

【文／測驗及評量研究中心助理研究員 林宜臻】

美國的「全國教育進展評估 (National Assessment of Educational Progress, NAEP)」被視為大規模教育評量的典範之一，一直扮演著重要教育資料的來源與方法論的革新 (methodological innovation) (Carr, 2004)。美國教育部專員 Francis Keppel 於 1962 年向國會呼籲建立評鑑學生基本知識與技能體系，獲得支持後，於 1963~1969 年間開始規劃相關事宜註 1。啟動之初，原將評估結果定位於促進各州與學校之教育改革，由於擔心評估結果用於不當的州之間與校之間比較，除了由美國各州教育協會 (Education Commission of the States, ECS) 執行外，採取不突出的區塊設計 (unfocused block designs)，不以州為單位，改以針對美國之西北部、東南部、西部及遠西部等四地區的教會學校、私立學校和公立學校的所有學生，並將調查對象以年齡鎖定註 2，概覽其學習表現。此外，以 p 值 (p-values) 檢驗群組之間有否顯著差異；彌封行政部門 (taped administrations)；以抽樣替代全體施測等之措施下，方於 1969 年正式實施。1970 年代初期，美國國會要求教育部減少對 NAEP 的財政預算，有些年，經費僅 200 萬美金，而且只有三位技術專家分析及執行 NAEP 的報告 (Carr, 2004)。因此，在教育部專員 Sydney P. Maraud 的主張下，NAEP 的監督由國家教育資源發展中心 (National Centre for Education Resource Development, NCERD) 移轉至國家教育統計中心 (National Center for Education Statistics, NCES)，NAEP 從此由教育研究轉變成資料的蒐集與分析。

1983 年，美國測驗服務社 (Educational Testing Service, ETS) 取得 NAEP 的指導與管理合同之後，NAEP 邁入精進設計的第二階段，外部技術顧問的心理測量師 (psychometricians) 和統計人員在 NAEP 的技術上發揮重要和關鍵的角色。該階段採用平衡不完全區塊設計 (balanced incomplete block designs)、試題反應理論 (item response theory models)、近似能力值 (plausible values) 等更複雜的矩陣抽樣 (matrix sampling) 設計與非古典的計分方式。

1988 年，公法 (Public Law 100-297 註 3) 的重新授權下，為 NAEP 建立一個新的管理結構。NCES 保留運作與分析技術之責，而由國家評量指導委員會 (National Assessment Governing Board, NAGB) 負責選擇學科領域進行評估，以及 NAEP 各學科領域評量架構 (Lazer, 2004)，NAGB 並被賦予報告及 NAEP 施測結果解說之責 (Elliot & Phillips, 2004)。NAGB 藉由更明確化的內容框架與成就維度 (achievement levels) 將 NAEP 推向標準參照系統 (Carr, 2004)。

1989 年 George Herbert Walker Bush 總統出席與全國州長協會 (National Governors Association, NGA) 的教育會議後，發表聯合聲明，要求對個人、學校和各州的教育進展進行測試。國會授權下，於 1990 年展開試行性州級的 8 年級生的數學測試，而 1992 年測試數學 (4、8 年級生)、閱讀 (4 年級生)，1994 年測試閱讀 (4 年級生)，州級 NAEP 經由試行性階段方於 1996 年正式實施數學 (4、8 年級生) 與科學 (8 年級生) 測試。

NAEP 於 2000 年總統候選人的辯論，開啟另一新階段。候選人都將 NAEP 列為優先的教育議題，但他們不將 NAEP 的角色單一化，而思索幾個可能的角色及與州參與者的新關係。州級 NAEP 原被懷疑能否持續，更不用說蓬勃發展。然事實上，對 NAEP 評估的需求甚至發展至大型學區也自願參加 2002 年城市地區試驗性評估 (Trial Urban District Assessment) (Carr, 2004)。

2001 年「沒有落後孩子 (No Child Left Behind)」的法案規範下，願意參加兩年一次州級 NAEP (State NAEP) 4、8 年級的閱讀和數學評估的州與學區，可獲得聯邦政府補助註 4，而是否參與州級 NAEP 的科學、寫作或其他評估，可依意願。美國國會為此擴增 NAEP 的預算，提供一年超過 100 萬美元的經費，支付州 NAEP 的相關費用。因此，NAEP 有近 3000 名測試管理員，並在每一個州有 NAEP 州協調員。

透明化是 NAEP 第三階段之特色。在 1980 年代初，只有極少數的分析者能瞭解 NAEP，而真正瞭解 NAEP 的複雜數據分析技術者為數更少。但因 NAEP 網站即時提供釋放的試題與所有 NAEP 報告，以及提供使用者容易操作的線上數據分析工具，讓大多數非技術型的使用者也可以自行產出州及州內部學生群體表現的圖表，讓使用者能隨時有所發現註 5。

NAEP 建立之初，不以州為單位，並將調查對象以年齡鎖定、彌封行政部門等措施，避免任何州層面、社群層面註 6，以及任何學生個人層面之相關比較。2001 年「沒有落後孩子 (No Child Left Behind)」法案，讓 NAEP 由全國性的「主要 NAEP (Main NAEP)」與「長期趨勢 NAEP (long-term trend NAEP)」，擴展至與州級 NAEP (State-NAEP)，甚至是 NAEP 城市地區試驗性評估。因此，經費由當初一年 200 萬美金擴增至超過 1000 萬美元。此外，NAEP 也由評量技術掛帥階段，邁向資訊透明化，以發揮最大的影響力。我國 TASA 的工具研發、抽樣、施測、計分等技術毫不遜於 NAEP，然目前報告及試題的釋出尺度，雖可遏止縣市間之不當比較及益於趨勢比較，卻也致使政策決策者及教育相關人員無法充分運用 TASA 資料於制度、課程、教學、評量等之改進。我國 TASA 宜借鏡 NAEP 讓挹注的經費用於訊息透明化，以發揮最大的影響力。

【註解】

註 1 規劃評量的目的、研發工具、抽樣以及資料搜集等事宜。

註 2 9、13 與 17 歲為對象。

註 3 指第 100 期國會所制定的法案中，美國總統所簽署公布的第 297 條法律。該條公共法要求每年須向總統及國會，針對教育指標進行報告。

註 4 Title I funding。

註 5 詳見 <http://nces.ed.gov/nationsreportcard/about/naeptools.asp>

註 6 例城市、市郊、農村等。