

專題評量：評量的類型

形成性評量的效度和信度

蒐集可靠的評量資料

教師總是在進行非正式的形成性評量，這是人類的本能，不斷評斷周圍的人和形勢，然而，絕大多數這類判斷都是無意識的，很可能造成錯誤的印象和理解。形成性評量中的資料要具備效度，必須能夠評量要評量的物件；要具備信度，則必須提供可複製的資訊。

有效度的評量精確指向具體的技能、策略和知識，例如以選擇題檢測數學中的問題解決能力，並不能真正給予教師有關學生問題解決能力的資訊，正確回答這些問題，只能說明學生記住了如何採取問題解決策略，或者說明他們非常善於猜測，但是卻不能說明在真實的問題解決情境中學生會怎樣表現。這些容易評分的評量，對評量 21 世紀技能沒有什麼效度。

Stiggins (2004) 告誡我們：「我們並未努力確保課堂評量的準確性，因此發生不準確評量的機率，以及隨之而來的所有層面錯誤決策，均明顯增加。」(p. 25) 當教師根據太少的資訊對學生知識與能力太快做出決策時，將抑制學生的成長而不是鼓勵成長。

Araison (2001) 描述了一些破壞效度的因素：

1. 刻板印象，根據自己的印象和原先的偏好來下結論。
2. 邏輯錯誤，依據不相干的某些特徵，如穿著打扮或兄弟姐妹的成績等，來評量學生的能力。（這些判斷常常是無意識的，教師本身並不察覺。）
3. 取樣不足，僅根據一次觀察或片斷的資訊做出判斷。
4. 過度概化，假設學生在某一情境下的行為，也會是其他情境下採取的行為。

有關學生表現的資料，也必須具備信度。有信度的資訊必須一致、具有代表性，舉例來說，在放長假前一天所蒐集到的學生思考評量資料，很可能不具信度，因為此時學生的行為必定與平常不同。

教師要做出有效的結論，評量資料必須同時具備效度與信度，亦即能夠展現重要的事物，同時也能夠展現常態。研究人員用「三角交叉檢視法」這個術語來描述從資料得出結論的過程，如同記者在披露犯罪案件的證據之前必須尋求確證，教師在對學生的能力做出結論之前，也需要多方面的資訊，即便如此，結論也是暫時的，能夠接納矛盾的資料，舉例來說，教師可能在小組專題或學習日誌中，發現某個孩子不會歸納，但是之後教師卻發現她在其他學科領域具有歸納能力，於是教師可以暫且得出結論：這個孩子不會歸納是因為學科領域知識不足，而不是因為缺乏思考能力。

絕大多數教師是很警覺的，持續關注他們的學生，因此總是會注意到學生在做什麼和說什麼，可惜的是，他們極少意識到這種非正式的觀察是形成性評量，沒有以系統的方式記錄下他們所見所聞。此類觀察如未經過仔細分析，可能導致偏離真相的視角或者錯誤的結論，因為未能考慮足夠的資料。根據偶然或者非系統方式蒐集到的資料規劃教學，可能阻礙學生的學習。從形成性評量中謹慎蒐集、考慮學生資訊，需要時間與計畫，但是這種評量方式對學生學習和動機的正面影響，讓我們覺得這種努力是很值得的。