不同，可能影響上肢次量表之近、遠端題目的分析結果。

二篇研究之交叉檢驗結果，顯示上肢次量表中有 6 題目具單向度，包含：舉手碰頭頂、手向後摸對側屁股、手臂舉高過頭、前臂旋前／後、握拳、手打開。使用該 6 題目對不同個案特性的受試者施測時，能評量到相同的潛在特質（此處為上肢動作功能），且將 6 題目分數加總即為個案上肢動作的整體功能表現。

對臨床人員而言，分數加總亦代表個案上肢動作能力，臨床人員會根據總分高低判斷個案能力強弱、並依此擬定適宜的治療計畫。但依據本研究之單向度交叉檢驗結果，若施測上肢次量表時直接測完全部 10 個題目，可能測量到上肢動作以外的潛在特質，因此該總分並不適合直接當作個案上肢動作功能的分數。對於研究人員也有類似的情形，若研究目的為個案整體上肢動作功能，則施測時便不應測量到個案的其他能力，意即只施測次量表中具有上肢動作單向度的題目。因此，根據臨床及研究目的所需，使用上肢次量表測量個案的整體上肢動作功能時，應選用具有單向度的題目，該群題目的分數加總才能代表個案的整體功能。

2. 下肢次量表：

二篇研究在下肢次量表各自刪除一題，但刪除題目並不相同。本研究刪除「沿床面彎曲下肢」、而 Hsueh 等研究則刪除「站立時患側腿外展」，各自刪除適配度不佳之題目後，其餘題目均具單向度。被刪除的二個題目均屬於下肢近端功能，不過其受測姿勢及難度均不同，「沿床面彎曲下肢」為在躺平的姿勢下沿床面彎曲髖關節、難度較低；「站立時患側腿外展」則為站立時髖關節外展、難度較高。而交叉驗證結果不同，亦可能是由於個案特性不同所造成。

依據交叉檢驗結果，顯示下肢次量表中有 8 題目具單向度，包含：大腿抬高、膝關節伸直、膝關節彎曲、腳板翹高且腳跟著地（於坐姿下）、腳跟翹高且腳尖著地、膝關節伸直且腳趾翹起、髖關節伸直且膝關節彎曲、腳板翹高且腳跟著地（於站姿下）。意即使用此 8 題目對不同個案特性的受試者施測時，均能評量到相同向度的建構（此處為下肢動作功能），且將分數加總即為整體的下肢動作功能表現。同上所述，若臨床及研究目的為整體下肢動作功能，則使用下肢次量表時應選用具有單向度的題目，該群題目的分數加總才能代表個案的整體功

能。

整體而言，STREAM 之單向度交叉驗證結果顯示：上肢次量表僅有 6 題目符合單向度，以總數只有 10 題來說略微偏低，未來可繼續探討到底是二篇研究樣本間的差異、或是該群「被刪除題目」確實測量到其他向度而使得適配度不佳等原因。而在下肢次量表共有 8 題目符合單向度，以總數 10 題的比例而言，結果令人滿意，與上肢次量表相比下能測量到更加單純、單一的動作功能，但仍可繼續分析被刪除的二個題目，其適配度不佳的原因。由於目前 STREAM 之羅序分析實證尚且不多，僅能就現有結果推論，未來相關證據越多、並具有合理之學理解釋後，便能更具信心地評斷不具單向度的題目是否繼續納入量表或者刪除。

二、羅序信度：

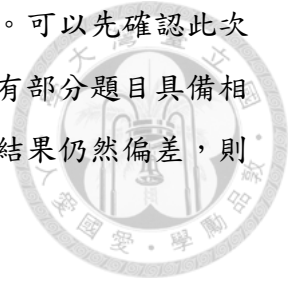
前面的章節曾經提過，羅序信度包含整體量表信度及個案信度，而個案信度會因為個案的測量標準誤不同而有所差異。羅序信度考慮個案能力的個別差異而有不同測量誤差的估計理念，相較於古典測驗理論中「所有個案均為相同測量標準誤」之假設更加準確、合理。

以量表整體信度而言，本研究上、下肢次量表羅序信度分別為 0.92、0.93，Hsueh 等研究則分別為 0.86、0.91，顯示在二個次量表中，本研究的羅序信度均較 Hsueh 等研究高出一些。羅序信度越高代表其測量標準誤越低、重複評量結果的一致程度越高。因此，若以信度作為判斷研究參數優劣之考量，本研究參數會比 Hsueh 等研究略微可靠一點，可供未來使用者參考。

以個案信度而言，本研究於上、下肢次量表之個案信度分布均偏高，在總數 302 位個案中，均有 243 人大於 0.90、59 人介於 0.80~0.90 之間。羅序分析憑藉個別差異的測量標準誤來估計每位個案的個案信度，當個案信度越高，越能精確估計該位個案的能力。一般將個案信度大於 0.90 視為具有相當良好之信度 (WJ, 1992)，因此本研究的個案能力估計值相當值得信賴。

最後，由於上肢次量表的交叉檢驗結果只有保留 6 題目，剩餘題目數較少，

可能使得信度降低，建議未來仍需再檢驗上肢次量表的單向度。可以先確認此次刪除題目之適配度是否確實不佳，若經重複檢驗之結果顯示仍有部分題目具備相同向度，則應考慮重新納入該題目的可能。反之，若重複檢驗結果仍然偏差，則考慮是否增加其他更適當、且具單向度的上肢動作題目等。



三、本研究及 Hsueh 等研究題目難度排序之相關性：

以下提到之難度排序為二篇研究於二次量表共同保留題目之排序，上肢次量表共同保留 6 題目、下肢次量表則為 8 題目。

1. 上肢次量表：

研究結果顯示二篇研究之上肢次量表題目難度排序為低度相關 (Spearman's rho = -0.31)，不符合原本預期，甚至出現負相關。其中以「前臂旋前／後」之難度排序相差最大，該題目於本研究 6 題目中為最簡單、難度最低，但在 Hsueh 等研究卻是第二困難的題目。此外，其餘題目之難度排序也不一致、沒有一定的趨勢或規律。

推測原因也許跟個案特性（如個案能力、動作恢復階段、恢復潛力）等因素有關。以個案能力而言，觀察本研究及 Hsueh 等研究之二群個案的原始分數，可看出在相同題目確實產生不完全相同之作答情形（本研究及 Hsueh 等研究上肢次量表平均分數分別為 12.9、9）。以動作恢復階段而言，本研究個案均屬於亞急性期、Hsueh 等研究則多為慢性期，一般來說如果個案恢復潛力差不多的話，恢復歷程到達慢性期者會有較好的動作功能表現；然而，位處亞急性期之本研究個案的上肢功能卻已經勝過 Hsueh 等研究個案，顯然該群樣本具有更好的預後。簡而言之，若我們對題目難度排序之預期僅考慮「動作恢復歷程」，忽略其他可能因素（如個案群體間的能力差異等），便可能出現不符合預期的結果。

事實上，關於中風患者發病後是否必然遵循「動作恢復歷程」之順序，也存有不同的見解。以出血性中風患者而言，大約 70% 的患者上肢動作會依照恢復歷程順序進步，但剩餘 30% 的患者。發病後 72 小時初評結果位於輕度至中度的患

者大多依照

就結果而言，二篇研究上肢次量表共同保留的 6 個題目（舉手碰頭頂、手向後摸對側屁股、手臂舉高過頭、前臂旋前／後、握拳、手打開），不具有正向、中度至高度相關性。建議未來可再進一步分析不同特性的個案，例如動作恢復階段介於本研究及 Hsueh 等研究之間的個案，探討難度排序是否會依據恢復階段的演進而有所變化的趨勢；或是分析二群個案能力更接近的受試者，把可能影響難度排序的因素單純化，以確認到底是個案能力或是恢復階段造成的差異。

2. 下肢次量表：

二篇研究之下肢次量表題目難度排序為中度相關 (Spearman's rho = 0.48)，約略符合原本預期。表示不論是亞急性期或慢性期之個案，於下肢次量表之作答情形呈現出一定的順序，該順序不會因為個案動作恢復階段或個案能力不同等而受影響。推測可能原因除了上述提過的個案特性因素之外，或許跟下肢動作功能本身也有關係，例如可能是下肢次量表題目間的動作難度差異原本就比較大，因此即便在其他因素的影響下，仍具有中度之難度排序相關性。

就結果而言，二篇研究下肢次量表共同保留的 8 個題目（大腿抬高、膝關節伸直、膝關節彎曲、坐姿下腳板翹高且腳跟著地、腳跟翹高且腳尖著地、膝關節伸直且腳趾翹起、髖關節伸直且膝關節彎曲、站姿下腳板翹高且腳跟著地），具有可接受的難度排序相關性。未來仍可繼續探討只有中度相關的可能原因，並確認是否有更高度排序相關的可能性。

四、本研究參數及 Hsueh 等研究參數各別估計個案分數之相關性：

1. 上肢次量表：

以本研究剩餘 7 題目及 Hsueh 等研究剩餘 8 題目參數估計（共同保留題目為 6 題），且難度排序只有低度相關的情況下，估計之上肢整體分數相關性仍相當高 (Pearson's r = .96)。依據圖 6、圖 7 推測，原因可能是二篇研究保留題目之難

度範圍均大致上涵蓋了多數個案能力分布，意即題目難度與個案能力之間沒有太大的落差，測驗題目對於多數個案而言難易度適中，以此情況下估計之個案分數較準確，使得二篇研究參數估計同一群個案分數之相關性極高。

圖 8 為本研究及 Hsueh 等研究之上肢次量表題目參數散佈圖，圖中每個點代表二篇研究共同保留 6 題目（第 4 題～第 9 題）之題目階難度，上肢次量表採取 0、1、2 的三點計分，三點之間共有二個階難度。以第 4 題為例，具有 4.1、4.2 二個階難度，其他以此類推。由圖 8 可知，每個題目之二個階難度分布各自聚集在一起、分布不均勻，且各題目間難度差異相當接近，能準確估計個案能力的範圍較侷限。然而，二研究參數估計之上肢整體分數相關性仍相當高，推測可能是因為本研究大多數樣本之能力值分布剛好與二個階難度分布範圍重疊，以至於雖然量表的題目難度分布不均，但仍足以估計該群個案的能力。由於二篇研究參數估計個案分數相關性高，因此若想以 STREAM 評量中風個案之上肢動作功能，本研究及 Hsueh 等研究的題目難度參數可以互相取代。

2. 下肢次量表：

二篇研究下肢次量表參數估計本研究個案之分數相關性為中度相關 (Pearson's $r = .62$)。以本研究及 Hsueh 等研究皆剩餘 9 題目參數估計，即便共同保留題目達 8 題之多，且題目難度排序有中度相關，但估計個案分數之相關性反而不如僅有 6 題相同的上肢次量表。依據圖 9、圖 10 可知，二篇研究題目難度參數及個案能力分布相當均勻，理論上分布越均勻，能估計之個案能力範圍越廣。而圖 11 為二篇研究於下肢次量表之題目難度散佈圖，可發現所有題目階難度大致呈現一定趨勢，表示剩餘 8 題目之階難度分布具有相似的線性關係，以此難度參數估計之個案能力應相當一致。然而，二研究參數估計之下肢整體分數僅有中度相關，原因也許跟樣本特性有關，例如二研究參數來自不同的中風恢復階段等。不過由於目前具備的線索仍然不足，研究者尚且無法掌握所有的可能原因。

以結果而言，若想直接使用難度參數估計另一群個案之下肢動作功能，本研

究及 Hsueh 等研究的難度參數是無法直接互相取代的。因此在尚且無法確認那個研究參數較佳的情況下，若想知道個案的能力及分數，建議仍需要重新執行羅序分析。

最後，本研究有二個優點：(1) 個案的恢復階段相當一致，均在發病後一個月內，樣本同質性高。(2) 為目前首個比較二篇 STREAM 羅序分析之心理計量特性交叉驗證研究，有助於評量使用者更加確認 STREAM 的優缺點、並提供未來相關研究的基礎。若心理計量特性之交叉驗證結果良好，即為該特性提供更加鞏固的實證，反之若交叉驗證結果不如預期，也能從可能的因素著手探討。

另一方面，本研究另有二個限制：(1) 僅比較二篇研究，即便知道二者之間具有差異，仍不易判斷誰對誰錯、或是否有二者皆錯之可能，只能就現有證據推論交叉檢驗結果不佳的原因。(2) 二篇研究樣本之 STREAM 原始分數有差異，本研究個案雖屬亞急性期個案、動作尚未恢復完全，但於上、下肢次量表之分數卻均已超過復健超過一年之 Hsueh 等研究個案，可知二群個案本身之動作功能或損傷程度並不在相同基準上。建議未來應盡可能選擇條件更加相近之樣本作研究，並多組比較以利進一步分析。

第五章 結論

本研究交叉檢驗 STREAM 上、下肢動作次量表之單向度、羅序信度、題目難度排序相關性、及二研究參數估計個案分數相關性。以單向度而言，上肢次量表僅共同保留 6 題目、題目數目偏少，而下肢次量表則保有 8 題目、較令人滿意。而本研究估計之上、下肢次量表羅序信度均大於 0.90、皆較先前研究更佳。題目難度排序相關性方面，上肢次量表僅具低度相關 (Spearman's rho = -0.31)，推測可能與樣本特性差異有關，而下肢次量表則有中度相關 (Spearman's rho = 0.48)，較符合研究者預期。最後二研究參數估計個案分數相關性，上肢次量表為高度相關 (Pearson's r = 0.96)、下肢次量表則為中度相關 (Pearson's r = 0.62)，雖然都符合研究者預期，但仍可深入分析下肢次量表僅有中度相關的原因。

整體而言，目前 STREAM 上、下肢次量表心理計量特性之羅序分析結果不完全一致。由於本研究仍存有部分限制，僅能就現有結果推論，建議未來可以繼續探討 STREAM 上、下肢次量表心理計量特性結果不一致的可能因素，並提供解決的方法。如此一來，除了有助於提升 STREAM 的評量價值、也讓臨床及研究人員更了解量表的優缺點。

參考文獻

- 王文中 (1997)。測驗的建構：因素分析還是 Rasch 分析？*調查研究*，3，129-166。
- 李素、林昭宏、謝清麟 (2006)。中風復健動作評估量表的心理計量特性分析。高雄醫學大學行為科學研究所。
- 邱弘毅 (2008)。腦中風之現況與流行病學特徵。*腦中風會訊*，15 卷 3 期，p. 2-4。
- 林嘉皇 (2013)。鏡像治療對慢性腦中風患者上肢動作功能之療效。*臺灣職能治療研究與實務雜誌*，9 卷 1 期，28-40。
- 胡漢華 (2008)。台灣腦中風防治指引 2008 第一版。台灣腦中風學會，2008，75-94。
- 祝旭東、王淳厚、謝清麟、陳美香、陳瓊玲 (1996)。布氏動作恢復量表之信度及同時效度研究。*職能治療學會雜誌*，14，1-12。
- Ahmed, S., Mayo, N. E., Higgins, J., Salbach, N. M., Finch, L., & Wood-Dauphinée, S. L. (2003). The Stroke Rehabilitation Assessment of Movement (STREAM): A Comparison With Other Measures Used to Evaluate Effects of Stroke and Rehabilitation. *Physical Therapy*, 83, 617-630. doi:10.1093/ptj/83.7.617
- Alt Murphy, M., Resteghini, C., Feys, P., & Lamers, I. (2015). An overview of systematic reviews on upper extremity outcome measures after stroke. *BMC Neurol*, 15, 29. doi:10.1186/s12883-015-0292-6
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573. doi:10.1007/bf02293814
- Biller J, L. B., Marsh E, et al. (1990). Spontaneous improvement after acute ischemic stroke. A pilot study. *Stroke* 1990;21:1008-12, 21, 1008-1012.
- Bond TG, F. C. (2001). Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Erlbaum, Mahwah. *Erlbaum, Mahwah*.
- Bonita R, B. R. (1988). Recovery of motor function after stroke. *Stroke*, 19, 1497-1500.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Person Reliability, Item Reliability, and More *Rasch Analysis in the Human Sciences* (pp. 217-234). Dordrecht: Springer Netherlands.
- Carr, J. H., Shepherd, R. B., Nordholm, L., & Lynne, D. (1985). Investigation of a New Motor Assessment Scale for Stroke Patients. *Physical Therapy*, 65, 175-180. doi:10.1093/ptj/65.2.175
- Chen, J., Liu, H. H., Li, Z., & Quiben, M. (2006). Multivariant Analysis of Stroke

- Rehabilitation Outcomes: A Pilot Study in a Chinese Population. *Physical & Occupational Therapy In Geriatrics*, 24, 1-14. doi:10.1080/J148v24n04_01
- Chen, J., Liu, H. H., Li, Z., & Quiben, M. (2009). Multivariant Analysis of Stroke Rehabilitation Outcomes: A Pilot Study in a Chinese Population. *Physical & Occupational Therapy In Geriatrics*, 24, 1-14. doi:10.1080/J148v24n04_01
- Daley K, M. N., Wood-Dauphinée S. (1999). Reliability of scores on the Stroke Rehabilitation Assessment of Movement (STREAM) measure. *Phys Ther.* 1999 Jan, 79, 20-23.
- Duncan, P. W., Goldstein, L. B., Matchar, D., Divine, G. W., & Feussner, J. (1992). Measurement of motor recovery after stroke. Outcome assessment and sample size requirements. *Stroke*, 23, 1084-1089. doi:10.1161/01.str.23.8.1084
- Duncan, P. W., Propst, M., & Nelson, S. G. (1983). Reliability of the Fugl-Meyer Assessment of Sensorimotor Recovery Following Cerebrovascular Accident. *Physical Therapy*, 63, 1606-1610. doi:10.1093/ptj/63.10.1606
- Fugl-Meyer, A. R., & Jääskö, L. (1980). Post-stroke hemiplegia and ADL-performance. *Scandinavian journal of rehabilitation medicine. Supplement*, 7, 140-152.
- G. Broeks, J., Lankhorst, G. J., Rumping, K., & Prevo, A. J. H. (1999). The long-term outcome of arm function after stroke: results of a follow-up study. *Disability and Rehabilitation*, 21, 357-364. doi:10.1080/096382899297459
- Goldstein, L. B., Bertels, C., & Davis, J. N. (1989). Interrater reliability of the nih stroke scale. *Archives of Neurology*, 46, 660-662. doi:10.1001/archneur.1989.00520420080026
- Hattie, J. (1985). Methodology Review: Assessing Unidimensionality of Tests and Items. *Applied Psychological Measurement*, 9, 139-164. doi:10.1177/014662168500900204
- Hsueh, I.P., Hsieh, C.L. (2002). Responsiveness of two upper extremity function instruments for stroke inpatients receiving rehabilitation. *Clinical Rehabilitation*, 16, 617-624. doi:10.1191/0269215502cr530oa
- Hsueh, I.P., Wang, W., Wang, C.H., Sheu, C.F., Lo, S.K., Lin, J.H., Hsieh, C.L. (2006). A simplified stroke rehabilitation assessment of movement instrument. *Phys Ther.*, 86, No. 7, 936-943.
- Huang, C.Y., Lin, G.H., Huang, Y.J., Song, C.Y., Lee, Y.C., How, M.J., . . . Hsieh, C.L. (2016). Improving the utility of the Brunnstrom recovery stages in patients with stroke: Validation and quantification. *Medicine*, 95, e4508. doi:10.1097/MD.0000000000004508
- Huang, Y.J., C. K., Chou, Y.T., Hsueh, I.P., Hou, C.Y., Hsieh, C.L. (2015). Comparison of the Responsiveness of the Long-Form and Simplified Stroke

- Rehabilitation Assessment of Movement Group- and Individual-Level Analysis. *Phys Ther.*, 95, NO. 8, 1172-1183.
- Koh, C.L., Pan, S.L., Jeng, J.S., Chen, B.B., Wang, Y.H., Hsueh, I.P., & Hsieh, C.L. (2015). Predicting recovery of voluntary upper extremity movement in subacute stroke patients with severe upper extremity paresis. *PLoS One*, 10, e0126857. doi:10.1371/journal.pone.0126857
- Kumar, K. V., Joshua, A. M., Kedambadi, R., & Mithra, P. P. (2017). Eclectic/mixed model method for upper extremity functional recovery in stroke rehabilitation: A pilot study. *Journal of Natural Science, Biology, and Medicine*, 8, 75-81. doi:10.4103/0976-9668.198357
- Kwok, T., Lo, R. S., Wong, E., Wai-Kwong, T., Mok, V., & Kai-Sing, W. (2006). Quality of Life of Stroke Survivors: A 1-Year Follow-Up Study. *Archives of Physical Medicine and Rehabilitation*, 87, 1177-1182. doi:http://dx.doi.org/10.1016/j.apmr.2006.05.015
- Lannin, N. A. (2004). Reliability, validity and factor structure of the upper limb subscale of the Motor Assessment Scale (UL-MAS) in adults following stroke. *Disability and Rehabilitation*, 26, 109-116. doi:10.1080/0963828032000157970
- Lin, G.H., Lu, Y., Wu, C.T., Chiu, E.C., Huang, S.L., Hsueh, I.P., & Hsieh, C.L. (2016). Psychometric properties of the Five-Digit Test in patients with stroke. *Disabil Rehabil*, 38, 97-102. doi:10.3109/09638288.2015.1031288
- Linacre, J. M., & Wright, B. D. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
- Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.
- Lyden, P. D., Lu, M., Levine, S. R., Brott, T. G., & Broderick, J. (2001). A Modified National Institutes of Health Stroke Scale for Use in Stroke Clinical Trials. *Preliminary Reliability and Validity*, 32, 1310-1317. doi:10.1161/01.str.32.6.1310
- Malouin, F., Pichard, L., Bonneau, C., Durand, A., & Corriveau, D. (1994). Evaluating motor recovery early after stroke: Comparison of the Fugl-Meyer Assessment and the Motor Assessment Scale. *Archives of Physical Medicine and Rehabilitation*, 75, 1206-1212. doi:10.1016/0003-9993(94)90006-X
- Miller, K. J., Slade, A. L., Pallant, J. F., & Galea, M. P. (2010). Evaluation of the psychometric properties of the upper limb subscales of the Motor Assessment Scale using a Rasch analysis model. *J Rehabil Med*, 42, 315-322. doi:10.2340/16501977-0519
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D.

- L., . . . de Vet, H. C. W. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63, 737-745. doi:10.1016/j.jclinepi.2010.02.006
- Rasch, G. (1960). Probabilistic Models for Some Intelligence and Attainment Tests. *Danish National Institute for Educational Research, Copenhagen*.
- Santisteban, L., Térémetz, M., Bleton, J.-P., Baron, J.-C., Maier, M. A., & Lindberg, P. G. (2016). Upper Limb Outcome Measures Used in Stroke Rehabilitation Studies: A Systematic Literature Review. *PLoS One*, 11, e0154792. doi:10.1371/journal.pone.0154792
- Twitchell, T. E. (1951). THE RESTORATION OF MOTOR FUNCTION FOLLOWING HEMIPLEGIA IN MAN. *Brain*, 74, 443-480. doi:10.1093/brain/74.4.443
- Vongsirinavarat, M., & Hiengkaew, V. (2014). *Factors determining functional ability of individuals with stroke in community* (Vol. 97 Suppl 7).
- Wang, C.H., H. C., Dai, M.H., Chen, C.H., Lai, Y.F. (2002). Inter-rater reliability and validity of the stroke rehabilitation assessment of movement (stream) instrument. *J Rehabil Med.*, 2002 Jan, 20-24.
- Williams, B. K., Galea, M. P., & Winter, A. T. (2001). What is the functional outcome for the upper limb after stroke? *Australian Journal of Physiotherapy*, 47, 19-27. doi:http://dx.doi.org/10.1016/S0004-9514(14)60295-6
- WJ, F. (1992). Fisher WJ. Reliability statistics. *Rasch Measurement Transactions*, 6, 238.
- Wohlk Olsen, L. (2003). *Essays on Georg Rasch and his contributions to statistics*.
- Wright, B. D., & Mok, M. (2000). Rasch models overview. *Journal of Applied Measurement*, 1, 83-106.

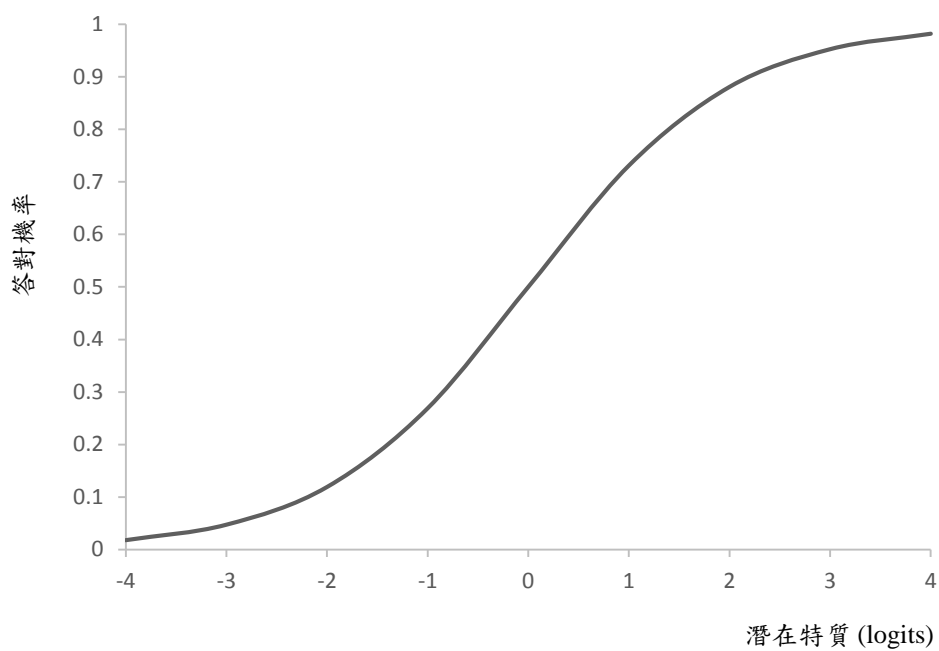


圖 1 試題特徵曲線 (ICC)

註：當潛在特質越大，答對機率越趨近於 1；反之潛在特質越小則答對機率趨近於 0。

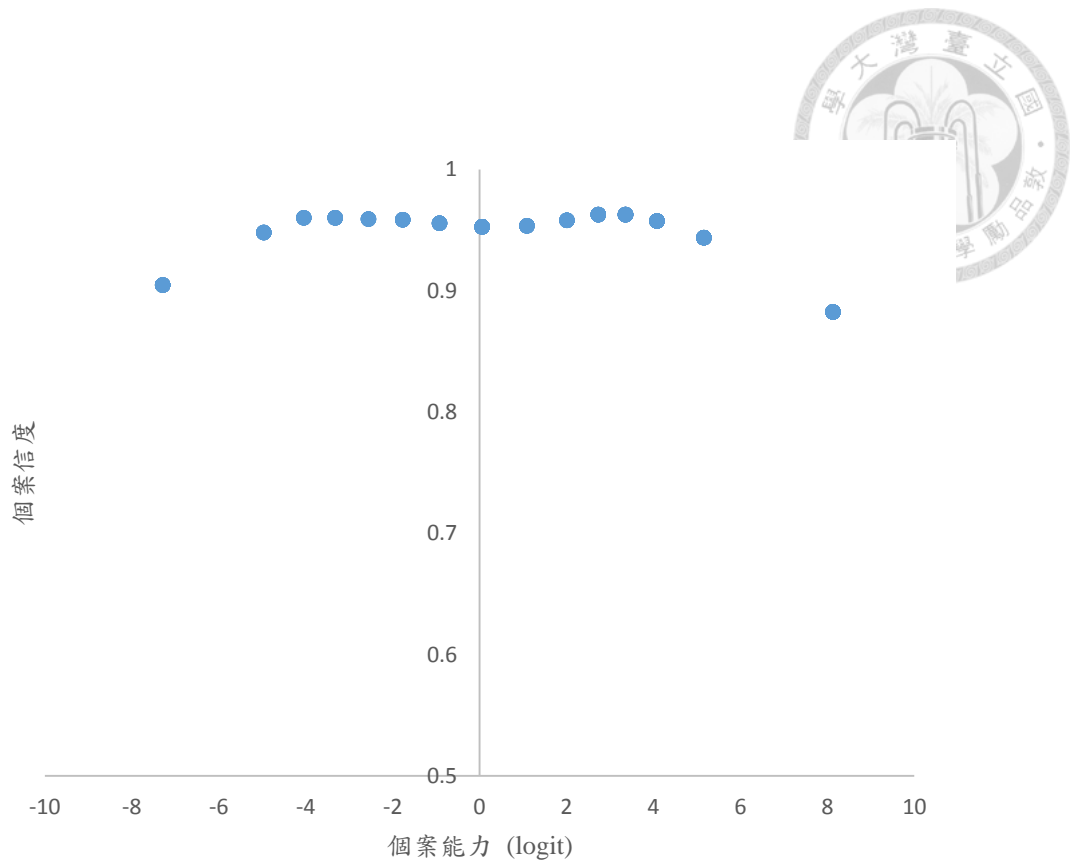


圖 2 上肢次量表一個案能力與個案信度散佈圖

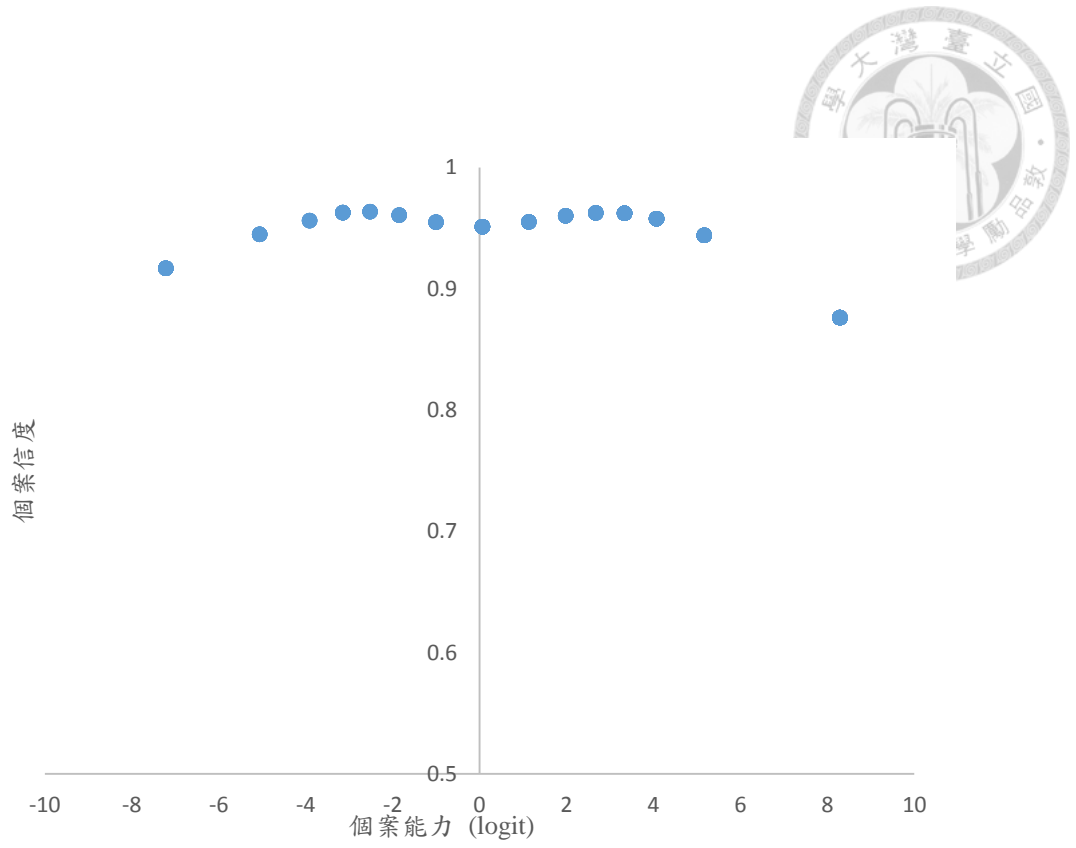


圖 3 下肢次量表一個案能力與個案信度散佈圖

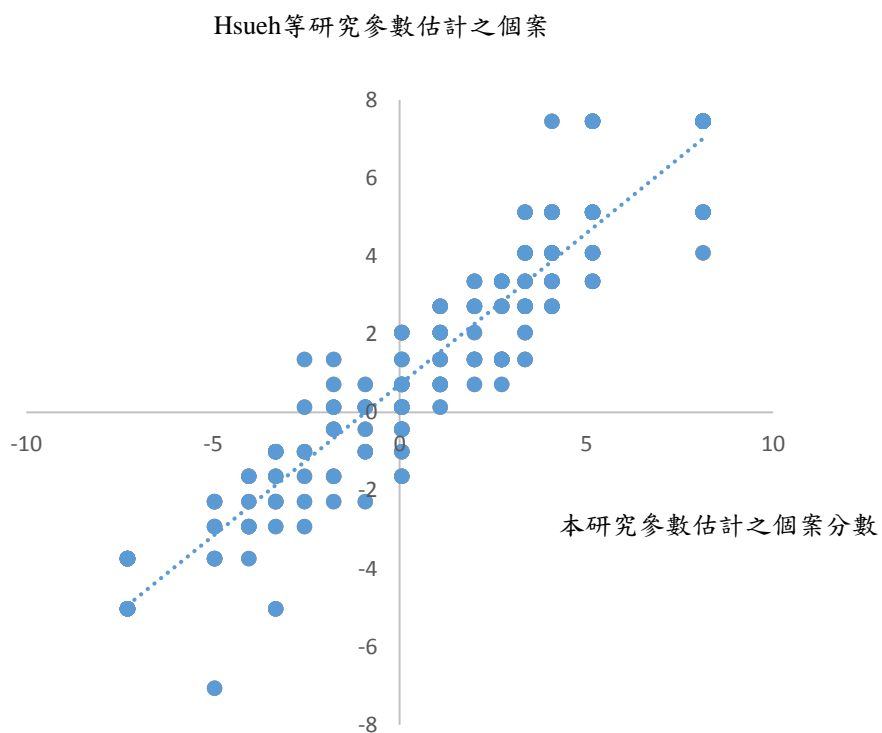


圖 4 上肢次量表一本研究及 Hsueh 等研究參數各別估計個案分數之散佈圖

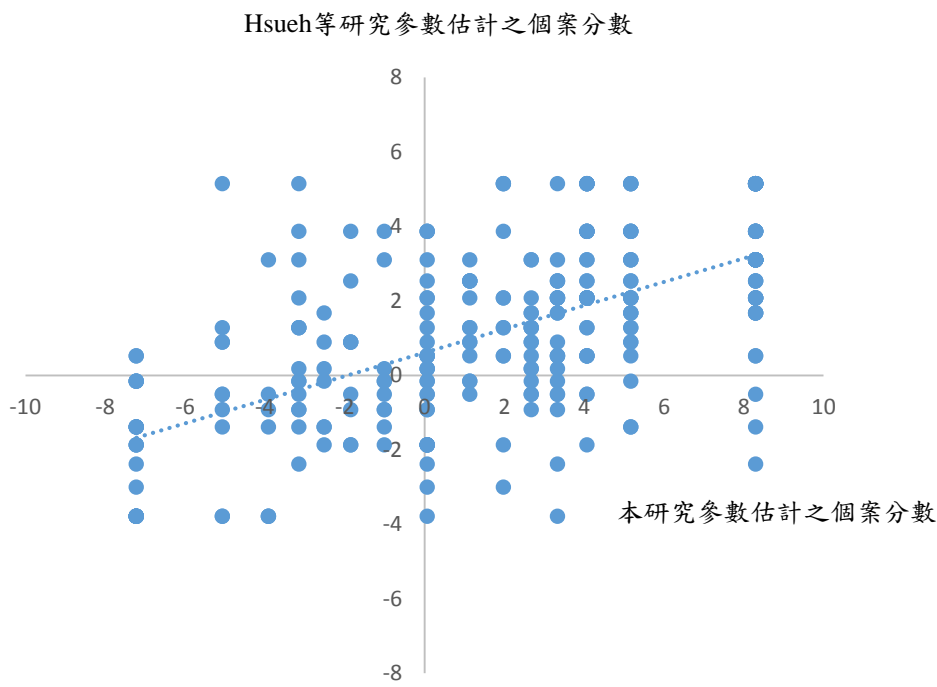


圖 5 下肢次量表一本研究及 Hsueh 等研究參數各別估計個案分數之散佈圖

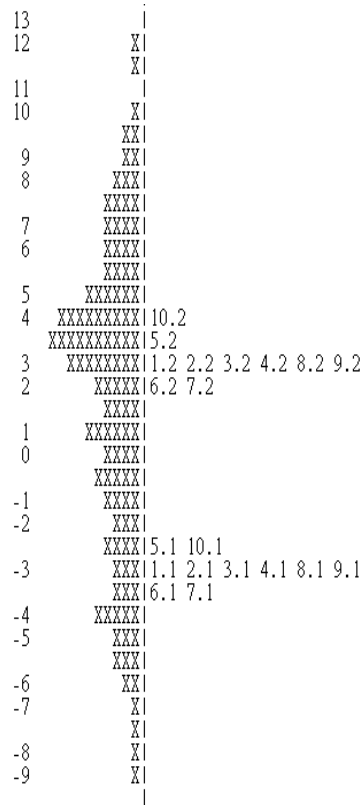


圖 6 上肢次量表一本研究題目難度參數及個案能力分布圖
 註：圖中左側 X 代表個案、右側數字代表題目。
 對應到最左邊數字為個案能力及題目難度之共用坐標軸（單位為 logit）。

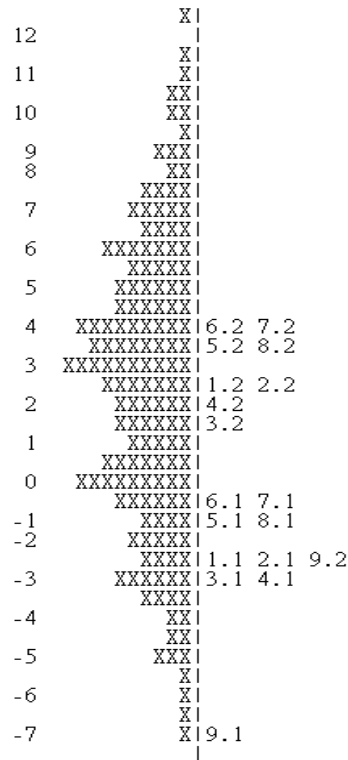


圖 7 上肢次量表—Hsueh 等研究題目難度參數及個案能力分布圖

註：圖中左側 X 代表個案、右側數字代表題目。

對應到最左邊數字為個案能力及題目難度之共用坐標軸（單位為 logit）。

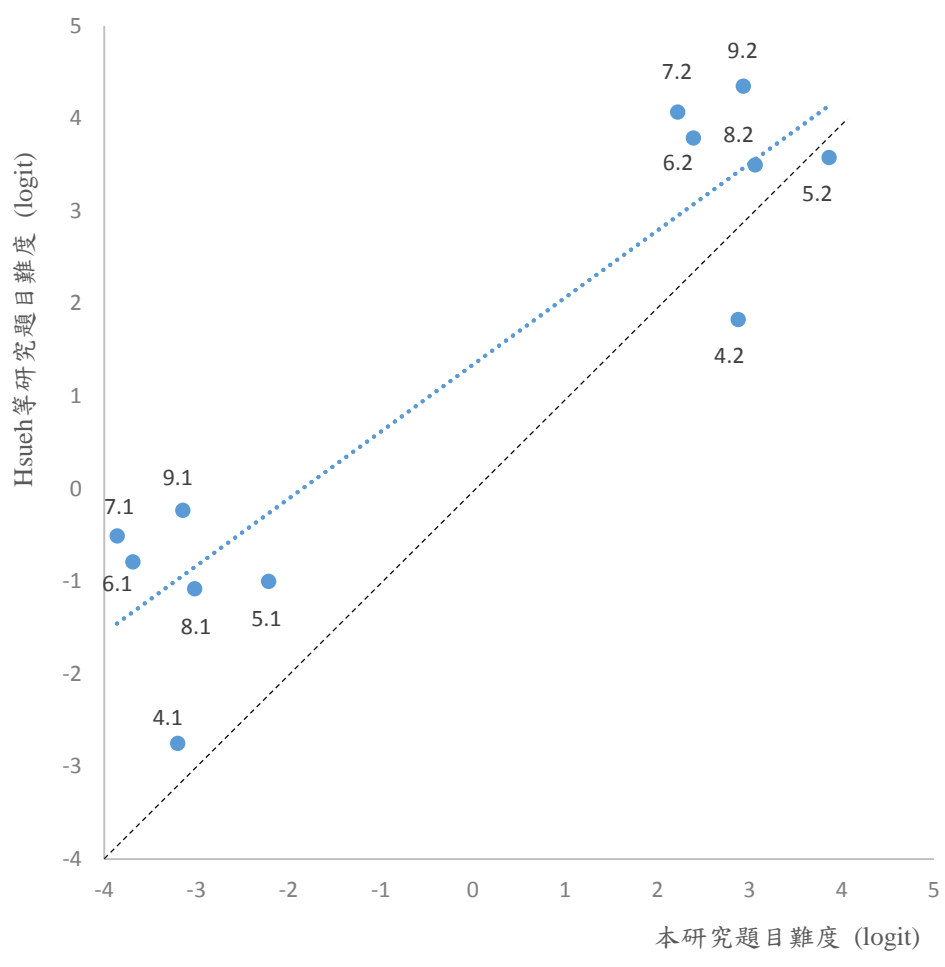


圖 8 上肢次量表—本研究及 Hsueh 等研究題目難度散佈圖
註：上方圓點虛線為資料點的趨勢線、下方 45° 虛線為理想狀態下的一致線。

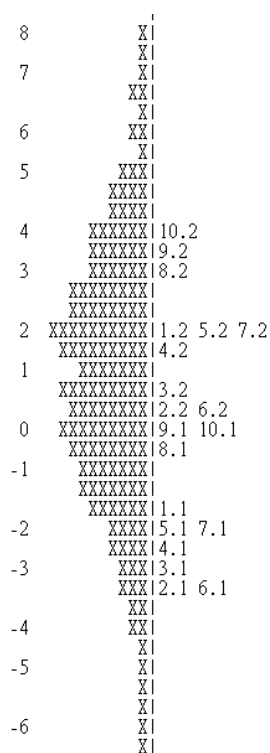


圖 9 下肢次量表—本研究題目難度參數及個案能力分布圖
 註：圖中左側 X 代表個案、右側數字代表題目。
 對應到最左邊數字為個案能力及題目難度之共用坐標軸（單位為 logit）。

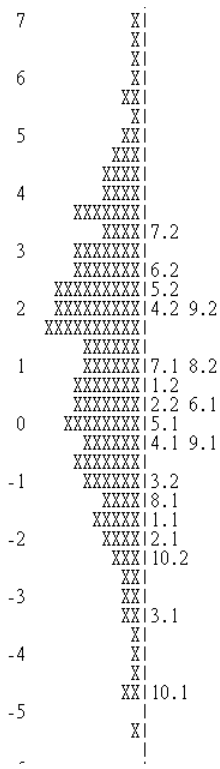


圖 10 下肢次量表—Hsueh 等研究題目難度參數及個案能力分布圖
 註：圖中左側 X 代表個案、右側數字代表題目。
 對應到最左邊數字為個案能力及題目難度之共用坐標軸（單位為 logit）。

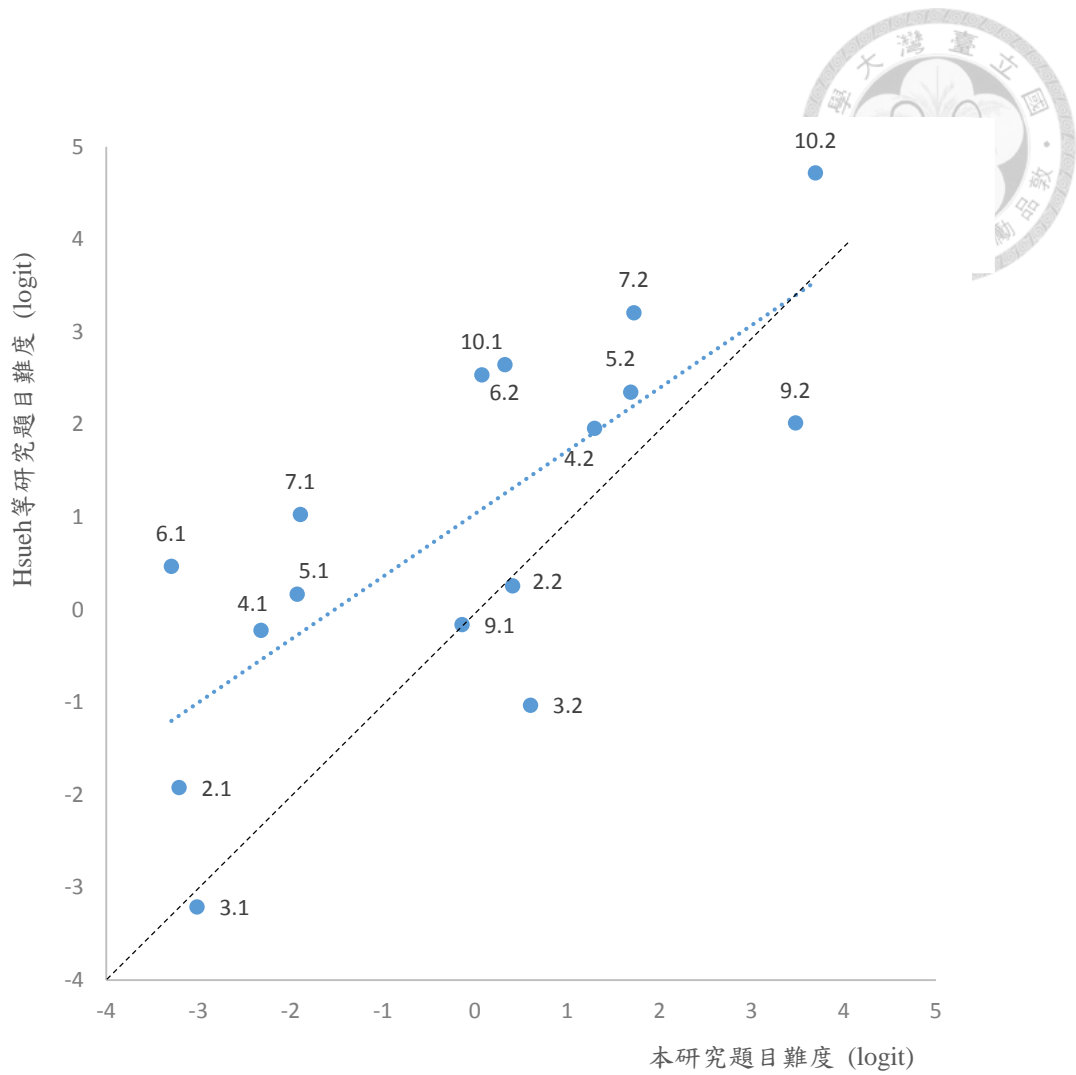


圖 11 下肢次量表—本研究及 Hsueh 等研究題目難度散佈圖
 註：上方圓點虛線為資料點的趨勢線、下方 45° 虛線為理想狀態下的一致線。

表 1 STREAM 之應用文獻分類表

| 功能 | 篇名、作者、出處 | 研究目的 | 研究設計 | STREAM 功能 | STREAM 相關研究結果 | 計分方式 |
|----------------------------------|---|---|--|------------------------------|---|--|
| <p>量化/變化 (包含上下肢動作、行動)</p> | <p>Eclectic / mixed model method for upper extremity functional recovery in stroke rehabilitation: A pilot study.</p> <p>Kumar, K.V., et al. (2017) (Kumar, Joshua, Kedambadi, & Mithra, 2017)</p> <p>“Journal of Natural Science, Biology, and Medicine”</p> | <p>驗證複合式介入法是否能改善急性中風患者的上肢功能。</p> | <p>1. 療效驗證研究 (STREAM 動作次量表)</p> <p>2. 25 位 BRS 2~5 單側中風患者</p> <p>3. 連續 6 天、每天 45 分鐘複合式上肢介入法(含 7 種不同療法、每種療法執行 5 分鐘)</p> | <p>STREAM-UE 當作評量療效的工具。</p> | <p>STREAM-UE 有顯著進步。</p> | <p>等距量尺 (interval)</p> <p>原本應為順序量尺 (ordinal)，為求方便直接加總計分。</p> |
| <p>量化/變化 (包含上下肢動作、行動)+預測</p> | <p>Multivariate analysis of stroke rehabilitation outcomes: A pilot study in a Chinese population.</p> <p>Chen, J., et al. (2009) (Chen, Liu, Li, & Quiben, 2009)</p> | <p>1. 比較上下肢的動作恢復。</p> <p>2. 探索中國中風患者年齡、性別、患側、中風類型及中風發病後至開始復健時間間距等因子與動作恢復程度。</p> | <p>1. 探索性研究、縱貫性研究、回溯性研究</p> <p>2. 從 1999 年六月至 2002 年六月間，以 STREAM 施測 92 位中風患者的動作恢復程度，</p> | <p>STREAM 當作蒐集動作功能程度之工具。</p> | <p>1. 中風發病後至開始復健時間間距為復原完成度及效率之單一最強預測因子。</p> <p>2. 若合併考量，則中風發病後至開始復健時間間距與中風類型佔復原完成度變異量之最大比例；而中風發</p> | <p>等距量尺 (interval)</p> <p>原本應為順序量尺 (ordinal)，為求方便直接加總計分。</p> |

| | | | | | | |
|----|--|---|---|---|---|---|
| | “Physical & Occupational Therapy In Geriatrics” | 復之間的關係。 | 再執行配對 t 檢定及多重迴歸分析。 | | 病後至開始復健時間間距與年齡為最佳復原效率預測因子。 | |
| 預測 | Factors determining functional ability of individuals with stroke in community. Vongsirinavarat, M., et al. (2014) (Vongsirinavarat & Hiengkaew, 2014) “Journal of the Medical Association of Thailand “ | 評估能夠解釋居住社區中風患者之功能能力因子。 | 1. 探索性研究 2. 758 位居住社區之急性期中風患者，且參與居家物理治療。 3. 評估 GCS, STREAM, PASS 及 BI 4. 逐步迴歸分析 | STREAM 當作評量工具，並與 BI 分數比較探討預測程度。(以 STREAM 當作預測因子，預測 ADL) | 1. PASS-maintain /change, STREAM, GCS 及年齡可解釋 BI 85.6%變異量 2. PASS-maintain 是其最中強預測因子，可解釋 81.5%變異量。可知維持坐/站姿是 ADL 關鍵。 | 等距量尺 (interval) 原本應為順序量尺 (ordinal)，為求方便直接加總計分。 |
| 預測 | Predicting recovery of voluntary upper extremity movement in subacute stroke patients with severe upper extremity paresis. Koh, C.L., et al. (2015) (Koh et al., 2015) | 1. 預測自主性上肢動作復原及變異。 2. 藉由重度上肢癱瘓患者之 STREAM 與臨床變項比較，提出自主性上肢動作復原之最佳預測模型。 | 1. 探索性研究 2. 140 位中風導致上肢重度癱瘓患者 3. 評估「出院時 STREAM-UE」、「出入院 STREAM-UE 變化」 4. 逐步迴歸分析 19 | STREAM-UE 當作評量工具，並將前後測數據差異與臨床變項比較、進一步探討預測因子。 | 1. 出院時 STREAM-UE：4 個預測因子（初始 STREAM-UE、出血性中風、初始 NIHSS 分數、初級運動皮質區外的皮質損傷） 2. 出入院 STREAM-UE 變化：3 個預測因子（出血 | 等距量尺 (interval) 原本應為順序量尺 (ordinal)，為求方便直接加總計分。 |

| | | | | | | |
|-------------------------|---|--|---|---|---|---|
| | “PLoS One” | | 個臨床變項，找出二個分數的最佳預測模型。 | | 性中風、初始NIHSS分數、初級運動皮質區外的皮質損傷) 3. 結論為重度上肢癱瘓患者的歧異性太大；該研究之臨床變項預測力偏弱。 | |
| 當作校標 (STREAM 驗證其他工具的效度) | Psychometric properties of the Five-Digit Test in patients with stroke. Lin, G.H., et al. (2015) (Lin et al., 2016) “Disability and Rehabilitation” | 驗證中風病人使用 Five-Digit Test 評量之生態效度、預測效度及反應性。 | 1. 心理計量特性研究 2. 三種特性各自徵招三組受試者。 3. 入院時評量： Five-Digit Test, BI 4. 出院時評量： Five-Digit Test, BI, PASS, STREAM-mobility | STREAM-mobility 當作用來比較 Five-Digit Test 預測效度的參考評量工具之一。 | Five-Digit Test 評量中風患者選擇性/轉換性注意力具可接受之生態效度及預測效度、以及良好的反應性。 | 等距量尺 (interval) 原本應為順序量尺 (ordinal)，為求方便直接加總計分。 |

註：STREAM = Stroke Rehabilitation Assessment of Movement, BRS = Brunnstrom Recovery Scale, STREAM-UE = STREAM-upper extremity subscale, GCS = Glasgow coma score, PASS = Postural Assessment Scale for Stroke Patients, BI = Barthel index, ADL = Activities of Daily Living, NIHSS = National Institute of Health Stroke Scale.

表 2 STREAM 之心理計量驗證文獻分類表

| 篇名、作者、出處 | 研究目的 | 研究設計 | STREAM 功能 | STREAM 相關研究結果 | 計分方式 |
|--|----------------------|---|----------------------|-------------------------|---|
| Reliability of scores on the Stroke Rehabilitation Assessment of Movement (STREAM) measure. Daley K., et al. (1999) (Daley K, 1999) “Physical Therapy” | 驗證 STREAM 的信度及內在一致性。 | 1. 施測者間信度：6 位有經驗的 PT 評量 20 位患者。兩兩一組，一個執行評量、另一個就近觀察。 2. 施測者內信度：20 位 PT 評量 4 位患者。正式評量前會先練習一次錄影式評量，接著團隊討論。 之後，10 人一組，同時且獨立地執行 4 次錄影式評量。 一個月後，再重複評量一次。 3. 內在一致性：使用直接觀察信度研究，補上 6 位能力較差的患者數據（由另外 4 位 PT 錄影式評量）。 | 驗證 STREAM 的信度及內在一致性。 | STREAM 具有高度的信度，支持其測量潛力。 | 等距量尺 (interval) 原本應為順序量尺 (ordinal)，為求方便直接加總計分。 |

| | | | | | |
|--|---|---|--------------------------------|--|--|
| <p>A simplified stroke rehabilitation assessment of movement instrument.</p> <p>Hsueh IP, et al. (2006) (Hsueh IP, 2006)</p> <p>“Physical Therapy”</p> | <p>使用多向度羅序分析將原版 STREAM (上肢、下肢、動作 3 個次量表, 各 10 題) 簡化為 S-STREAM</p> | <p>1. 受試者為 351 位初次中風患者, 中風後時間之中位數為 12.5 個月。</p> <p>2. 首先以羅序分析驗證 STREAM 三個次量表的單向度。再根據專家意見及分析結果刪去多餘的項目。</p> <p>3. S-STREAM 的羅序信度、S-STREAM 與 STREAM 的共同效度, 以多向度羅序分析及 ICC 檢驗。</p> | <p>STREAM 發展成更方便、快速評量的簡化版。</p> | <p>1. 原版符合羅序模式項目: 上肢 8 題、下肢 9 題、動作 10 題 (S-STREAM 從中各取 5 題)</p> <p>2. S-STREAM 具有高度的羅序信度、單向度、共同效度。</p> | <p>等距量尺 (logit)</p> <p>有別於傳統上直接將順序量尺當作等距量尺使用, 為真正使用等距之 logit 量尺。</p> |
|--|---|---|--------------------------------|--|--|

註：PT = physical therapist, S-STREAM = simplified STREAM, ICC = intraclass correlation coefficient.

表 3 STREAM 之上下肢動作次量表

| 上肢動作 | | 下肢動作 | |
|-----------------|------------------------|---------------------------|----------------|
| 仰臥 (Supine) | U1 舉手朝向天花板並出拳 | L1 沿床面彎曲下肢 | |
| | U2 伸直手肘 | | |
| 坐姿 (Sitting) | U3 聳肩 (兩側) | L2 大腿抬高 | |
| | U4 舉手碰頭頂 | L3 膝關節伸直 | |
| | U5 手向後摸對側屁股 | L4 膝關節彎曲 (患側腳跟置於健側腳尖，向後滑) | |
| | U6 手臂舉高過頭 | L5 腳板翹高、腳跟著地 | |
| | U7 前臂旋前/後 (elbow 90°) | L6 腳跟翹高、腳尖著地 | |
| | U8 握拳 (大拇指在外，手腕微翹) | L7 膝關節伸直，且腳趾翹起 | |
| | U9 手打開 | | |
| | U10 拇指尖碰食指尖 | | |
| | 站立 (Standing) | 無 | L8 患側腿外展 |
| | | | L9 髖關節伸直、膝關節彎曲 |
| | | L10 腳板翹高、腳跟著地 | |

註：最左邊欄位的「姿勢」表示受試者須在該指定姿勢下受測。

表 4 本研究對象之基本人口學資料 (N = 302)

| 變項 | 數值 |
|---------------------|-------------|
| 性別 (男/女) | 180 / 122 |
| 年齡：平均 (標準差) | 65.3 (13.7) |
| 中風類型 (%) | |
| 缺血性腦中風 | 198 (66%) |
| 出血性腦中風 | 104 (34%) |
| 發病後時間 (月) | 1 |
| 中風復健動作評估量表：平均 (標準差) | |
| 上肢次量表 | 12.9 (6.2) |
| 下肢次量表 | 12.2 (5.7) |

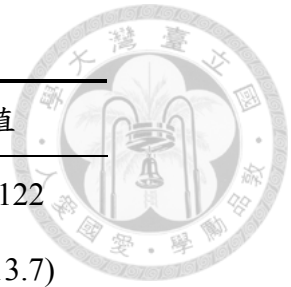


表 5 上下肢次量表之羅序信度

| | 題目數目 | 信度 |
|-------|------|-----|
| 上肢次量表 | 7 | .92 |
| 下肢次量表 | 9 | .93 |



表 6 上肢次量表之個案信度分布情形 (N=302)

| 信度 | 人數 (%) |
|-------------|-------------|
| > 0.90 | 243 (80.5%) |
| 0.80 ~ 0.90 | 59 (19.5%) |



表 7 下肢次量表之個案信度分布情形 (N=302)

| 信度 | 人數 (%) |
|-------------|-------------|
| > 0.90 | 243 (80.5%) |
| 0.80 ~ 0.90 | 59 (19.5%) |



表 8 本研究及 Hsueh 等研究之上肢次量表題目難度參數與排序

| | 本研究 | 排序 | Hsueh 等研究 | 排序 |
|-----|-------|----|-----------|----|
| U1 | - | - | -0.02 | - |
| U2 | - | - | -0.77 | - |
| U3 | - | - | - | - |
| U4 | -0.16 | 3 | -0.46 | 1 |
| U5 | 0.82 | 6 | 1.29 | 3 |
| U6 | -0.65 | 2 | 1.50 | 4 |
| U7 | -0.82 | 1 | 1.78 | 5 |
| U8 | 0.02 | 5 | 1.21 | 2 |
| U9 | -0.11 | 4 | 2.06 | 6 |
| U10 | 0.90 | - | - | - |

註：難度排序僅列出二篇研究共同保留之題目。數字大者，代表較困難。

表 9 本研究及 Hsueh 等研究之下肢次量表題目難度參數與排序

| | 本研究 | 排序 | Hsueh 等研究 | 排序 |
|-----|-------|----|-----------|----|
| L1 | - | - | -0.40 | - |
| L2 | -1.40 | 2 | -0.83 | 2 |
| L3 | -1.21 | 3 | -2.12 | 1 |
| L4 | -0.51 | 5 | 0.87 | 3 |
| L5 | -0.12 | 4 | 1.26 | 5 |
| L6 | -1.48 | 1 | 1.56 | 6 |
| L7 | -0.09 | 6 | 2.12 | 7 |
| L8 | 1.26 | - | - | - |
| L9 | 1.67 | 7 | 0.93 | 4 |
| L10 | 1.88 | 8 | 3.63 | 8 |

註：難度排序僅列出二篇研究共同保留之題目。數字大者，代表較困難。

表 10 上下肢次量表—本研究及 Hsueh 等研究題目難度排序之相關性

| | Spearman's rho |
|-------|----------------|
| 上肢次量表 | -0.31 |
| 下肢次量表 | 0.48 |



表 11 上下肢次量表—本研究及 Hsueh 等研究參數各別估計個案分數之相關性

| | Pearson's r |
|-------|-------------|
| 上肢次量表 | 0.96 |
| 下肢次量表 | 0.62 |

