

# 判斷本位評量初探

林文田

## 摘要

本文旨在介紹判斷本位評量的內涵及其信效度，以提供另一評量特殊需求學生的方法與實施建議。在面對評量困難或無法施測之幼童時，專業人員或特教教師可藉由幼童多元資料的蒐集（包括標準化測驗、非標準化測驗、觀察及訪談等）及其他各種的評量方式，採用臨床專業判斷的方式，為學生做出正確的教育診斷。

**中文關鍵詞：**臨床判斷、判斷本位評量

**英文關鍵詞：**clinical judgment, judgment-based assessment

## 壹、前言

大家一定都有過這樣的經驗：當發現自己喉嚨痛或開始流鼻涕時，一定猜想，喉嚨可能發炎了或是可能感冒了，於是便開始注意自己的飲食，儘量不吃油炸的食品，並開始大量的喝溫開水，若是仍然覺得不舒服，方才求診。醫師則根據病患所描述的症狀，再依其生理所呈現的各種現象做一判斷，最後開出處方。而若病患回家依照指示服藥後，有效則病狀改善乃至痊癒，反之醫師則再依病患的症狀判斷可能的病症。上述這種醫學診斷—處遇的模式相信大家都不陌生，醫師經由其臨床經驗，再根據其所受醫學訓練的專業判斷對病情所做的一種評量模式。

利用臨床判斷(clinical judgment)的方式在我們生活的周遭亦時常被用到；例如我們常說：「七坐八爬九長牙。」就是利用以往父

母在養育幼童發展上臨床判斷的運用。對於幼童來說，在教育與醫療上的不同是：醫療上可從病理看出其隱含的病情來對症下藥，但在教育上卻需借助於一些測驗工具來看出其能力上的表現。然而在實施傳統的測驗工具時，由於幼童的認知、語言、情緒、動作發展等能力的障礙，若是實際依照標準化的施測程序與解釋，往往無法完全描述幼童的能力，因此常需參照幼兒身邊相關人員（例如：父母、教師等）的描述及在不同情境下（學校、家庭）的表現來做一正確的判斷。然而，身心障礙幼童常受限於本身能力的障礙，一些標準化的測驗工具對他們來說常常無法施測。於是乎，利用相關專業團隊，蒐集幼童在所處環境中呈現出來的資料（包含父母所提供的資料）做正確的判讀，便是評量幼童相關能力十分重要的依據。而判斷本位評量(judgment-based assessment)模式便是基於此而發展出來，以下就判斷本位評量的

內涵、信效度及省思上做相關的說明與建議。

## 貳、判斷本位評量的內涵

判斷本位評量在廣泛評量幼童上扮演一個很重要的角色，藉由判斷本位評量，我們可以蒐集、量化個體所處環境表現出來的印象及概念。由此可知，判斷本位評量是以臨床判斷為基礎而發展出來。Simeonsson、Huntington 和 Parse (1980)將「臨床判斷」界定為「父母或專業人員在平日運用直覺和/或個人經驗對個案的各種推斷」。他們認為，為了要「真實地」了解兒童環境中重要關鍵人物，平日累積的觀察結果和印象是很重要的。尤其，對一些以標準化測驗評估卻「無法施測」的嚴重障礙者而言，更是如此。

### 一、判斷本位評量的目的

礙於一般標準化測驗對身心障礙者造成不當的評估效果，判斷本位評量的目的是要蒐集除了一般標準化測驗所提供的訊息外，以及兒童相關人員所提供的各種訊息與資料來源，以對其做出正確的判斷。因此，利用專業間與多元的評量方法（觀察、晤談、等級量表評定等），結合各種不同工具測驗的結果，以提供判斷的依據。然而，判斷本位評量所提供的不僅是資料的蒐集而已，它更強調父母與專業人員之間的溝通，以幫助訂定相關計畫及評估療育計畫的執行，並能對兒童的資料有一致性及區分的功能(Bagnato & Neisworth, 1990)。

### 二、判斷本位評量的特徵

Thurman 和 Widerstrom(1990)認為對許多診斷者而言，統合及綜合個體在家庭生態

體系複雜的資訊，將更能廣泛著重於評量的焦點。有經驗的臨床工作者依據許多的診斷工具及個人判斷來診斷當事者。判斷本位評量已被定義為蒐集、組織、及將專業人員和照顧者對孩子所處環境上的特徵印象予以量化，這些能力均需藉由觀察及累積一段時間的經驗，以做出立即的判斷(Bagnato & Neisworth, 1988)。因此，利用聯合處遇的方法對那些在標準化測驗表現不佳或在一些評量情境上無法引出孩子最佳的表現能力時，這種方法就特別的重要。LeLaurin(1990)提到臨床上常使用的方式有：

1. 首先，判斷(judgments)是由專業意見所形成，諸如物理治療師判斷肌肉狀況、心理治療師判斷孩子或家長的情緒狀況等，這種專業判斷是基於判斷者先前的訓練與經驗、對適切行為熟練的觀察與摘要及過去相似個案來取得相關訊息。
2. 其次，判斷本位評量與孩童所處環境或孩童所處的社會生態體系的觀點及印象有關。這種判斷方式可從孩童及家庭的需求上，提供令人理解的描述，以確保個案在跨情境（學校及家裡）的各種表現上能蒐集到更準確的訊息。判斷本位評量就是要在有足夠環境的安排機會下，蒐集個案的各種表現。
3. 第三，判斷本位評量對確認低頻行為及更細微特質或反應類型上更有用，例如孩童對新奇事物的反應或對突然聲音的反應等。這些資料有助於把特定功能性的表現納進來，亦可幫助診斷與計畫的訂定。
4. 最後，當在評估行為的類型或模式時，判斷常用於特定的等級量表上。例如：用來

評量孩童特質或家裡環境的相關等級量表等。

因此，診斷者必須對測量要有概念，也必須對行為有所警覺。當正確的使用所蒐集到的訊息或資料時，判斷本位的訊息才會具有內部一致性及團隊一致性，而這些資料正確的研判方能符合孩童及家庭的需求。臨床判斷最重要的一個要素就是能對隱含(implicit)的資料做出正確的判讀。

### 三、判斷本位評量的對象與教育應用

由判斷本位評量的目的與特徵不難發現，評量的對象主要是在認知、語言、情緒、動作發展等能力發展障礙，且又無法僅以一般標準化測驗結果來進行評估的兒童。其對象通常包括智能障礙、溝通障礙、情緒障礙、泛自閉症(autism spectrum disorders)及發展遲緩等幼童。

在教育的應用上，可用來鑑定特殊教育的學生。在教學上，特教教師可依臨床專業判斷與經驗，適當評量學生的能力，以解決無法以一般標準化測驗工具來得知學生能力的情形。

### 參、信度與效度

在實施判斷本位評量時，其信度與效度所遭遇到的考驗與其他另類評量(alternative assessment)相似，均是質疑於它的不夠客觀、未具足夠的信度與效度。Fleischer 和 Belgredan(1990)指出許多的文獻資料顯示，在使用判斷本位評量時均著重在評分者間信度(inter-rater reliability)及效標關聯效度(criterion-related validity)的運用，且許多實徵

性研究也發現父母與專家對兒童的評定上有很高的相關。國內王天苗、范德鑫(1998)以臺北市國中小普通班學生、資源班學生與啓智班學生及被轉介接受臺北縣、市國中特殊教育鑑定之學生，以判斷本位評量為基礎所編製而成的「學校適應能力量表」進行研究。研究結果發現「學校適應能力量表」中五大向度、十四項能力足以區別普通班和特殊班學生、資源班和特殊班學生，大多數適應能力得分和智力之間有低度相關、教師和家長間之評定結果並不一致、不同地區、教育年段和教育安置之學生學校適應能力有差異。由此可知，專家與父母之間的評量並不一致，許多專家將這種差異性歸因於父母親過於高估孩子的能力(Fleischer & Belgredan, 1990)。Gradel、Thompson 和 Sheehan(1981)提出以顯著差異(apparent discrepancy)的方法來評斷，並認為造成這種不一致的原因可能是專家過於低估，而不是父母過於高估。

王天苗、范德鑫(2000)進一步以臺北縣特殊學生鑑定安置的學生及就讀臺北市國中小特殊班與資源班的學生為樣本。研究中包括參與計畫的特教教師，探討運用「專業判斷」結果與以標準化智力測驗和適應行為量表結果鑑定智障學生和安置決定間的一致性，並且探討教師對運用專業判斷評量方式決定教育安置的看法。結果發現：老師利用「學校適應能力量表」專業判斷所得結果與依智力和適應行為兩項標準化評量所得結果間一致的比率約為 96%，班級教師或特教教師依專業判斷建議的安置與經鑑定人員正式鑑定後決定的安置結果一致的比率約為 86%，而特教教師對以專業判斷決定學生的

特殊教育安置的意見紛歧。因此，評量者間在信度的呈現上仍未有定論。

判斷本位評量的效標關聯效度或同時效度(concurrent validity)和判斷本位量表與其他各種常模及標準本位測量有關；許多研究使用判斷本位量表已經反應出，評量者行為或特徵的判斷與表現測驗上的得分有類似的相關。Bagnato 和 Neisworth 於 1985 年的研究中，使用 Preschool Attainment Record(PAR) 及 Perceptions of Developmental Skills(PODS) 兩種判斷本位量表以及 Gesell Developmental Schedules-Revised 及 HICOMP Track Record 兩種發展表現測量(developmental performance measures)來檢測同時效度，結果指出在身心障礙幼兒的發展程度測量上，判斷本位與發展表現量表有很大的關聯及一致性，這種利用區辨與輻合的判斷方式對療育專家是相當重要的(Fleischer & Belgredan, 1990)。因此，蒐集各種與幼兒相關發展程度的資料亦是不可或缺。

Fleischer 和 Belgredan(1990)認為判斷本位評量亦可提高、增進及證實正式或標準化測驗的效果。Kleinman 等人(2008)以 77 名幼兒接受二次的診斷（第一次診斷時幼童年齡約為 2 歲，第二次診斷時幼童已約為 4 歲），以期瞭解自閉症是否能在 2 歲時便診斷出來。研究中以 DSM-IV 診斷標準的專業判斷和使用兒童自閉程度評量表 (Childhood Autism Rating Scale, CARS)、自閉症診斷觀察表(Autism Diagnostic Observation Schedule, ADOS)及自閉症診斷晤談(Autism Diagnostic Interview, ADI)的方式來檢驗診斷的穩定性，且又以不同的標準化測驗工具（例如：

文蘭適應量表、貝萊嬰幼兒發展量表）來相互驗證，結果指出在診斷自閉症上均達顯著性( $p < .01$ )；而在診斷的穩定性上，臨床專業判斷則高達 80%。

因此，若能搭配、蒐集各種不同的資料與評量方式將有助於對資料的判讀。不過，筆者認為若僅考量到評量的信效度尚不夠嚴謹。評量應同時考量到效度的責任歸屬，除了要評量出所欲評量的能力外，更應考慮評量後所造成的社會後果，不同的文化、家庭、種族背景均是評量後需考慮到的問題。以判斷本位評量來說，可能由於評量者（專家）與孩童父母社會文化背景的差異而有不同的詮釋，因而造成評量的差異，其結果便須謹慎解釋，以免造成對幼童不當的標記作用。

## 肆、省思與建議

對任何診斷者或處遇專家來說，臨床判斷是其專業技能很重要的一部分。目前在特殊需求學生的鑑定安置上，大都是由特教教師扮演教育診斷的角色，其目的是要藉由教師平時教學及實務上的經驗來對學生進行適當的診斷與處遇。台北市在學情障的鑑定安置上，便是要求特教教師須接受學情障的初階專業訓練。為了讓診斷更加準確，進一步辦理進階班及高階班，其目的除了使鑑定人員須具備高的專業知能外，另一方面可藉由多元資料的蒐集，運用專業判斷，以證明判斷有根據，防止診斷或處遇的錯誤以及保護當事人。因此，臨床工作上多元資料的蒐集應是不可或缺的一項要素，這點在早期療育的領域上尤其重要。這是由於早期療育需要

各種不同專業人員專業技能及觀點的整合，以提供測驗及各方觀點的一致性(LeLaurin, 1990)。

一般而言，我們都過於強調標準化測驗的施測結果，而忽略一些對兒童主觀印象描述的資料。因此，建立與評量相關的元素對身心障礙者而言是相當重要且必須的。而透過專業人員對兒童本身及家庭資料的蒐集可達到評量的目的。蒐集的方式可透過電話訪談、家庭拜訪、臨床訪視（結構化的測驗、半結構化的評量、兒童玩耍情境、親子互動等）及個案討論來進行。因此，若是我們盡信標準化測驗所呈現出來的資料，而不去參照非正式評量所提供的資料，就會陷入「買履信度」、「刻舟求劍」的處境，而無法完整對評量的相關資料進行正確的判讀。

在進行判斷本位評量時，評量者能否對資料進行驗證與查對、對資料做有效的處理、資料是否具多元性以及資料是否具客觀性等，均是影響專業人員在進行判斷時相當重要的因素。再者，由於判斷本位評量需要多位專業人員共同進行判斷，也因為並非每位評量者均受過相關專業的訓練，因此在評量時會遭遇相當大的阻力。事實上，有些診斷者對他們自己精明的判斷很自豪，也認為在這個團體裡唯有專家才能針對相關的領域作出評量，於是便產生不良的「專家效應」。這種自以為是，不兼顧團體專業協調，只顧自己說法的偏見，有些學者稱為「角色釋放」(role release)失敗的一種，指的便是團隊專家們難以分享他們專業知識的領域(Foley, 1990; Peterson, 1988)。

有鑑於此，筆者提出下列幾點建議：

- 一、除了標準化測驗的結果外，將孩童所處環境的生態體系納入評量的項目，才能正確的評估孩童的真實能力。
- 二、專家除了是專業分工外，可透過個案會議討論，來達到專業整合。專業人員應貢獻所學互相協調，並進行資訊互通，避免單一向度的診斷，如此才能對孩童有正確的判斷能力，而不致造成專業間相互牽制的不良影響。
- 三、在評估孩童時，應了解評量的目的為何？有什麼樣的辦法可以蒐集到個案的各種資料（資料是否多元）？資料是否具參考價值？評量的人員有哪些？其專業能力是否足夠？等問題。
- 四、蒐集評估孩童的各項能力時，應提供相關人員遵循的實施程序及方法，以提高診斷的準確性。

## 伍、結語

無論何種評量方式，所評量的應是兒童所接受「教育」的適切性，而不是標記作用；因此，評量應考慮到評量後的後果。若我們不考慮兒童未來所提供教育服務的適切性，則再多的評量亦無實質上的意義，最終不過留於看到孩子的「缺陷」而已。另外在「專家」與「生手」的評量判斷上亦應作適當的解釋，專家之間應以「角色釋放」為目的，以確實在使用判斷本位評量上能真正做出相互分享並做出正確的判定。最後，判斷本位評量的成功有賴於多元資料的蒐集與相關專業人員的參與，我們亦不可忽略幼童的父母親在整個評量上的重要性，適時的參考各方

資料方能做出最正確的教育決定。

(本文作者為臺北市立南港國小教師)

## 參考文獻

王天苗、范德鑫 (1998): 智障學生學校適應能力之探討。《特殊教育研究學刊》，16，109-129。

王天苗、范德鑫 (2000): 運用專業判斷鑑定智障學生之研究。《特殊教育研究學刊》，19，127-148。

Bagnato, S. J., & Neisworth, J. T. (1988). An assessment typology in early intervention. In S. L. Odom & M. B. Karnes (Eds.), *Early intervention for infants and children with handicaps: An empirical base* (pp. 23-49). Baltimore: Brookes.

Bagnato, S. J., & Neisworth, J. T. (1990). *Developmental system to plan early childhood services*. Circle Pines, MN: American Guidance Service.

Fleischer, K. H., & Belgredan, J. H. (1990). An overview of judgment-based assessment. *Topics in Early Childhood Special Education*, 10(3), 13-23.

Foley, G. (1990). Portrait of the arena evaluation. In E. Gibbs & D. Teti (Eds.), *Interdisciplinary assessment of infants: A guide for early intervention professionals* (pp. 271-286). Baltimore, MD: Paul H. Brookes.

Gradel, K., Thompson, M. D., & Sheehan, R. (1981). Parental and professional agreement

in early childhood assessment. *Topics in Early Childhood Special Education*, 1(2), 31-39.

Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis A. D., Barton, M., Hodgson, S., Green, J., Dumont-Mathieu, T., Robins, D. L., Fein, D. (2008). Diagnostic stability in very young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 38(4), 606-615.

Lelaurin, K. (1990). Judgment-based assessment: Making the implicit explicit. *Topics in Early Childhood Special Education*, 10(3), 96-110.

Peterson, n. L. (1988). *Early intervention for handicapped and at-risk children: An introduction to early childhood special education*. Denver: Love.

Simeonsson, R., Huntington, G. S., Parse, S. A. (1980). Assessment of children with severe handicaps: Multiple problems — multivariate goals. *Journal of the Association for the Severely Handicapped*, 5(1), 55-72.

Thurman, S. K., & Widerstrom, A. H. (1990). *Infants and young children with special needs: A developmental and ecological approach*. Baltimore: Brookes.