

測評中心研究員參加「測驗理論與應用工作坊」心得分享



von Davier 博士主講測驗理論與應用工作坊並指導參加者上機操作測驗軟體

【文、圖／測驗及評量研究中心助理研究員 林奕宏】

新年新氣象，學術界的活動也欣欣向榮！今年 1 月 12 日，院內測評中心曾建銘副研究員率領吳慧珉、謝佩蓉、林奕宏、蕭嘉偉等研究同仁參加由臺灣師大頂尖大學計畫辦公室主辦的測驗理論與應用工作坊。該工作坊是由美國教育測驗服務社（Educational Testing Service, ETS）研究發展部門的首席研究科學家 Matthias von Davier 博士主講。

von Davier 博士以試題反應理論（Item Response Theory, IRT）的基本模式 Rasch Model 為解說的起點，隨後延伸到較複雜的模式如：Mixture Models、Diagnostic Models 等，同時指導學習者上機操作由 ETS 研發的測驗軟體 MDLTM（Multidimensional Discrete Latent Trait Models），以實例來呈現這些測驗模式的可應用性。當日的參加者超過五十位，都是測驗評量領域的研究者。

von Davier 博士是國際知名的心理計量學家，除了服務於美國 ETS 之外，也是英國牛津大學教育系的名譽研究員。他是 IRT 認知診斷模式的專家，對認知診斷模式有獨到的見解，曾多次應邀發表與認知診斷模式議題有關的演講。

認知診斷與教育評量有何關係？事實上，教育評量的目的，除了測量出學習者的學習現況外，同時也應該提供學習者學習成效的診斷訊息，以利教學者根據診斷訊息進行有效的補救教學。國內外已有許多教育學者主張教育評量的實施，應該以認知心理學對學習歷程的研究成果，作為編製測驗的實質理論依據。此種將教育評量與認知心理學的實質理論相結合的評量方式，即是所謂的「認知診斷評量」。

認知診斷評量之目的在於了解學生學習困難的原因，以便進行補救教學。評量主要是在決定學生的成熟度、預備狀態、起點行為，以及其與學習有關的特質，予以分組安置，診斷其學習困難的原因。進行的時間點，通常是在教學之初或是學生學習有困難的時候，由教師安排進行評量。評量項目包含認知、情意、技能方面的行為，必要時也包含學生身心或是其家庭環境等因素的評量。評量工具多是標準化測驗、診斷測驗，或是教師自編測驗、相關的檢核表等等。簡而言之，認知診斷評量至少具有以下三項特點：

- 一、 目的在於診斷學生的困難所在，並針對其困難，予以必要的補救教學。
- 二、 是一種具備綜合性和精密性的評量。
- 三、 學習困難特別嚴重者，需要進一步去探討其真正的原因，有時需要使用不同的診斷工具，或轉介尋求專業人員的協助。

在認知診斷評量的發展中，對於技能或認知屬性等是如何影響測驗的結果有許多不同的假設，從而衍生出了許多種不同的測量模式，如 DINA 模式 (deterministic input, noisy "and gate)、NIDA 模式 (noisy input, deterministic "and" gate model)、RUM 模式 (reparametrized unified model)、DINO 模式 (deterministic input, noisy "or" gate model)、G-DINA 模式 (generalized DINA) 等。這些不同的認知診斷模式，彼此之間是否具有關連性及各模式適用的情境為何？國外已有許多學者投注於此方面的研究，但國內目前仍較少見這一方面的探討，是一項新興的議題；von Davier 博士正是此領域的專家。因此 von Davier 博士此次來訪，與我院內研究同仁交流研究心得，對於促進國內測驗學術及實務的發展，將產生相當正面的影響。