

# 「全港性系統評估」(TSA) 政策實施研究

## 概要

2016年10月31日



葉建源  
議員辦事處

Office of IP KIN YUEN  
Legislative Councillor  
(Education Constituency)



# 目 錄

1. 前言：為何進行這研究	2
2. 研究方法	3
3. TSA 的原意與實施之初	4
4. TSA 政策是否達到預期成效？	5
(A) 學校改進 (School Improvement)	
(B) 系統監察 (System Monitoring)	
(C) 問責 (Accountability)	
5. TSA 的副作用	8
(A) 教育局是濫用數據的禍首	
(B) 壓力層層輸送，教師臣服於 TSA	
(C) 試題難度進一步令情況惡化	
(D) 課程狹窄化，教學過程測考化	
(E) 教師「去專業化」(Deprofessionalize)	
(F) 前線對教育局信任的崩解	
6. 改善建議	11
7. 附錄	14
(一) TSA 學校報告範例	
(二) 參考文件《公開考試、學習評估及中小學入學辦法研討會 擬議的核心能力評估》	

# 「全港性系統評估」(TSA) 政策實施研究

## 概要

葉建源議員辦事處

2016 年 10 月 31 日

(本概要扼要敘述報告的要點，供社會人士參考)

## 1. 前言：為何進行這項研究？

### TSA 是什麼？

1.1. 「全港性系統評估」(Territory-wide System Assessment, 簡稱 TSA)是本港由 2004 年開始實施的評估政策，於三個主要學習階段結束前（小三、小六、中三），對學生的中、英、數三科基本能力的評估，而基本能力是指學生在課程中各主要學習階段內必須掌握的知識和能力，以有效地在下一個學習階段學習。理論上，TSA 是「保證學生達到基本水平」的量度，是為普及基礎教育的目標之一，並「旨在反映全港學生水平，而並非個別學校或學生的表現，因此不帶有任何風險」。(《教育制度檢討教育目標諮詢文件》，1999)

### TSA 的爆發

1.2. 2015 年，TSA 爆發成重大的社會議題，互聯網上出現多個數以萬計支持者的關注專頁，批評 TSA 為低年級學生帶來巨大壓力。而在此之前，香港教育專業人員協會（教協）早已進行了三次教師意見調查，也早已指出問題的嚴重性，可惜當局一直不予理會。面對龐大社會壓力，當局宣佈在 2015-16 年度暫停實施小三 TSA 評估一年。

### 官方檢討的不足

1.3. 面對市民、教師以至立法會的壓力，教育局終於在 2015 年底借用一年前已成立的「基本能力評估及評估素養統籌委員會」對 TSA 進行檢討。該委員會很快便發表檢討報告，認為 TSA 方向正確，把問題歸咎在試題過深之上。因此，政府認為可於停辦一年後，於 2016-17 年度全面恢復實施小三 TSA。

- 1.4. 該委員會並沒有包含主要教師團體代表，也沒有包含不同意見的家長群組成員，被批評為缺乏代表性。更重要的是，該委員會並沒有時間和機會，就問題開展深入的研究，被學者批評為「沒有提供任何一般正規公共政策檢討和評鑑研究應有的政策效果實證分析與論證」，就「肯定系統評估的設立原意和價值」，非常荒謬。<sup>1</sup>

### 進行這項研究的目的

- 1.5. 有見及此，立法會教育界代表葉建源議員聘請研究員，專門研究 TSA 政策問題，並結合教協在此之前的三次意見調查，以期更深入分析問題所在，並提出建設性的建議。

## 2. 研究方法

### 2.1. 本研究涵蓋三方面

- (a) 文件研究 (Documentation analysis)
- (b) 數量化研究 (教師意見調查) (Quantitative study)
- (c) 質化研究 (持分者深入訪談) (Qualitative study)

### 數量化研究

- 2.2. 數量化研究包含教協過去五年的三次教師調查，這些調查為本研究提供了持續性的數據，有助確定 TSA 對師生所帶來的壓力程度。

年份	對象	收回問卷數量
2011	小學課程發展主任	142 份
2013	(a) 小學課程發展主任	(a) 88 份
	(b) 小學中/英/數科老師	(b) 1415 份
2015	(a) 小學課程發展主任	(a) 139 份
	(b) 小學中/英/數科老師	(b) 1916 份

### 質化研究

- 2.3. 2016 年展開的是一項質化研究，除檢視相關的政策文件及回顧相關文獻之外，主要是開展了持分者的深入訪談。被訪者包括：辦學團體成員、校長、副校長、主任、教師、學者、香港考試及評核局高層成員，以及前教育局官員，共 18 人。(研究框架從略)

<sup>1</sup> 曾榮光 (2016)。〈全港性系統評估檢討報告的批判：政策評鑑研究的視域〉，《教育學報》，第 44 卷第 1 期，香港中文大學。

## 研究的限制

2.4. 本研究隊伍採取嚴謹的態度，重視研究操守。研究由一位議員推動進行，力求客觀持平，但讀者閱讀時仍須留意。

# 3. TSA 的原意與實施之初

## TSA 的原意

3.1. 在 2000 年梁錦松先生擔任主席的教育統籌委員會（教統會）全面教育改革藍圖之中，TSA「旨在反映全港學生水平，而並非個別學校或學生的表現，因此不帶有任何風險」。<sup>2</sup>由於**不帶任何風險，因此應該不帶有任何壓力**，不會影響正常教學，更不會影響教統會當時鼓吹的「樂於學習，善於溝通，勇於承擔，敢於創新」的改革目標。

## 昔日政府文件一早提出預警

3.2. 在翻查相關文件的過程中，我們發現了當時的教育署在 1999 年撰寫了一份公開文件<sup>3</sup>，就「擬議的核心能力評估」提出一系列考慮因素。該文件仿如「預言書」一樣，點出了 TSA 可能產生的問題，例如：

- 「一項測試的整體設計，取決於測試的目的。……不同的目的，不一定能夠兼容配合。……以『切合目的』為原則，研究有關測試形式、目標學生及測試次數等事宜。」
- 「問責是一個重要的考慮因素。……我們必須充份考慮應否或能否根據測試結果，就學生的成績，向學校、校董會、辦學團體或其他有關方面問責。……問責所產生的第一個問題，是如果測試成績直接引致獎懲，這個測試便關係重大，成為一個高風險測試。……測試可能對學校的教學產生反流效應，例如縮窄／扭曲課程、採納不理想的教學模式，以及進行無意義的操練。」
- 查閱數據方面，須考慮「『不必要的傷害』原則（即公布有關結果或以其他方式傳遞有關結果，不應對當中提及的人士造成不必要的傷害）與『查閱資料的權利』原則（可供查閱的資料應該準確和適當），兩者之間的取捨。」以及「資料被濫用的機會有多大？……當局須避免測試結果遭到濫用，這點至為重要。」
- 「應注意測試數據有其局限性。……不會分辨出表現理想或未符理想的原因；……沒有提供迅速或妥善的方法，以助學校改進。」

<sup>2</sup> 教育統籌委員會（2000）。《終身學習、全人發展：香港教育制度改革建議》。  
(<http://www.e-c.edu.hk/tc/reform/annex/Edu-reform-chi.pdf>)

<sup>3</sup> 教育統籌委員會核心能力測試小組（1999）。《擬議的核心能力評估（參考文件）》，教育署《教師在教育改革中的角色與承擔》。見附錄（二）。

這份由當局撰寫的文件所提出的預警顯然沒有獲得當局的重視，以致當局在 TSA 推行之初，已經走上「歧途」。就以查閱數據一項為例，據受訪的前教育局人員指出，局方並沒有採取任何具體措施（例如指引或守則）防止數據被濫用。（這份預言書至今仍極有參考價值，但已難以尋得，故將全文列為附錄以供讀者參考。）

## 4. TSA 政策是否達到預期成效？

- 4.1. 綜合當局的意見，TSA 的主要目的有三項，前二項較明顯，後一項較隱晦：
- **學校改進**——藉 TSA 評估提供的反饋訊息（Feedback），改進學校的教學。
  - **系統監察**——藉 TSA 提供的評估訊息，讓政府全面掌握整個學校系統的教學表現。
  - **問責**——藉由 TSA 提供的評估訊息，讓政府掌握個別學校的表現。

### （A）學校改進（School Improvement）

#### 缺陷源自設計

- 4.2. 當局一直聲稱 TSA 有助改進學校，學校可以利用 TSA 評估的反饋訊息提高自身的教與學水平，「為未能達標的同學及早進行支援及輔導」。<sup>4</sup>但事實上，並無證據可以證明 TSA 對提高學校的教與學水平有明顯幫助。而這個缺陷，基本上是由 TSA 本身的設計決定的。

#### 回饋訊息僅屬相對表現

- 4.3. 從個別學校的微觀層面看，學校從 TSA 報告獲得的回饋訊息只有兩項：

- 本校學生在某一題的答對率為多少
- 全港學生在同一題的答對率為多少

由此各校可以得出一個結論，即本校學生在該題的答對率到底是高於全港平均，還是低於全港平均，如此而已。這種回饋訊息除了把一半學生「比上去」，把另一半學生「比下去」之外，並無其他特別意義。

- 4.4. 少數校長和老師表示 TSA 報告至少能幫助他們知道學生的相對強弱，從而思考如何改善。然而更多的老師指出，TSA 報告無法讓他們了解學生出錯的原因，只能「靠估」。對於具體改善教學的回饋，如學生如何犯錯，教學

---

<sup>4</sup> 教育局，2015，「全港性系統評估正面睇（三）——評估目的及設計理念」，新聞公報。  
(<http://www.edb.gov.hk/tc/about-edb/press/cleartheair/20151027.html>)

如何改善等，TSA 報告一概欠奉。

#### 對改進學校幫助極微

- 4.5. 從宏觀的層面考察，香港中文大學教育行政與政策學系曾榮光教授曾利用考評局每年發布的報告提供的數據，追蹤分析過去十多年的趨勢，發現學生小三不達標的話，能夠在小六達標的比例極微，每屆僅徘徊於 2.3-5.1% 之間，並逐年下滑，顯示 TSA 對整體的教學改進的幫助極微，跡近於無。如果明白 TSA 報告所提供的反饋訊息極其有限，則這個結果就不難理解了。

#### 專業支援計劃作用未明

- 4.6. 值得一提的是教育局有舉辦一些「校本專業支援計劃」，聲稱是配合 TSA 進行的學校改進手段。可惜教育當局婉拒我們訪談的要求，令我們無法深入了解有關情況。據其他受訪者透露，這些支援計劃到校時會與教師進行「學生前測」、「共同備課」、「後測」等程序。問及為何已有 TSA 數據，但仍需進行前、後測，受訪者表示 TSA 不是單獨存在的工具，而是作為輔助角色，支援人員仍需到校了解個別學校及學生的情況，然後與老師進行共同備課，來改進教學。由此可見，TSA 的功能極為單薄。

#### 效果不彰

- 4.7. 總括而言，並無證據顯示 TSA 實質地達到「學校改進」和「促進學習的評估」的目的。

## (B) 系統監察 (System Monitoring)

- 4.8. 據 2000 年教育改革藍圖的定義，所謂「系統監察」的意義在於：

「為政府及學校管理當局提供全港學校學習範疇水平的資料，以便政府為有需要的學校提供支援和監察教育政策執行的成效。」

(教育統籌委員會 (2000)《香港教育制度改革建議》，頁 58。)

這包括兩點：

- 了解——藉此掌握全港學校學習範疇水平的資料
- 跟進——藉以為有需要的學校提供支援，及監察教育政策執行的成效

#### 能夠了解全港學校表現的資料

- 4.9. 就前一點（了解），當局每年編寫的 TSA 報告應可達到目的。

#### 沒資料顯示有實質的跟進

- 4.10. 但就後一點（跟進），沒有任何資料顯示教育局曾利用 TSA 報告的資料採

**取實質的跟進行動。**議員曾多次在立法會會議上追問，均不得要領。

4.11. 一位非前線教學人士向我們提供資料，教育局官員及學者曾利用 TSA 數據進行一些研究，但沒有向外公開，包括：

- 追蹤不同性別學生的數學成績，隨成長階段的變化
- 追蹤學生家境情況，與成績的關係
- 學習語言（Medium of Instruction）和考試語言（Medium of Test）與成績的關係
- 吃早餐與成績的關係

到底這些資料是否真確，而這些研究又與 TSA 的系統監察功能是否有關，均有待考究。

### **成效有限**

4.12. 總括而言，TSA 有助當局掌握全港學校學習範疇水平的資料，但在為有需要的學校提供支援和監察教育政策執行的成效兩方面而言，沒有證據顯示 TSA 有任何實質成效。

## **(C) 問責 (Accountability)**

### **隱晦的目的**

4.13. 教育局並無任何文件表示 TSA 的功能之一是藉此向學校問責，但根據多方面的跡象，我們有理由相信這是當局推行 TSA 的一項隱晦的目的。

### **曾經是正式問責工具**

4.14. 證據之一，是當局把 TSA 列為質素保證機制的一部分。自 2004 年開始實施，TSA 一直是質素保證機制的評量指標之一。直到 2014 年，因面對很大的社會壓力，政府才在 TSA 一項加上「小學不適用」字眼，並於公開場合多次表示不以 TSA 衡量學校表現，但學生仍須每年進行 TSA 評估。由此可見，**TSA 於 2004-2014 年間，一直是「正式地」作為學校問責指標之一。**

4.15. 受訪的不少校長和老師都認為當局利用 TSA 監控學校，例如：

- 「好肯定的是，教育局的地區辦事處一定有去看（指學校的 TSA 數據），否則他們不會說我們的成績差，不可能是看我們校內的考試吧。」
- 「我理解教育局是用這些數據去 rank 學校（將學校排名次），我自己這樣覺得。」

4.16. 總括而言，「問責」並非 TSA 的原意，但實施之後，成了教育當局隱性的目的，這與下文所述的壓力有極大關係。



## 5. TSA 的副作用

- 5.1. 上文所述，為 TSA 能否達到其預期成效（Intended consequences），這部分則是關於 TSA 實施後的副作用問題（Unintended consequences）。TSA 衍生的副作用非常巨大，導致最終在 2015 年爆發成為嚴重的社會議題。關鍵在於，**TSA 由最初聲稱的「低風險評估」，變成前線校長和老師深信不疑的「高風險評估」。**

### （A）教育局是濫用數據的禍首

- 5.2. 據受訪者表示，TSA 實施之初，校長和老師確信政府的宣稱這是「低風險評估」，不會有任何後果，是「不用準備的評估」。

#### 高官濫用數據

- 5.3. 但早在 2006 年之前，情況已出現變化。有多於一位辦學團體人士回憶指，當時有教育局高官在與多個辦學團體見面時，**拿出各團體屬校的 TSA 成績，非常不客氣地批評個別辦學團體成績不濟，最後更拂袖而去，氣氛極為惡劣。他們憤怒又沮喪，感到教育局無視學校之間的收生差異，對中、英、數成績之外的學校教育更毫不重視。自此之後，他們就感受到來自教育局的強大壓力。**

#### 學齡人口下降與「殺校」

- 5.4. **適齡學童人口下降，「殺校」危機，加深了 TSA 的壓力。**有老師回憶，在 2009 年，其任教學校瀕臨「被殺」，教育局分區辦事處人員向校長直指該校 TSA 低於全港水平，影響派位人數，最終就會「殺校」。另一老師回憶，約 2007 年，任教學校「被殺」，向教育局上訴被駁回，理由有二：不受家長歡迎，TSA 成績差。甚至收生理想、沒有危機的學校的同工也指出，人心惶惶，也不能不重視 TSA。

- 5.5. 在 2004-2014 年間，TSA 成了學校外評的表現指標之一，上文已述，不贅。與此同時，有辦學團體成員更認為，教育局**審批學校申請新校舍時，會以 TSA 成績作為衡量辦學團體的指標。**雖然這一點未能完全證實，但林林總總的利害關係，使辦學團體、校長、老師都飽受來自教育局的壓力。

## (B) 壓力層層輸送，教師臣服於 TSA

### 壓力層層輸送

- 5.6. 教育局向辦學團體施壓，辦學團體則向屬校施壓。有些辦學團體製作排名榜 (League Table)，把學校按 TSA 成績進行排名，向學校施壓。校長又向教師施壓。壓力最終一層一層去到基層老師。無論成績好壞，都難以不受影響。

### 教師的臣服與轉變

- 5.7. 年復一年，局方雖沒有「規管」教師怎樣進行日常教學，但 TSA 的評量方式，逐漸地成為教師行為的指揮棒。一所中產家庭背景學校的校長指出：
- 「我們的老師做事很認真，逐個範疇逐個題目去看的時候 (TSA 報告)，你看得很仔細的話呢，於是你會催促自己，催促自己於這一方面，是否要再做好些呢？於是老師就自不然，慢慢地進行自我調較，自我調較是指例如『這一類型的題目，我們做得不是太好吧，去年也是這樣，那麼我們是否需要於我們的課堂裏面，於這方面加強？』這樣地，於是呢，經過一段日子，時間短，兩三年呢你不會發覺的，但經過現在這十多年，當你再回望 TSA 早期實施的時候，我會看到，其實老師的心態是會有不同，自然會變成追趕著 TSA 題目去走。」

## (C) 試題難度進一步令情況惡化

- 5.8. 上述濫用數據的壓力傳遞是 TSA 異化的主要因素，而試題難度令情況進一步惡化。

### 理論上 TSA 應屬基本能力範圍

- 5.9. 按照既定政策，TSA 屬於「基本能力評估」(Basic Competency Assessment)，題目應該屬於基本能力的範圍。但事實上，不少受訪教師都認為 TSA 題目相當深，並非只有基本能力的學生所能回答。

### 現實上 TSA 題目越來越深

- 5.10. 綜合受訪者的意見，早期 TSA 題目的確較容易，設題確以「基本能力」為範圍。然而其後逐漸變化，設題者希望把不同能力的學生分出來，因此把部分題目加深了。由於校長和老師認定 TSA 屬於基本能力範圍，當學生表現不理想時，便認為他們連基本能力也未能達到，便大為緊張。

5.11. 但由於種種原因，例如 TSA 原先理念無法傳承，出題者出題能力參差不齊，再加上題目不可重複、題型要有變化的原則，甚至為了防止商營出版社的補充練習「捉路」，題目便越出越深，題型越出越刁鑽。這趨勢令坊間的補充練習也不斷攀比難度，而前線教師也要教得越來越深，操練也越來越嚴重。

## (D) 課程狹窄化，教學過程測考化

### 課程狹窄化

5.12. 普遍學校的教學過程和考測均受到 TSA 的干擾，而程度上有不同。曾經面對嚴重收生壓力的學校，受到最大的影響，**TSA 主導了整個課程、教學和考測，是 TSA 對教育帶來的最嚴重的扭曲**。幾位受訪校長教師如此分享：

- 「向著 TSA 進發，譬如我校就差不多 100% 向準 TSA 方向，基本上每天的教學就是教 TSA。」
- 「我自己最著重的是中英數三科裡面將課程狹窄化，或者 TSA 化，……教學的內容狹窄到只是看著 TSA，這個是我自己覺得最可悲和最心痛。」
- 「例如，要教一些技巧，好像現在去閱讀一條題目，首先要『問』了些 main point (重點)，到底問題問甚麼，然後要小朋友拿支筆去 locate (找尋) 個答案出來，於一篇文章裡面，叫 Q1, question one，個答案原來 question one 是從這裡找到，小朋友呢，就寫個 multiple choice (多項選擇題) 的答案，我強調是用 multiple choice 的答案，佢用支筆，去填個洞洞，日日都是這樣做。可是，例如我們以前讀書那些 sentence making，根本是少了好多。他們現在只懂得填洞洞，填洞洞，填洞洞。如果要寫句子呢，他們不懂得寫。」

### 測考 TSA 化

5.13. 教是如此，考也一樣，校內的考試測驗，無論題型、格式，都變得與 TSA 一模一樣，以便學生適應 TSA 的考法，令整個教育從低年級開始，便跌入「應試教育」(Teach to the test) 的局面，難以自拔。

## (E) 教師「去專業化」(Deprofessionalize)

### 失去原有方向感

- 5.14. 隨著教師被引領進入機械式的「應試教學」後，教師的教學變得簡化，並且失去自身原有的方向感，以 TSA 全面代入。久而久之，造成教師「去專業化」(Deskill / Deprofessionalize) 的局面。
- 5.15. TSA 令不少教師花大量時間和精力於「TSA 專業」上，而壓縮了以「全人發展」為目標的教師專業及教師價值。幾位老師表示自己和周圍發生這樣的轉變：
- 「開始不是去想如何教書了，而是怎樣去考好這個試（指 TSA），如何於教書裡面去追這個 TSA。……已經將個 syllabus 放在一旁了。」
  - 「後來我發現，原來是因為之前一兩年，校長逼得他們（指同事）很緊要呀，要有成績呀，我說這樣其實不會有成績。簡單說，就是很急於要小朋友懂得答題目，其實要學生懂得答一條題目，老師要準備很多，去建構這個知識，現在沒有了，整個數學教育最重要的，具體到抽象，沒有了。」

## (F) 前線對教育局信任的崩解

- 5.16. 隨之而來，是對教育局信任的崩解。多位受訪校長和學者均不約而同地指出，**比較 TSA 的成績並不公平，因為學校本來就收取背景非常不同的學生。**不只是受到施壓的學校，TSA 成績良好的學校校長雖然沒受到施壓，但卻於校長之間的群體中知道教育局或辦學團體利用 TSA 數據向其他學校施壓，因而感到教育局不再是可以依賴的後盾，因而對教育局失去信任。
- 5.17. 2015 年 12 月 11 日教育局發出的通告把操練的責任通通推給學校，引起掀然大波，校長們群情洶湧，也與此有直接關係。而檢討 TSA 的委員會組成不公，無意認真檢討，也看在前線校長和老師的眼裡，對教育局的信任也進一步土崩瓦解。

## 6. 改善建議

### 6.1. 開展對 TSA 的深入研究

本研究的目的，是嘗試有系統地分析 TSA 問題的成因和現狀，進而尋求解決之道。但由於種種的局限，本研究只能視作引玉拋磚，當局應認真委託有

公信力的學術機構展開深入研究，作為決定 TSA 何去何從的堅實基礎。其實需要研究的課題很多，例如一個 TSA 工具能否同時負起促進教學和系統監測兩大任務，又例如 TSA 報告是否有可能及如何提升學校的教與學，更重要的是整個制度的壓力如何消弭。

## 6.2. 不應恢復小三 TSA 評估

TSA 三個階段評估之中，爭議最大的是小三評估，其次是小六評估。小三評估之所以特別受關注，是因為學生年齡很小，評估令整個壓力和變質問題向小學低年級蔓延，禍害極大。因此，在未能改變異化變質的惡果、有可靠的對策和所有持分者放心之前，不應恢復小三 TSA 評估。

## 6.3. 重整「應試文化」

事實上，TSA 是由教育局主導的「總結性評估」(Summative assessment)，它是標準化的測驗，無法照顧學校之間的差異，也不能向學校提供個別學生的成績資料，根本不適宜作為「促進學習」的工具。因此應讓位於由老師主導、貼近學生情況的「進展性評估」(Formative assessment)，以完成促進學習的重要任務。事實上，「總結性評估」和「進展性評估」各有功能，但在初小階段，後者無疑應佔主導角色。

	現時 TSA 的做法	真正能夠做到「促進學習的評估」的做法
主導者	教育局及考評局	教師
進行評估的時間	小三、小六、中三結束前	教學過程中
獲得回饋的時間	進行評估後數月才能獲得學校報告 <sup>5</sup>	隨即，及時
教師獲得的回饋	評估性的回饋 (Evaluative)： 以百分率和評級為主	形容性的回饋 (Descriptive)： 除百分率和評級外，還可透過試卷明白學生如何犯錯，跟進學生不明白之處
跟進個別學生	不能	能夠
數據的使用	將學生成績與全港對比，可能被教育局或辦學團體濫用，造成侵害專業的壓力輸送	由教師自己使用，藉以了解學生的學習

<sup>5</sup> 見附錄（一）TSA 學校報告範例。

#### 6.4. 重新思考「問責」

一向以來，教育界的「問責」往往是單方面的，即學校向教育局問責。但越來越多學者指出，應更重視「智慧型問責」(Intelligent Accountability)，其精粹在於：

- 強調「信任」
- 強調「相互的責任」(Mutual Responsibility)
- 結合和平衡「內在問責」(Internal accountability)與「外在問責」(External accountability)(從略)

#### 6.5. 重建信任

TSA 引發的危機，是教育局管治威信受到嚴重質疑，教育界對局方的信任崩解，為此，本研究提出以下兩個重點：

**對過多的監控放手**，簡言之，是減少各種操控和問責的手段，降低進行各種評量的頻率與規模。

**尊重前線，尊重不同意見，認真諮詢**主要的持分者，包括校長團體、教師團體達成共識。

#### 6.6. 以兒童健康成長為最高原則

這是不證自明的重點，在過去一段時間裡，這項原則似乎被忽略和輕視了。我們希望今後的政策釐定，勿以行政方便、監察方便等等為先，而應以兒童身心的健康成長為首先考慮。

附錄 (一)

TSA 學校報告範例

中國語文 Chinese Language (Sub-paper: Reading and writing)

題號 Item No.	選項 Options	得分 / 等級 Score/Grade	學校百分率 School Percentage	全港百分率 Territory-wide percentage
<p>Q1 和 Q2 是封閉式題目，多項選擇題，考核學生能否選擇正確答案。以 Q1 為例，正確答案為「D」，甲校的答對率是 60.4%，而全港的平均答對率是 71.8%，故甲校在此一題的答對率低於全港平均水平。</p>	<p>Q1</p> <p>A B C D* U#</p>		<p>7.3% 12.9% 19.4% 60.4% 0.0%</p>	<p>7.2% 11% 9.3% 71.8% 6.8%</p>
	<p>Q2</p> <p>A* B C D U#</p>		<p>63.4% 12.2% 17.1% 7.3% 0.0%</p>	<p>50.6% 25.8% 17.0% 5.7% 0.9%</p>
<p>Q3 和 Q4 是開放式題目，以等級為標準的題目(如作文)。「4」是最高的評級，「0」是最低的評級。以 Q3 為例，甲校有 19.5%的學生於此題目取得最高的第 4 級，相對於全港的 17%。</p>	<p>Q3</p> <p>0 1 2 3 4 U#</p>	<p>0 1 2 3 4 U#</p>	<p>2.4% 12.2% 35.4% 30.5% 19.5% 0.0%</p>	<p>3.1% 7.9% 42.6% 29.4% 17.0% 0.0%</p>
	<p>Q4</p> <p>0 1 2 3 U#</p>		<p>0 1 2 3 U#</p>	<p>2.4% 24.4% 57.3% 15.9% 0.0%</p>

\*正確答案 correct answer

#沒有答題或答案無效 no or invalid answer

## 附錄(二)

### 參考文件 公開考試、學習評估及中小學入學辦法研討會 擬議的核心能力評估

#### 目的

本文件旨在提出下列有關核心能力測試的問題，並載列小一至中三中、英、數核心能力測試的構思，以徵詢公眾意見：

- 核心能力測試的目的為何？
- 測試結果將作何種用途？由何人使用？
- 測試的形式及內容為何？測試的對象是誰？
- 測試應由哪個機構主辦？

#### 核心能力測試的目的為何？

2. 一項測試的整體設計，取決於測試的目的。然而我們須特別注意，不同的目的，不一定能夠兼容配合。因此，我們必須就建議測試的目的達成共識，並了解各個測試方式的利弊，然後以“切合目的”為原則，研究有關測試形式、目標學生及測試次數等事宜。

3. 測試的其中一個主要目的，是在教與學的過程中，讓教師和學生接收「回饋」。透過測試，可得悉學生的強項和弱項，為成績好的學生安排精修課程，需要幫助的學生則可獲得輔導支援；而教學策略亦可因應學生的情況予以改善。

4. 問責是一個重要的考慮因素。隨著當局推行校本管理，以及把更多職責下放給學校，我們必須制訂一個問責架構。就透明度而言，市民(特別是家長)都希望得知更多關於個別學校的資料。因此，我們必須充份考慮應否或能否根據測試結果，就學生的成績，向學校、校董會、辦學團體或其他有關方面問責。

#### 與問責有關的問題

5. 問責所帶出的第一個問題，是如果測試成績直接引致獎懲，這個測試便關係重大，成為一個高風險測試。因為這個測試結果代表了學校能否教導出優秀學生，並會被用來評定學校的等級或衡量其表現。測試越重要，學校便會把注意力放在測試的準備上，而所花的時間也越多。如測試不是用來量度重要而有意義的內容、能力及知識，便會浪費了寶貴的時間及資源。由此而引起的是測試可能對學校的教學產生反流效應，例如縮窄/扭曲課程、採納不理想的教學模式，以及進行無意義的操練。



6. 美國最近的研究結果凸顯了一個問題：不少學校把教學重點放在提高測試分數上，致令測試結果失去真正效用。研究結果亦顯示，雖然測試分數有所提高，但學生的知識可能是完全沒有增進的；同時，由於高風險測試通常會引致測試失準或測試分數謬誤，研究指出應避免讓全國性的測試出現這些情況。

7. 另一重要問題是學校之間的比較。我們須制定措施，確保學校之間的比較是基於公平及準確的原則，即是我們必須考慮學校的增值表現。為此，我們須知悉學生的社會經濟狀況，以前的成績，以及基線的訂定等資料，但要持平地查究這些資料是極端困難的。

8. 測試結果亦能客觀而真確地反映學生的學習結果。換言之，問責機制亦可輔助學校進行自我評估和作出定期滙報。這會有助學校監察在達致全港教育目標及校本目標方面的進度，以及在有需要時，對教與學作出改善。

**測試結果將作何種用途？由何人使用？**

### 系統層面測試

9. 在系統層面的監察方面，核心能力測試是要描繪本港現時的教育發展狀況，以便能解答一些問題，例如：

- 學生學到多少知識？他們有哪些強項和弱項？
- 有關科目的整體水平正在上升還是下降？學生現時學到的東西是否較五年或十年前為多？
- 把兩類學校(例如官立學校與資助學校)的學生比較，他們的表現有什麼差別？
- 在什麼程度上，現行教育制度能達到所訂的教育目標？

### 學校層面測試

10. 在學校層面的監察方面，核心能力測試的目的可以是評估學生的學習進度，使學校獲得客觀的測試數據，從而透過自我評估，改善教與學的成效。測試結果可解答下列問題：

- 學校本身有沒有進步？
- 與其他背景相若或學生質素相近的學校比較？有什麼差別？
- 與全港所有其他學校比較，有什麼差別？

此外，學校層面的數據亦可用以向學校問責或提示學校/教師致力提高教學水平。英國和法國公布類似學校排名表的資料，而美國和澳洲維多利亞省則公開全

國測試中學校層面的數據。因此，這些學校層面的數據可能非常敏感，而第 4 和 7 段提及的問題亦將會出現。

### 學生層面測試

11. 藉著結合測試與學習，為學習者提供「回饋」，以助確定目標，並鼓勵他們為自己的學習負責。所收集的數據亦會用以協助未能發揮潛能的學生。測試數據有助解答一些問題，例如：

- 某一所學校、某一班別或某一學生在有關科目中有哪些強項和弱項？
- 哪些學生有學習困難及需要輔導？
- 教學大綱及工作安排中，哪些範疇需要修訂，以及須改善哪些教學策略和方式？

### 核心能力測試結果的使用者

12. 下列人士可能會利用測試結果作所述用途：

- 教師可分析測試結果，以改善教學方法及幫助個別學生(學生及學校層面的數據)。
- 校長可分析測試結果，以協助個別教師，並與各教師合作策劃改善校內教與學的質素(學生及學校層面的數據)。
- 決策者可透過測試結果，找出有關系統或系統其中一部分的優點和缺點，並引導各有關方面致力改善教與學的質素(系統及學校層面的數據)。
- 教育署的督學及其他人員可以測試結果作為指引，以決定如何向學校及教師提供技術協助及支援(系統、學校及學生層面的數據)。
- 負責課程發展及編寫課本的人士可從測試結果中找出教材及課程的優點和缺點(系統層面的數據)。
- 師資訓練人員可透過測試結果找出師資訓練課程的優點和缺點，並作出適當的改善(系統層面的數據)。
- 測試設計人員可分析測試結果，以改善測試的設計，並匯報學生的成績是否有所改變(系統、學校及學生層面的數據)。
- 家長可利用有關資料輔助子女的學習(學生層面的數據)。

從核心能力測試中獲得的資料，對校本管理、質素保證視中的自我評估，以及課程整體檢視，均十分重要。因此，應在上述各項工作間設立一個介面，以確保測試結果運用得宜。

## 查閱數據

13. 一般意見均贊同，系統層面的數據可以向所有有關人士公布，但學生層面的數據則不能公開。然而，下列與學校層面數據有關的問題仍有爭議：

- 應否容許公眾和家長查閱學校層面的數據？
- 在什麼程度上，校董會能夠根據學校層面的數據來判斷校長是否稱職，以及在什麼程度上校長可以根據這些數據來評估教師的表現或確定哪些教師不稱職？
- 政府應否用學校層面的數據，把各學校評級或加以比較？
- 應否容許學校公開一些資料，使學生之間可進行比較，或提供任何刊物把學校比較或評級？

14. 關於上述問題，須考慮的事項如下：

- (a) 對學校表現的知情權與保護學校和學生免受不公平比較的責任，兩者之間的取捨；
- (b) 「不必要的傷害」原則(即公布有關結果或以其他方式傳遞有關結果，不應對當中提及的人士造成不必要的傷害)與「查閱資料的權利」原則(可供查閱的資料應該準確和適當)，兩者之間的取捨；
- (c) 應在哪個層面透露資料？(例如學校層面的數據只限學校查閱，不得公開班級、教師及學生的數據等)；
- (d) 應由何人收集和提供資料？
- (e) 資料被濫用的機會有多大？
- (f) 有關表現的資料在多大程度上有助提高教與學的成效？

## 一般意見

15. 當局須避免測試結果遭到濫用，這點至為重要。關於這方面，使用者應注意測試數據有其局限性。這些數據：

- 只是在所根據的原則和基準範圍內才可使用；
- 不會分辨出表現理想或未符理想的原因；
- 沒有說明日後的理想表現應如何；
- 只採用一種評估工具，因此必須與其他定性和定量資料一併研究；
- 沒有提供迅速或妥善的方法，以助學校改進。

因此，測試數據，特別是學校和學生層面的數據，應與其他定性及定量證據來源一併研究，例如學校的背景、歷史和學生質素，學生的生理和心理狀況等。更重要的是，為免把焦點過分集中於學業成績，應制定範疇指標，以更全面反映教育的成果。

## 核心能力測試的形式及內容為何？測試的對象是誰？

16. 測試題目的形式大體上分兩類：構建式回應題目及選擇式回應題目。在構建式回應題目中，學生必須通過構思，作出回應(例如論說文式測驗或表現能力的課業)；在選擇式回應題目中，學生則有多個答案可以選擇(例如選擇題、搭配題或是非題)。

17. 很多人主張測試必須涉及高層次的技巧和才能，例如解決問題、探究及分析的能力等，因而必須採用遠較傳統測試更為”真實”或切合實際的課業。我們也確實難以從選擇題中，評核學生在這類高層次技巧方面的表現。在統一測驗中廣泛採用的選擇題形式，一直備受嚴厲批評。然而舉例來說，雖然英國的全國測試系統採用能力表現測試，但要全國劃一施行所有測試，以及判斷學生的表現，是相當困難；因而令到測試數據的質素也有問題。另一方面，在某些國家舉行的全國或全省測試中(例如澳洲新南威爾斯省的基本技能測試[Basic Skills Test]、維多利亞省的學習能力測試計劃[Learning Assessment Project]，以及智利的國家基礎教育質素測試計劃[National Program for Quality Assessment of Basic Chilean Education]，則只採用或主要採用選擇式回應題目。

18. 其實，每種形式都各有優劣。我們在決定測試題目或課業的形式時，往往要作出取捨。在權衡各種形式的利弊時，必須考慮有關測試對教與學和財政方面的影響、以及測試的可靠性和效力。

19. 若以有三個層面數據(即系統、學校及學生層面)的假設下，可能採用的題目形式、目標學生及測試次數，見於下表：

系統層面數據		
題目形式 ● 以選擇題較為可取(省時、成本低、較客觀、可包括較多題目)	目標學生 ● 有代表性的樣本，例如每級的 1/30(全港每級約有 2500 個樣本) ● 測試通常每隔二至三年舉行一次，而非每年一次	測試數目 ● 測試甚少在每級或每班舉行，但會在小學階段舉行一至兩次及在中學階段舉行一次 ● 可採用矩陣設計
學校層面數據		
題目形式 ● 同時採用構建及選擇題目	目標學生 ● 一所學校的全體學生或每所學校取一	測試數目 ● 所有關鍵階段基本上每年一次

	個有代表性的學生樣本 ● 一些關鍵階段結束時，例如小三、小五及中三	● 可每年 1/3 的學校或每年都抽樣本或三年內每年 1/3 樣本，然後一次全體測試
學生層面數據		
題目形式 ● 同時採用構建及選擇題目	目標學生 ● 所有級別(小一至中三)的學生	測試數目 ● 每年一次或兩次

20. 倘若資源許可，在全國測試中同時採用兩種形式的題目(可分兩次舉行)，是較為可取。總括而言，測試形式應取決於測試的目的、技術上是否可行(例如測試對象的數目)、測試密度，以及對財政的影響。

### 內容及標準

21. 用以監察表現的核心能力測試(學生、學校及系統層面)是與課程有關，因此其內容涵蓋方面須較為全面。測試應包括課程內所有重要範疇，而不應像傳統考試一樣，只抽取少部分內容進行測試。此外，亦必須訂立測試規格，列明核心能力測試所涵蓋的主題及各種技能。在這方面，舉辦核心能力測試的一個重要先決條件，是製備一套有關基本能力的核心/主要元素，以及有關科目中較低層次及較高層次能力的資料。另外亦須訂立內容標準(列出教什麼及何時教)和有關目標(界定標準適用於哪些學生、符合標準的學生比例及達致標準的時間)。長遠來說，我們須要制定一套附有各等級範例，界定不同表現水平的標準。

### 測試應由哪個機構主辦？

22. 應委派哪個機構主辦建議的核心能力測試？應該是政府機關、外界機構，還是其他機構？

23. 教育署在舉辦標準測驗方面具有專業知識和經驗，例如香港學科測驗，即涵蓋小一至中三的中文、英文及數學科目。本港學校多年來一直採用香港學科測驗，作為教與學成效的分析判斷工具。

24. 把核心能力測試的設計交由信譽卓著的評核機構負責，例如香港考試局，並由外界技術及課程專家提供協助，亦是一個可行的辦法，以兼取兩類人士的專長，同時可以增加核心能力測試的一般認受程度和可信性。此外，若本地沒有所需的專才，便可能要在海外聘請。在智利，國家測試(SIMCE)是由教育部與一所

大學(Pontificia Universidad Catolica)聯合舉辦，在三至四年內分三個階段推行。該三個階段分別是：由大學主辦測試、把測試轉交教育部負責，以及教育部管理測試。

25. 不論由什麼機構主辦測試，該機構必須有良好聲譽，即工作質素高、具專業技術及公平正直。該機構需要有多方面的專才，包括計劃管理、研究設計、課程分析、測驗及問卷設定、抽樣測試、印刷及分發、蒐集、處理和分析資料與及撰寫報告。

### 核心能力測試的建議模式

26. 教統會核心能力測試小組(該小組)經過多番深入研究後，建議首先制定中文、英文及數學科的核心能力測試(有關測試)。有關測試包括：

(a) 小一至中三各級的網上測試。這些網上測試由電腦輔助進行，目的是：

- 評估小一至中三學生的中文、英文及數學科成績；
- 分析學生在這些科目中的強項和弱項；
- 找出能力稍遜的學生，為他們安排輔導，而優秀的學生，則為他們開辦精修課程；
- 讓學校可監察各級每年在這三科達致的成績水平。

網上測試採用互動形式，並會按照學生的能力，調節題目的難度。網內會建立一個大型資料庫，儲蓄大量客觀，以及與課程有關的題目，以支援網上測試。同時，我們會鼓勵學校制定課業式試題，以作補充。有關部門應為學校提供足夠的樣本，用作參考或經修訂後採用。

(b) 在重要的教育階段，例如小三、小五、中一和中三，進行由中央統籌的保密筆試。這些測試圍繞學校課程，目的是監察學生在系統和學校層面的表現，因此，只須抽選部分學生參加測試。(請參閱第 19 段)

27. 待有關測試準備妥當後，便會訂定中文、英文及數學科的最低限度成績水平。長遠而言，亦會為這些科目訂立表現水平。

28. 小組展望日後應為其他學習範疇，例如科技及情意和社會範疇，設計核心能力測試。

### 有關測試的目的

29. 有關測試與學位分配無關，所以不屬高風險測試。有關測試的作用是監察學生中、英、數三個學習範疇系統層面表現，以及找出學生的強項及弱項，以便學

校調整教學策略，改善教學。小組並未就是否設立學校層面測試一事達成共識，並會在諮詢公眾後提出最後建議。假若學校層面測試會在將來舉行，小組認為有關測試資料只可供政府參閱，以便由政府考慮為有需要的學校提供額外資源，改善質素。

## 推行策略

30. 香港尚須培養出使用測試數據以改善教與學成效的文化。我們須制定策略性的推行計劃，列出每個推行階段應達致的最終目的及目標，以推廣這個文化。並確保各有關方面，例如教師、校長及家長，在設立核心能力測試後，懂得妥善使用測試數據。

31. 網上測試起初只在小三推行，但最終會在五年內擴展到小一至中三所有級別。小三的網上測試將於二零零一年以試驗形式推行。中學方面，中一的網上測試會在二零零四年推出。在網上測試推出前，各中、小學仍會沿用香港學科測驗。長遠來說，全面拓展及推行核心能力測試需要八至十年時間。

## 意見及建議

32. 請各界踴躍提出意見及建議，並於一九九九年十二月十五日之前送交教育統籌委員會秘書處，地址是香港中區政府合署中座 714 室(傳真號碼：2537 4591)，或以話音郵件(中文留言請電 1833 133，英文留言請電 1833 122)或電子郵件(電郵地址：div3@emb.gcn.gov.hk)發表意見。

教育統籌委員會  
核心能力測試小組

一九九九年十月

---

教育統籌委員會(1999)：〈參考文件公開考試、學習評估及中小學入學辦法研討會擬議的核心能力評估〉。《教育統籌委員會教育改革建議專題研討會參考文件》，頁 46-55。

鳴謝

參與研究的一位不願透露姓名的學者

協助審閱本報告的多位學者

接受訪問的18位教育同工

葉建源議員辦事處的同工

教協教育研究部

以及其他曾提供各種協助的朋友

全港性系統評估 (TSA) 政策實施研究 · 概要

葉建源議員辦事處, 2016

作者：葉建源

研究員：林璐茜

地址：香港中區立法會道1號立法會綜合大樓919室

電話：(852) 3468 7222

傳真：(852) 2771 8752

網頁：[www.ipkinyuen.org.hk](http://www.ipkinyuen.org.hk)

臉書：[www.facebook.com/ipkinyuen](http://www.facebook.com/ipkinyuen)

© 版權所有 可作非商業用途轉載或複印，敬請註明出處。