

TSA、國際調查：目的及方法（懶人包）

侯傑泰 中文大學教育心理學卓敏講座教授

因應 TSA 在社會上的爭論，一些特首候選人具體提出以抽樣方法替代，究竟是否可行。我們以簡單問答形式作分析。

TSA 導致學生壓力？

問：TSA 究竟有什麼功能？

答：TSA 讓我們知道全港學生學業的強弱，及升降趨勢，這是全港報告；也讓個別學校得知他們全級整體學生在不同學習範圍（如，四則運算、應用題、文法）的強弱，這是學校報告。

問：學生要獲得好成績，豈不是壓力很大？

答：香港與大部分國家和地區相同，進行有關調查時，都不發學生個人成績表，所以 TSA 對學生基本上沒有壓力。

問：TSA 引起爭議，一些人認為學校要獲得好成績，會「操練」學生，這樣不是增加學生壓力嗎？

答：就如我們上學，學校也會發個人成績表。好讓我們知道學習的成果及強弱，得以改善。同理，學校依據 TSA 報告，改善教學，專業地調整某類內容的練習，最終獲益是學生。將所有練習描繪為「有害的操練」，並不合理，也對學校不公平。

問：現在香港學生的讀書壓力是否已經太大？

答：在近年多個大型國際調查研究中，高居學業成績第一名的新加坡，決定於 2021 年在他們的升中派位，不再分分計算成績，改用 8 個能力級別，以減低學生的壓力。香港 1979 年已用 5 組分級，2000 年更改為 3 組。讀書壓力何謂適中，難定標準界線，每個父母標準不同。

人家也有 TSA 嗎？

問：TSA 是否香港獨有？

答：美國在 1960 年已開始建立這類調查，現在世界上差不多所有國家地區都進行 TSA，小如太平洋島國斐濟、亞努亞圖都有。

問：現在 TSA 對學校壓力是否太大，不合理？

答：縱觀世界各地，如澳洲、英國等地，家長在網上可得知所有學校歷年 TSA 成績。在亞洲，韓國及日本部分地區，公眾亦可知學校成績。美國一些學校可以因表現欠佳而被迫關閉，以負責任。香港自 TSA 實施以來，從不公開個別學校成績，現在香港教育局已將學校 TSA 成績從視學項目中剔除。算是國際間令學校壓力最小的類別。

TSA 抽樣可行嗎？

問：一些特首候選人建議抽樣，可行嗎？

答：他們忘記除了全港報告外，TSA 仍有個別學校的報告。抽樣可提供前者，但學校將無法知道自己教學上的問題。

問：學校不是有自己測驗及考試嗎，何需 TSA？

答：要求學校編製一份測試基本能力的試卷，並進行深入複雜的分析，絕不可能。個別學校的測驗及考試，只能反映學校內的學習差異，但不能反映該校與全港小三學生，在同一把量尺上的基本能力學習情況。

問：TSA 還可再優化嗎？

答：尤其是大數據年代，綜觀世界各地，都在不斷利用測評協助教學。改善給個別學校的 TSA「體檢」報告內容、多利用問卷協助學校了解自己學生的特性、及探索發掘一些更合宜的題目種類，都是一些已經開始，但仍需要不斷改良前進的重要課題。當然，欠缺互信，大家寸步難移，苦了的是香港的教育。

取消 TSA，政府獲得掌聲，不是更好嗎？

問：取消 TSA，政府將獲得掌聲，為何仍要堅持？

答：誠如我們監控所有醫院收到的傳染病人，手術失敗率等，是負責任政府的做法。難道教育比醫療更不重要嗎？政府不向每所學校提供「體檢」(TSA)，是負責還是不負責，公眾自行判斷。學校因為 TSA 而有一些壓力，也是自然不過的，就如我們因為體檢發現問題，而去找方法增強體魄一樣。學校應該專業地去改善教學。

問：TSA 真的有助保證教育質素？

答：2015 年在七十多國家地區參加的國際學生能力評量計劃（PISA）中，香港名列第二，顯示香港貧富學生成績差距甚少，扶貧保底做得好。依據 PISA 總結分析，政府應該對所有學校，都共用一把尺（TSA），去展示他們的成績要求；不對學校提供這些信息，可能令一些學校教學滑落而不自知。

問：若 TSA 能提升教學，為什麼我們仍有那麼多能力欠佳的學生？

答：影響學生學習的因素極多，我們進行課程改革、提供學生輔導、及進行 TSA 等，都是盡量處理學生的學習問題，並不代表能解決所有問題。就正如香港建了地鐵，塞車仍是等閒常見。

政府更應善用國際數據及經驗

問：與國際研究相同，TSA 加了學生等問卷。將數個研究結合或做追蹤研究，會有什麼好處？

答：TSA 加了學生等問卷，我們可得知其他影響學生學習成效的因素（如，早餐及與子女閒談的影響）。若能將數個數據結合更有很多其他好處。舉一個最簡單的例子。香港曾有多次調查研究，發現參與補習學生的成績比沒有補習的較差，有人因而下結論：補習浪費金錢，害了孩子。但結合更多數據分析，上述結果的較合理解釋可能是：成績差學生更常去補習而已。

問：為瞭解上述兩種解釋中何者才對，我們為何不用實驗方法？

答：在自然科學領域，我們可以隨機將學生分為兩組，分別接受或不接受補習，再看數年後他們的成績改變。這種隨機控制組實驗的研究方案，可能違反研究倫理規範，對參與研究的學生產生負面的影響，難以彌補。

問：不做實驗，可以怎樣？

答：我們常用追蹤（longitudinal）研究，不作實驗干預，觀察相同成績的學生，一些有參加補習，一些沒有，在數年後對比兩者的成績表現。當然有需要時，再以統計方法調整學生個人背景的差異，從而分析（如，回答補習與成績，誰是因，誰是果）。

問：除了補習，追蹤研究還可解答什麼問題？

答：追蹤研究也可以幫助我們研究教學策略（包括：教學語言）、學童成長（如，睡眠時間）、教學效能等常見教育議題。很多國際研究都是橫斷(cross-sectional)一

次性數據搜集，局限甚大，是這些國際研究的致命傷，也是為甚麼它們需要結合其他數據分析。

國際研究如何結合本港數據

問：香港參加數個大型國際研究，可否利用他它們改善香港教學呢？

答：香港參加了 PISA、TIMSS(學生能力國際評估計劃、國際數學與科學教育成就趨勢調查)等國際研究近二十年。TIMSS 是由國際教育成就評價協會主辦，近 60 個國家及地區小四和中二學生參與，了解他們數學和科學的表現，找出影響學習的因素，去改進教與學成效。經濟合作與發展組織 (OECD) 則每隔三年籌辦 PISA，2015 年共七十多個地區及國家，每地區數千名十五歲學生參與，評估他們閱讀、數學和科學的能力，以及影響學習的因素。香港及很多亞洲地區都表現不錯。這些大型調查對教育很有啟發性，廣受各地政府重視。在香港，教育局分析這些大型國際結果，全面檢討我們的課程、教學策略，從而改善教學；例如，從 PISA 2009 我們知悉香港中學生網路上閱讀的表現後，教育局透過加強與老師的分享，促進電子閱讀的重要，並提升電子閱讀素質的教學策略。

問：香港可否進一步利用這些國際研究去改善教學呢？

答：每次國際研究發布研究結果後，由於與本地其他數據未能結合使用。數據未有進一步善用。所以將國際數據與本地數據合併分析，是向前邁進一步。

問：在進行這類身份配對研究時，具體如何做？有什麼守則？

答：我們須清楚告訴參與的學生及學校，搜集這些數據的目的及日後用途（包括與其他數據合併進行研究）。對個別同時參加兩個研究的學生，我們依他們的姓名班別學號等配對，如，將兩或多份問卷／成績合併。當我們將數個數據合併後，每個學生（或學校）只留下一個虛擬的編碼（如 105 號），供分析時作區分學校及學生之用，研究並非針對某個學生或某間學校，根本無需描述個別學校或學生的行為，又怎會泄露私隱呢？

問：其他先進國家都在這樣做嗎？

答：其實國際經濟合作與發展組織 (OECD)，十分鼓勵利用這些研究數據。如澳洲、加拿大等地均已發布該國的追蹤研究。香港也有主持這些國際研究的學者，自行設計小型追蹤研究。

問：若將同一個學生在不同研究中的問卷等資料合併，需要配對學生的回應。為保障隱私，研究隊伍向教育局交出學生及學校身份以配對學校及學生，這符合研究倫理嗎？

答：首先，這類研究只涉及學校和學生均在一個沒有風險的（不影響升學）及自願參與的考評中的表現及對學習的問卷調查，一般來說（如中文大學），只要徵得中學生及學校同意，已可以搜集。再者，在很多國家是由政府教育局自行進行研究，所以也自然有學生身份，可與其他政府數據配對進行研究。

問：國際上，這樣處理合符研究道德嗎？

答：歐美國家向來對私隱要求更高，他們尚可以進行這類身份配對。很多東西方國家，政府自己收集，或從所聘請的大學研究隊伍得到相關資料配對，國際上也有用這些數據發表文章，且並無引起國際學術界譴責。所以這合併數據仍是符合國際研究規範和道德的。

教育局能否利用國際研究去監控學校？

問：教育局有了學校編碼，便可以用來監控學校，這可能發生嗎？

答：學校的表現通常以兩次考評成績的差異（即增值），作為進步幅度的指標。香港約 500 間中學及 500 間小學。在 PISA 或 TIMSS 等國際研究，每次由主辦的國際機構隨機抽取約三分一學校（150 間）參加。政府若要監控某所學校在兩次研究中（約三、四年間）成績的升降，實際並不可行，因為同一間學校在兩次研究中均被抽中的機率僅得 11%（ $1/3 \times 1/3 = 1/9$ ）。同理，三次均被抽中的學校只佔 1%；隨機抽取 11%、1% 學校，不能作為監控全港每所學校的機制。

問：若某校「幸運」地在兩次研究中均被抽中，我們能否用該校學生樣本去監控該校？

答：這是不可能的，因為學生樣本不能代表該校。例如，按 TIMSS 研究的抽樣，某校的小四（或中二）年級，只有一班會被抽中，整班應考。顯而易見，假設 A 至 D 班是按能力分班，那麼該校被抽到 A 班還是 D 班應考，對學校成績影響甚大。若果 TIMSS 抽到 A 班，則高估該校成績，抽到 D 班則低估該校成績。那麼某一班（如 2A）學生成績怎可能用來監控學校？

問：要是每一間學校的樣本都沒有代表性，那麼整個研究不是沒用嗎？

答：不是的，我們可以由此知道全港的表現。舉另一例子，我們從甲校抽取一兩個籃球隊員，乙校一兩個乒乓球隊員等，全港不同運動員相加，只要取樣正確，足以瞭解香港的學童運動水平，但政府絕不能用這個研究，比較全港不同個別學校的運動水平。同理，在 TIMSS 研究中，我們將全香港不同班別相加，所得的總分仍是充分代表香港的平均成績，但這些數據是不能用來監控個別學校。

問：另一研究 PISA 也能代表香港嗎？

答：PISA 也是用同一道理，在學校中一至中五年級中抽取 42 個 15 歲學生（一般讀中三或中四），在抽樣設計上不能代表該校千多名學生的成績。不過全港樣本合併，便可代表香港 15 歲學生的表現。簡而言之，因為參與學校被抽中用來計算“增值”的樣本量極小，更重要的是每所學校被抽中的學生並不能代表該校，所以教育局也無法用這類國際數據去監控學校。

國際研究數據為何珍貴？

問：就算技術上 TIMSS 及 PISA 不能用作監控學校，這些數據可能被亂用嗎？

答：這是外行的猜想。TIMSS 及 PISA 成績不影響學生升學，參與學生或不一定全力應考。所有分析均是了解全球或香港趨勢，而不是找出那所學校成績更好。

問：就算 TIMSS 及 PISA 數據技術上不能代表個別學校表現，教育局會否錯用工具去比較學校表現？

答：這也是過分擔憂。假若教育局要求前綫人員用直尺去測量溫度、用校長姓氏去決定學校表現等，這些錯誤工具，若真的用於大規模獎賞、懲罰、或監控學校，能不成為國際教育界的笑柄嗎？

問：TIMSS 及 PISA 比香港的數據更珍貴有用嗎？

答：TIMSS 及 PISA 考核的題目及問卷，是世界最頂尖學者編製（如全球勝任力、合作學習能力），高質量且領導全球教育路向，我們不善用 TIMSS 及 PISA 數據，與本港其他數據合併研究，實屬浪費。香港應該與其他先進歐美國家相同，利用 PISA、TIMSS 等數據進行更有用的追蹤研究，為本地教育服務。

後記：TSA、TIMSS 及 PISA 等大型調查研究，涉及頗專業的課題，很易產生誤解。希望我們仍能共同努力，更深入討論，研究如何善用數據，引導我們的教育改革。