

英國國定課程評量的沿革 與其對十二年國民基本教育課程評量的啟示

盧雪梅*

英國 1988 年的教育改革法案通過國定課程與評量的實施，按照學生年齡分為四個關鍵期，在各關鍵期末（7 歲、11 歲、14 歲和 16 歲）安排學生接受評量，並公布學校的表現，供家長作為選擇學校的參考。本文藉由文件與文獻的分析，回顧並闡述自 1987 年國定課程評量規劃期以降，各朝政府重大的評量政策、實務與課題，並探討其脈絡與影響。本文同時檢視與評析當前國內配合十二年國民基本教育改革推出之重大的評量方案與措施，在國內既有的基礎上，參酌英國國定課程評量實施的經驗，提出十二年國民基本教育課程評量之規劃與實施建議，包括：根據課程綱要訂定各學習階段的期望表現或標準，強化「臺灣學生學習成就評量資料庫」(TASA) 的學力監控功能，整合 TASA 協助縣市辦理學力檢測計畫與補救教學科技評量，重新思考國中教育會考的目的與定位，及規劃與推動全方位的教師評量。

關鍵詞：十二年國民基本教育、十二年國民基本教育課程評量、英國國定課程評量

* 盧雪梅：國立臺灣師範大學教育心理與輔導學系副教授
(通訊作者：smlu@ntnu.edu.tw)

Evolutions of the National Curriculum Assessment in England and Its Implications for the Assessment of the 12-year Basic Education Curriculum in Taiwan

Sheue-Mei Lu*

The British Education Reform Act of 1988 introduced the national curriculum and national assessment in England and Wales. The curriculum was structured into four key stages, and pupils were assessed at or near the end of each key stage (ages 7, 11, 14 and 16). The school performance tables were published to help parents make informed decisions. This article aims to present an account of the development and changes in the England national curriculum assessment from its original blueprint in 1987 to the current time through reviewing official papers, documents, and related academic articles. This article also examines the current assessment projects and programs for the 12-year Basic Education in Taiwan. Based on the recent assessment innovation efforts of the 12-year Basic Education and the assessment experience of England, the author proposes suggestions for the design and implementation of the assessment of the 12-year Basic Education Curriculum as follows: 1) Establish expected achievement standards for each learning stage based on the 12-year Curriculum Guidelines. 2) Enhance the long-term trends monitoring function of National Taiwan Assessment of Student Achievement (TASA). 3) Integrate regional TASA and remedial instruction-technology-based testing programs. 4) Re-clarify the purposes of the Comprehensive Assessment Program for Junior High School Students. 5) Plan and implement comprehensive teacher assessment policies.

Keywords: *12-year Basic Education, assessment of 12-year Basic Education Curriculum, national curriculum assessment in England*

*Sheue-Mei Lu: Associate Professor, Department of Educational Psychology and Counseling, National Taiwan Normal University
(corresponding author: smlu@ntnu.edu.tw)

英國國定課程評量的沿革 與其對十二年國民基本教育課程評量的啟示

盧雪梅

壹、前言

英國（指英格蘭與威爾斯）在 1988 年通過教育改革法案（Education Reform Act 1988），其中頒布國定課程與評量的實施，擬透過中央集權與市場競爭機制，提升國家教育水準與經濟競爭力，是英國二戰之後影響最深遠的教育改革。國定課程規定對義務教育階段學生，學校應該教授的基礎科目和學習內容，按照學生的年齡分為四個關鍵期（Key Stages），在各關鍵期末（7 歲、11 歲、14 歲和 16 歲）安排學生接受評量，並公布學校表現，作為家長選擇學校的參考（Daugherty, 1995）。由於威爾斯的評量政策在 2002 年之後逐漸與英格蘭分道揚鑣，舉例來說，威爾斯在 2002 年廢除第一關鍵期（7 歲）的全國測驗（Daugherty, 2009）。本文焦點放在英格蘭的國定課程評量政策與實施的沿革。

英國國定課程評量採外部全國評量與教師評量雙軌並行，一般而言，學生學習成就監控的評量，都採外部評量，然英國亦將教師評量納入，此設計在國際間堪稱獨樹一格。其自規劃期迄今，歷經保守黨（~1997）、工黨（1997~2010）、保守黨和自由民主黨組成的聯合政府（Conservative and Liberal Democrat coalition government）（2010~2015）執政，當前由保守黨（2015~）執政，各朝政府的政策與實務雖各有其思量與優先，但以國定課程評量為國家教育水準的監控機制，始終如一。國內專文探討英國國定課程評量並不多見，陳明印（2004）介紹國定課程評量制度；吳麗君（2000）以五個「交互損益」為架構，回顧國定課程評量前十年的發展；吳麗君（2001）比較蘇格蘭和英格蘭小學階段的國家考試。至於探討近期評量改革的研究論文近乎闕如，舉例來說，在 2014 年登場的新課程，做了實施以來最具突破性的改革，不再以成就層級（Level of Attainment）來架構課程的成就目標（Department for Education [DfE], 2013a），連帶的影響到評量結果報告的方式，自 2016 年起，全國測驗報告以量尺分

數 (scaled scores) 取代沿用二十餘載的成就層級 (DfE, 2014a)。

國內十二年國民基本教育 (以下簡稱十二年國教) 的多項改革正在進行中，評量改革依然是眾所關注的一環。英國國定課程評量一路走來，已超過四分之一世紀之久，他山之石，可以攻錯，本文主要藉由文件和文獻的分析，回顧並闡析自 1987 年規劃期以降，各朝政府重大的評量政策、實務與課題，探討其發展脈絡與影響，並從中擷取可供國內當前教育改革借鏡與反思之處。

貳、保守黨執政期 (1987 年~1997 年)

1988 年教育改革法案頒布時的執政黨是保守黨，本節先概述國定課程的重點，再闡述保守黨政府重大的評量變革及其脈絡。

一、國定課程概述

英國教育與科學部 (Department of Education and Science) 在 1987 年 7 月公布《5 到 16 歲的國定課程：一份諮詢文件》(*The national curriculum 5-16: A consultation document*) (Department of Education and Science and Welsh Office [DES/WO], 1987)，此文件內容成為 1988 年教育改革法案的重要依據。國定課程規定對義務教育階段公立學校 5 至 16 歲學生，學校應該教授的基礎科目和學習內容，按學生年齡分為四個關鍵期，其中 5 至 7 歲為第一關鍵期 (KS1)，7 至 11 歲為第二關鍵期 (KS2)，11 至 13 歲為第三關鍵期 (KS3)，14 至 16 歲為第四關鍵期 (KS4)。國定課程有十科基礎科目 (foundation subjects)，包括三科核心科目 (core subjects) ——英文、數學和科學，七科其他基礎科目 (other foundation subjects) ——歷史、地理、技術、音樂、藝術、體育和現代外語，現代外語從中學開始修習 (黃光雄、周淑卿，1992；DES/WO, 1987)。各科目的課程綱要包含成就目標 (Attainment Targets [AT]) 和學習方案 (programmes of study)。成就目標指對不同能力與成熟度之學生在各關鍵期結束時所具備的知識、技能和理解能力 (knowledge, skill & understanding) 的期望。各成就目標用十個成就層級 (level of attainment) 界定 5 到 16 歲學生的學習進步 (progress)，一般學生預期每兩年進步一個層級，每一個層級以成就敘述 (Statement of Attainment [SoA]) 描述學生應知、應了解和應該能做的為何。學習方案陳述各個關鍵期必須教

給學生的重要事項 (matters)、技能 (skills) 和程序 (process)，以協助他們達到期望的成就層級；藉由各關鍵期末的評量，了解學生在成就目標達到層級為何 (黃光雄、周淑卿，1992；DES/WO, 1987)。在 1989 年 9 月，由 KS1 (5 歲) 全部科目與 KS3 (11 歲) 核心科目率先登場 (Daugherty, 1995)。

二、國定課程評量的設計

教育大臣 Kenneth Baker 在《5 到 16 歲的國定課程：一份諮詢文件》(DES/WO, 1987) 公布不久後，成立評量與測驗任務小組 (Task Group on Assessment and Testing [TGAT])，由 Paul Black 擔任的召集人，負責規劃國定課程評量系統，TGAT 正式報告書 (DES/WO, 1988) 於 1988 年元月公布。TGAT 功成身退，課程業務交由新成立的國家課程諮詢局 (National Curriculum Council) 負責，評量則由學校考試與評量局 (School Examination and Assessment Council [SEAC]) 負責。TGAT 規劃之評量系統要點摘述如後 (DES/WO, 1988)。

TGAT 提出四項國家評量設計的優先準則：(1) 評量結果應對學生達成目標的成就狀況，提供直接的資訊，故評量應採標準參照 (be criterion-referenced)；(2) 評量結果應對學生進一步學習的需求提供做決策的依據，故評量應是形成性的 (be formative)；(3) 不同班級與學校的成績或等級若要相互比較，教師、學生和家長須使用共同的語言與標準，故評量結果需要校準或調整 (be calibrated or moderated)；(4) 標準和量尺的建立與使用應與預期的教育發展進程相關聯，以賦予不同年齡學生評量的連續性，故評量應該與進步 (progression) 有關聯。TGAT 特別強調，「除非符合這些準則，否則評量支持學生學習和教師專業發展的潛在價值將無從實現」(DES/WO, 1988)。

TGAT 兼顧專業理念與教育大臣的指示，提出國家評量資訊應該服務下列目的：(1) 形成性的 (formative)：使學生的成就得以被辨認和討論，以利適當計畫下一步的學習；(2) 診斷性的 (diagnostic)：檢視學生學習困難，提供適當的補救性協助和輔導；(3) 總結性的 (summative)：透過系統性的方法來記錄學生的整體成就；(4) 評鑑性的 (evaluative)：藉以評估並報告學校、地方教育局的某些業務與教育服務 (DES/WO, 1988)。

基於上述原則與目的，TGAT 將國定課程評量設計為具多重目的，採用多元評量和標準參照取向的系統。成就層級不只是課程目標，也是標準參照評量的基礎，作為報告學生成就和監控 (track) 其學習進步的架構。國定課程評量採雙軌並行，包括教

師評量 (Teacher Assessment [TA]) 和各關鍵期末的全國評量，稱為標準評量作業 (Standard Assessment Tasks [SATs]) (Daugherty, 1995)。TGAT 指出，若分別就不同目的發展工具與收集資料，師生的負擔不可能維持在合理的範圍內，其建議 16 歲之前的評量 (KS1~KS3)，以形成性評量和教師評量為優先，簡要的說，教師在日常教學過程中運用多元策略對學生進行觀察與評量，在關鍵期末，根據成就目標的成就敘述，判斷最適合學生的成就層級為何，並藉由團體校準程序 (group moderation) 提升教師評量的一致性，以全國評量結果調整教師評量結果，彙整和統計各學校和班級的評量資料做為評估學校和教師效能的資料，如此，國定課程評量可以達成服務形成 (含診斷)、總結和評鑑的目的 (DES/WO, 1988)。此外，為降低全國評量對學校課程與教學造成回浪效應 (wash-back effect)，也就是，考試引導教學的效應，TGAT 建議 SATs 的形式採用操作、口語表達和書寫活動，也可以用小團體來進行。SATs 由外部機構研發，但由教師在課堂內實施，教師可以彈性將評量活動和課程結合，也由教師自己計分，使課室的教師評量與外部 SATs 的特徵，沒有明顯的差別 (DES/WO, 1988)。

TGAT 報告書公布後，多重目的迎合教育大臣對學校評鑑的期待，也受到教師團體的掌聲，因為比他們原先預期的好多了 (Daugherty, 1995, 1997)，然而沒有獲得當時首相 Margaret Thatcher 的青睞，因為系統過於複雜且昂貴，而且太重視教師的判斷 (Daugherty, 1995)。在 1988 年六月，教育大臣公布國家評量的七項原則，接受 TGAT 大部分的建議，只有團體調整這一項受到質疑，教育大臣認為太複雜且所需經費過於龐大，但也沒有立即否決，TGAT 大部分的建議成為國定課程評量的政策 (Daugherty, 1995)。

三、國定課程評量重大的事件與變革

根據 TGAT 的建議 (DES/WO, 1988)，若要充分獲得國家評量的效益，至少需要五年的準備期，不過教育部已迫不急待促成全國評量上路。在 1989 年，SEAC 開始研發 KS1 和 KS3 核心科目的評量工具 (SATs)，原則上按照 TGAT 的建議，採取實作取向，由教師負責施測、計分和記錄，期望教師能從中獲得有益的回饋 (Daugherty, 1995)。以下回顧並闡述保守黨政府重大的評量事件及其脈絡。

(一) 第一關鍵期 (KS1) 全國評量登場

在 1990 年 5 月，KS1 核心科目舉行預試，由於採取標準參照評量，部分教師就

三科核心科目成就目標 (AT) 的成就敘述 (SoA)，對每一位學童進行將近 200 個判斷和記錄，直呼吃不消 (Black, 1994; Daugherty, 1995; Nuttall & Stobart, 1994; Torrance, 1993)，以首版的課綱為例，英文有 5 項 AT，157 個 SoA；數學有 14 項 AT，296 個 SoA (Black, 1994)。在 1991 年 5 月，KS1 首次正式施測，雖然評量內容已經過瘦身減量，然而在一個約 25-30 人的班級中，教師平均仍要花 44 個小時在評量上。教師除了抱怨工作負擔重，也抱怨學童受教時間被打擾得太久 (Torrance, 1993)。在 1992 年，KS1 全國評量又再度簡化，減少活動式評量，增加團體考試部分 (Daugherty, 1995)。

(二) 第三關鍵期 (KS3) 全國評量和教師杯葛

KS3 核心科目在 1991 年的預試，也遇到類似 KS1 的情況，同年 6 月，教育大臣 Kenneth Clarke 在某公開場合中批評 SATs 預試材料複雜到沒有道理的程度 (elaborate nonsense) (Black, 1994; Daugherty, 1995, 1997; Whetton, 2009)。教育大臣以教師抗議為由，將評量方式改變成固定時間的筆試，1992 年的預試即改成筆試 (Daugherty, 1995; Nuttall & Stobart, 1994)。由於 SAT 與美國教育測驗服務社 (Educational Testing Service) 發展的學術性向測驗 (Scholastic Aptitude Tests) 之簡稱雷同，國定課程評量自 1992 年底起不再用 SATs 一詞，KS1 評量改稱標準作業 (standard tasks)，KS2 和 KS3 評量則稱為標準測驗 (standard tests) (Daugherty, 1995)，不過民間和學校教師仍沿用 SATs 來稱呼全國測驗。

KS3 首次正式測驗預定在 1993 年 5 月登場，然而教師團體在 3 月發動集體罷考，教育部採取法律行徑來遏止罷考，但法院判決教師團體勝訴，認定罷考是合法的勞資之爭，不是意圖煽動違法，教師不能被要求做額外無酬勞的計分和記錄工作 (Black, 1994; Daugherty, 1995; Whetton, 2009)。雖然教育部堅持考試照常舉行，但超過 95% 的中學罷考 (Nuttall & Stobart, 1994)。

(三) 國定課程與評量的檢討與改革

由於教師團體的杯葛，教育大臣在 1993 年 4 月請 Ron Dearing 組織跨黨派小組，負責重新檢視國定課程與評量，Dearing 在當年 12 月提出報告書 (Dearing, 1993)，當中關於評量改革的要點說明如後：(1)KS1 至 KS3 的全國評量繼續維持，但科目只限於英文、數學和科學三個核心科目；KS4 評量仍以「中等教育證書」考試 (General Certificate of Second Education [GCSE]) 的成績為據；(2)減少課程綱要成就目標的數量，並停止用成就敘述 (SoA) 來描述對學生的期望，改用層級描述 (level descriptors)

對學生期望成就提供廣泛統整性的描述，並將十個層級改成八個層級，再外加一個特殊層級 (special level)，且層級描述只適用 KS1 到 KS3 的學生。KS4 學生之 GCSE 通過的成績增加 A*一級，從原來的七級 (A 到 G) 改成八級 (A*到 G)，期望標準為 C 級以上；(3)學校表現榜 (school performance tables or league tables) 賦予教師評量與全國測驗相等的地位 (equal standing)，以強調教師的支持與投入對國家評量的重要性。

在 Dearing 改革之後，各關鍵期學生成就層級範圍和期望層級 (expected level) 整理如表 1。KS2 首次正式全國測驗在 1995 年登場，一開始就採用固定時間的筆試。由於教師杯葛事件，自 1995 年起，KS2 和 KS3 全國測驗改由外部機構計分，但 KS1 全國評量仍然由教師實施和計分。學校表現榜公布 KS2 和 KS4 的測驗結果，提供給家長作為選擇學校的參考，KS1 和 KS3 學校的表現雖沒有對外公布，但個別學生和學校的結果會通知家長 (Whetton, 2009)。

表 1 英國國定課程的關鍵期、年齡與成就層級

關鍵期	學生年齡	年級	成就層級	期望層級
KS1	5-7	1-2	1-3	2
KS2	7-11	3-6	2-5	4
KS3	11-14	7-9	3-7 (8)	5 或 6
KS4	14-16	10-11	GCSE 的標準	C 以上

(四) 教師評量

教師評量 (Teacher Assessment [TA]) 是國定課程評量系統的一部分，國定課程實施後，有幾位教育大臣曾在一些公開場合強調教師評量在國家評量的地位，不過事後證明，都是口頭之惠而已 (lip services) (Daugherty, 1996)。TA 如何實施？如何根據成就敘述來記錄學生表現，並將之轉化為成就層級，以利評量資料符合國家評量的目的？是教師們關切的問題，教師們寄望政府舉辦在職訓練並給予支持性資源以利達成目的 (Daugherty, 1995, 1996；McCallum, McAlister, Brown, & Gipps, 1993)。SEAC 初

期將大量資源挹注在 SATs 的研發和實施上，對於教師評量幾乎沒有明確的政策和指引 (Daugherty, 1995, 1996; McCallum et al., 1993)。在 1990 年底，SEAC 終於出版三套教師評量指引和輔助材料，後來又陸續出版一些成就層級評定的樣本供教師參考，對於這些速成且未經試驗的材料，教師們反映，對他們的幫助有限 (McCallum et al., 1993)，舉例來說，SEAC 出版的輔助指南和材料，有的將 TA 界定為教學的一部分，提供回饋做為後續教學的計畫，著重 TA 形成性評量角色，有的強調 TA 收集學生表現做為其關鍵期成就層級判斷的依據，著重 TA 總結性評量的角色，或者兩者兼顧，教師們的困惑仍然未解，換言之，SEAC 對 TA 的目的與性質也沒有確切的掌握 (McCallum et al., 1993)。此外，在公布學校表現一事上，教師評量有球員兼裁判之嫌，也面臨到一些角色張力，由於政治人物對 TA 資料信任度不足，TA 在國家評量報告的地位被邊緣化 (Black, 1998)。後來雖然根據 Dearing 報告書 (Dearing, 1993) 的建議，學校表現榜將 TA 和全國測驗的結果並列，強調教師的支持與投入對國家評量的重要性，不過在工黨執政不久後，TA 再度從學校表現榜消失。

四、小結

TGAT 企圖建造一個能提升學生學習，尊重並強化教師專業角色，及滿足公共績效法令要求的評量系統 (Black, 1994)，規劃了具備多重目的，運用多元評量和採用標準參照的評量系統。TGAT 的設計充滿專業理想但未經實徵試驗，再加上主政者迫不及待推動改革，在未準備充分的情況下，課程評量實施後，問題陸續浮現。以全國評量為例，因採實作取向，首當其衝的問題就是可行性 (manageability)，教師抱怨評量負擔沉重，此外，主政者也擔憂教師 SATs 的計分品質，全國評量不久後轉為時間固定和計分標準化的筆試。國定課程評量實施經不斷修正後，與 TGAT 的藍圖漸行漸遠，簡而言之，重全國測驗而輕教師評量，評鑑性目的遠遠凌駕在其他目的之上 (Daugherty, 1995; Whetton, 2009)。Black (1997) 在十年後的回顧中坦承，對一個以考試做為市場選擇機制的政府來說，寄望其支持這樣複雜的評量系統，是個過於天真的想法。

參、工黨執政期（1997 年～2010 年）

在 1997 年 5 月大選後，工黨上台執政。大選期間，首相 Tony Blair 宣稱其施政的三大優先將是「教育、教育、教育」（“Education, Education, Education”），宣示其提升教育水準的決心（Issacs, 2010；Whenton, 2009）。在工黨執政期間，國定課程經過三次改革，分別在 1999～2000 年、2007～2008 年（針對中學課程），及 2008～2009 年（針對小學課程）（House of Commons: Children, Schools and Families Committee, 2009），不過沒有影響到評量安排。以下就工黨政府較重要的評量政策、配套與改革做回顧與說明。

一、國定課程評量與教育目標管理

工黨政府在 1997 年 7 月公布教育白皮書《追求卓越的學校教育》（*Excellence in Schools*）（Department for Education and Employment [DfEE], 1997），提出未來五年的教育發展計畫，其中將提升小學生的語文和數學水準，列為優先項目之一，並定下 2002 年 KS2（11 歲）全國表現的目標：至少 80% 學童的英文成就達到期望層級（層級 4），75% 學童的數學成就達到期望層級。為達成此目標，《追求卓越的學校教育》白皮書提出小學每天至少安排一節語文課和一節數學課，並分別訂在 1998 年和 1999 年啟動國定語文策略（national literacy strategy）和國定數學策略（national numeracy strategy），以全國測驗結果做為評鑑學校和地方教育當局之語文和數學策略實施成效的指標（DfEE, 1997）。對於被列為表現不佳的學校，教育標準署（Office for Standards in Education [Ofsted]）將加強視導，學校將被要求限期改善，甚至遭到改管或關閉的命運（DfEE, 1997）。Reed 與 Hallgarten（2003）以擁抱 3T 文化：測驗、目標和表現榜（Tests, Targets and Tables），來形容工黨政府的評量政策。

二、學校表現榜的變革

自 1999 年起，小學（KS2）學校表現榜只公布全國測驗的結果，在各校達到期望層級以上之學生百分比之外，增加改進指數（improvement measure）一項資料，也就是各校前三個年度，三科核心科目在期望層級以上之學生百分比的總和；同樣的，KS4

公布各中學前三個年度，五科 GCSE（須包含英文和數學兩科）成績在 C 級以上的學生人數百分比。自 2003 年起，開始公布 KS3 學校表現榜，此外，因考量各校輸入條件（intake）的差異，中小學的學校表現榜再增加的增值分數（value-added scores）一項資料，也就是，各校學生在某特定關鍵期的進步指標（progress measures）。自 2004 年起，按照《14~19：擴充機會、提升標準》（14~19: *Extending opportunities, raising standards*）綠皮書的建議（Department for Education and Skills [DfES], 2002），KS4 學校表現榜採納更多職業課程的證書，將職業證書的成績轉換成 GCSE 的等級，公佈各中學五科 GCSE 或 GCSE 等值證書（GCSE equivalence）（五科中須包括英文和數學）成績在 C 級以上的百分比。Reed 與 Hallgarten（2003）指出，政府宣稱學校表現榜公布的資訊只會增加，不會減少（more information, not less）。補充說明的是，學校表現榜的資訊雖一再增加，但仍以達到期望層級的學生百分比為主要的績效指標，此外，教育部只公布學校的資料，並沒有排序，但媒體會自行將學校排名，這又給學校增添一重壓力。

三、自選測驗的實施

自選測驗（Optional Tests[OP]）於 1997 年登場，最先推出的是小學 3、4 和 5 年級的測驗，稍後推出中學 7 和 8 年級的測驗，目的在幫助教師與家長評估與監控在關鍵期評量之前，學生在各年級朝向期望層級進步的情形。OP 由資格與課程局（Qualifications and Curriculum Authority）研發，但不屬於法定的測驗（statutory tests）。OP 有英文和數學兩個科目，以 3、4 和 5 年級的測驗內容來說，英文科包括閱讀、寫作和拼字三種測驗，數學科包括數學和心算數學（mental math）兩種測驗，均採時間固定的筆試方式。OP 材料由學校自行下單訂購，包括測驗材料（含心算數學播放的 CD）、教師施測和計分指引、成就層級切點資料，計分和登記的程式等等，材料可以重複使用。此外，OP 也發行視障和聽障學生適用的調整測驗材料（Qualifications and Curriculum Authority, 2008）。補充說明的是，OP 約 3~4 年設計一套全新版本。在 2008 年 10 月，教育大臣宣布廢止 KS3 全國測驗後，為了協助學校與家長了解學生的成就，OP 又再增加 9 年級英文、數學和科學三個科目（Isaacs, 2010）。

四、第一關鍵期（KS1）評量改革

各方對全國測驗的批評聲浪從未停止過，尤其是讓 7 歲學童面對測驗的壓力，格

外受到詬病。在 2004 年，QCA 試辦 KS1 評量安排的調整，全國測驗施測時間由學校在規定的期限內自行安排，測驗結果用來輔助教師評量的判斷，KS1 國定課程評量以教師評量結果為據，由於試辦反映良好，自 2005 年起全面實施。值得一提的是，此項政策仍然延續至今，主要原因是，KS1 評量結果不公布，也不做為學校績效指標用（Isaacs, 2010；Whetton, 2009）。

五、單一層級測驗的試辦

教育與技能部（Department for Education and Skill）在 2007 年元月公告《大躍進：學校如何幫助每位學生提升進步？》（*Making good progress: How can we help every pupil to make good progress at school?*）諮詢文件（DfES, 2007），提出大躍進試驗方案（Making Good Progress [MGP] Pilots），原因之一是 KS2 在 2005 年後之全國測驗的達標率增加緩慢（Torrance, 2011；Whetton, 2009）。MGP 試驗方案意圖用新的方式來評估、報告並激勵學校進步，以使沒有孩子在任何階段停滯不前。MGP 有五個主軸，單一層級測驗（Single Level Tests [SLTs]）是其中一軸，自 2007 年 9 月起跑，試驗期訂為兩年（DfES, 2007）。

單一層級測驗（SLTs）是全國測驗的變通方案，只要學生經教師判斷已達到某特定成就層級，即可接受外部測驗，不必等到關鍵期末。SLTs 對象為 3 到 8 年級的學生，科目為英文科（包括閱讀與寫作）和數學科。SLTs 一年舉辦兩次，分別在六月和十二月，學生若被教師判斷已到某個成就層級，即可登記參加該層級的考試，結果只有通過或不通過兩種情況。全國測驗是由多個層級試題組成，例如 KS2 測驗包含層級 3~5（或 6）的試題，經由標準設定程序設定各層級的切點分數，根據學生整體答題表現賦予成就層級，如同國內國中教育會考的做法。

六、「促進學習的評量」相關政策與配套

MGP 試驗方案有兩軸與評量有關，在 SLTs 之外，另一軸為「促進學習的評量」（Assessment for Learning [AfL]），此軸的目的在協助教師運用 AfL，定期追蹤學生在英文和數學次層級（sub-level）的進步情形，了解他們的學習現況，以計畫下一步的學習（DfES, 2007）。在此補充說明，由於層級描述涵蓋範圍頗大，實務操作上將層級 2、3 與 4 三個層級又各再細分 a、b 和 c 三個次層級（Department for Children, Schools and Families [DCSF], 2010），以利學生短期內的進步更容易彰顯出來。

值得一提的是，AFL 當前在國際間已蔚為一個新興評量典範 (James, 2011)，在此略說明其源起，以利了解政策制定的背景。英國評量政策壓力團體「評量改革團體」(Assessment Reform Group [ARG]) 的兩位成員，Paul Black 與 Dylan William 在 Nuffield 基金會的贊助下進行研究，他們回顧自 1988 年之後以英文發表之形成性評量相關研究與文獻約 250 篇，進行統合分析，舉實徵證據闡述與論證課堂的形成性評量才能真正提升學生的學習成就，而且對弱勢族群的效果量更是明顯，研究成果以〈評量與教室裡的學習〉(*Assessment and classroom learning*) 為篇名發表 (Black & William, 1998a)。他們建議政府，若要提升學生的成就，改進課堂的形成性評量才是正確之道，而不是用全國測驗 (Black & William, 1998b)。此後，ARG 試圖用淺顯易懂的語言幫助非學術圈人士了解課堂評量如何提升學生學習，改用平易近人的用語——「促進學習的評量」(assessment for learning) 與「對學習的評量」(assessment of learning) 取代「形成性」與「總結性」的學術性語言。ARG 對 AFL 的界定是：

「促進學習的評量是一個尋找與詮釋證據的歷程，學習者和其教師可以用這證據決定學習者在學習軌道的何處，需要再往何處前進，以及如何採取最佳的方法抵達那裡。」(ARG, 2002, p.2)。

官方文件首次使用 AFL 一詞是在 2003 年公布的《卓越和樂趣：小學國家策略》(*Excellence and enjoyment: A strategy for primary schools*) (DfES, 2003)。AFL 相關政策的高峰是教育部在 2008 年四月公布的《促進學習的評量策略》(*The assessment for learning strategy*) (DCSF, 2008)，預計在 2008 年到 2011 年間，提撥 1.5 億英鎊的經費補助學校發展與運用促進學習的評量 (Whetton, 2009)。《促進學習的評量策略》經費主要挹注在擴大推動「評估學生進步」方案 (Assessing Pupils' Progress [APP])，APP 是一種結構取向的教師評量 (structured approach to teacher assessment)，目的在幫助教師評量能有效且可信地與國定課程的成就層級連結，以監控個別學生的學習進展並給予適當的介入 (Stanley, MacCann, Gardner, Reynolds, & Wild, 2009)。

相較於 ARG 關注的是課堂形成性評量的實施和品質，官方關切的是，教師對學生表現成就層級的判斷和進展監控，協助教師提高總結判斷的品質與評定結果的一致性。雖然官方對 AFL 的詮釋與實施重點與 ARG 的定義不盡相同，但 AFL 對政治人物的影響力逐漸擴增，是不爭的事實 (Daugherty, 2007; James, 2011)。

七、第三與第二關鍵期（KS3 與 KS2）評量的改革

在 2008 年 10 月 14 日，教育大臣 Ed Ball 在無預警的情況下，宣佈廢除 KS3 的全國測驗，其認為 KS4 學校表現榜的資訊已足以讓家長和學生作為選校的參考，也能作為評估中學教育績效之用（Isaacs, 2010；Whetton, 2009）。Ed Ball 提議 KS3 學校表現榜改成公布教師評量結果，並增加 KS3 階段的自選測驗（OP），協助學校監控學生的成就。一般認為 KS3 測驗廢除的導火線，是當年暑假負責 KS2 和 KS3 全國測驗計分機構耽延將測驗結果送回給各學校（Isaacs, 2010；James, 2011；Whetton, 2009）。

KS3 全國測驗廢除後，要求 KS2 測驗一併廢除的聲浪頗大，教育大臣在當年 11 月由成立專家小組，對未來國定課程評量安排進行研究。專家小組在 2009 年 5 月提出報告與建議，要點包括：(1)強化教師評量的品質與對家長的報告；(2)KS2 的全國測驗應該繼續維持，因為小學教育品質仍需要績效監管，但只考英文和數學兩科，科學改成抽樣施測（DCSF, 2009）。自 2010 年起，KS2 科學的全國測驗即採用抽樣施測，這項政策沿用至今。

八、小結

在 2008 年公布廢除 KS3 全國測驗之前，國定課程評量的實施大致在穩定中求改變。工黨政府將全國測驗結合目標管理策略，做為評估學校績效責任與提升全國教育標準的主要手段，使得全國測驗成為高風險測驗（high-stakes tests），對學校的課程與教學造成不少負面效應，為考試而教學（teaching to the test）是相當普遍的現象，非測驗科目被忽略或排擠，課程學習內容被窄化，此與提供學生廣泛且均衡（broad and balanced）課程的宗旨反而背道而馳。要求全國測驗改革的聲浪，從未停止，雖然自 2005 年起，KS1 全國測驗鬆綁，自 2009 年起，KS3 全國測驗廢止，但主政者仍不肯放手小學（KS2）與中學（KS4）學校課責政策。在教師評量方面，相較於保守黨政府，工黨政府有較明確的教師評量政策與配套，例如關鍵期間年級的自選測驗、促進學習的評量與評估學生進步方案等，目的在藉由教師評量監控關鍵期間學生的進步情況，以協助其達到關鍵期的期望層級，提高全國測驗達標學生的比率，至於課堂的形成性評量仍然非官方關注所在，James（2011）批評道，政治人物對學生的表現（performance）的關切仍超過對其學習（learning）的關切。學校表現榜仍然獨尊全國測驗結果，除非全國測驗停止實施時，教師評量才取而代之。

肆、聯合政府以降（2010 年～）

在 2010 年 5 月大選後，保守黨和自由民主黨組成的聯合政府上台執政，其在當年 11 月公布教育白皮書《教學的重要性》（*The importance of teaching*），提出未來的教育藍圖，其中第四章針對課程、評量和證書提出改革方針（DfE, 2010）。在 2015 年大選後，由保守黨執政，國定課程與評量基本上延續聯合政府的政策。聯合政府對國定課程做了自 1994 年後最大幅度的改革，本節先概述國定課程的改革，再闡述重大的評量改革暨相關政策的脈絡與要點。

一、國定課程改革

聯合政府上台後，即宣布對國定課程進行全面改革，擱置工黨政府原訂 2011 年 9 月起實施的小學新課程。國定課程審議在 2011 年元月啟動，此次課程改革的目的有三：(1)確保新課程持著嚴謹且高的標準，並提高學校教學內容的一致性；(2)所有學生都學習到各學科最重要的知識；(3)除核心科目外，其他基礎科目賦予教師更大的自由度以發展學生的潛能（DfE, 2013a）。在 2014 年 9 月，KS1 到 KS3 所有年級同步實施新課程，唯小學 2 年級與 6 年級的核心科目晚一年登場，因為這兩個年級的學生在次年 5 月要接受關鍵期評量。KS4 的英文和數學自 2015 年起逐年級實施，科學自 2016 年起逐年級實施（Roberts, 2014），非核心科目之 GCSE 課程與證書考試也自 2016 年起陸續改革（Office of Qualifications and Examinations Regulation, 2014）。

新課程最突破性的改革，是課綱不再以層級和層級描述來架構 KS1 到 KS3 的成就目標，不同科目成就目標的組織方式也不再統一規定。具體言之，KS1 核心科目的成就目標分 1 年級和 2 年級陳述；KS2 的英文和數學則分為前期（3 和 4 年級）和後期（5 和 6 年級）兩階段敘述，科學則各年級分開陳述。小學（KS1 與 KS2）的其他基礎科目和中學（KS3 和 KS4）所有科目，都以關鍵期為單位來陳述，而且只規範基本的學習內容，賦予學校和教師更多的彈性（DfE, 2014b）。此項改革不只影響法定評量（包括全國測驗與教師評量）的報告方式，也影響到學校對學生關鍵期內學習成就報告與監控的方式，教育部成立無層級評量委員會（Commission of Assessment Without Levels），協助各校發展其評量和監控學生學習進展的方式（Commission of Assessment

Without Levels, 2015)。

二、評量改革與相關政策

(一) KS2 評量與學校課責機制檢視

為降低小學為因應全國測驗而過度練習的負面效應，同時維持堅實的小學課責機制，《教學的重要性》白皮書提議，對 KS2 全國測驗和學校課責機制進行獨立審視(DfE, 2010)。教育部在 2010 年 11 月委請 Lord Bew 成立小組負責此任務。Lord Bew 在 2011 年 6 月提出報告書，指出學校外部的績效衡鑑機制對提升教育水準與學生進步有其重要性，然而也應該提高法定教師評量的重要性和地位。在評量方面，較重要建議包括：數學、英文的閱讀與部分的寫作能力（如拼字、標點符號和文法）維持全國測驗，英文的寫作（作文）與科學採用教師評量，科學的全國測驗改成抽樣施測。在學校課責方面，建議學校績效責任指標，在成就目標達成情形外，也應該兼顧進步的情形，以提高公平性。自 2012 年起，KS2 評量如 Bew Lord 的建議實施（DfE, 2011）。

(二) 一年級學生拼讀能力檢測

《教學的重要性》提議對 6 歲學生進行拼讀能力檢測，以即早篩出有閱讀困難的學童。自 2012 年起，將 1 年級生的拼讀能力檢測（phonics screening check）納入法定評量。拼讀測驗在每年六月中旬舉行，目的在確認學生拼讀解碼能力（phonic decoding）是否達到期望標準，採個別施測，由教師負責實施和計分，評量結果通知家長，也呈送給地方當局。沒有通過檢測的學生，第二年須再接受評量（DfE, 2013b）。

(三) 英格蘭中學文憑和 KS4 學校表現榜證書的改革

《教學的重要性》指出，和 2000 年相比，KS4 學生修習學術性課程證書的人數比率大幅下滑，修習 GCSE 等值職業課程的人數比率大增，因此，提議創立英格蘭中學文憑（English Baccalaureate [EBacc]），以鼓勵學校提供更廣泛的學術性課程給學生修習（DfE, 2010）。EBacc 指英文、數學、科學、歷史、地理和外國語文等六類學術性的 GCSE 科目。自 2011 年起，學校表現榜增加公布各校 EBacc 在 C 級以上的學生百分比。

此外，根據 Wolf (2011) 的職業教育評鑑報告書，部分中學為提高表現榜的成績，

鼓勵學生修習容易過關但對未來就業或訓練無益的職業科目。因此，教育部在 2012 年將上千種 GCSE 等值的職業證書從學校表現榜移除，且自 2014 年起，唯有經過教育部審查且核可的職業證書才能納入表現榜的採計，以確保學校提供學生貨真價實的課程 (DfE, 2015)。

配合課程改革，教育部在 2013 年宣布，自 2016 年起，學校表現榜將採計八個科目證書 (DfE, 2013c)，所謂八科目包括：英文和數學兩科證書，三科 EBacc 證書，三科經核可的證書，稍後對此將有較詳細的說明。

(四) 建立學校績效責任的樓地板標準

《教學的重要性》提議，建立中小學績效責任的樓地板標準 (floor standard)，此政策自 2013 年起實施。樓地板標準界定 KS2 和 KS4 各校學生成就和/或進步的最低期望標準，對於未能達到標準的學校，將視情況施予處置，包括書面警告、增加 Ofsted 的視導次數、撤換校長 (head teacher)，或將學校轉換成特許學校 (academy) (DfE, 2010)。配合新制評量的改革，2016 年將啟用新的樓地板標準，稍後將以 2016 年為例，說明標準如何界定。

(五) 小學的評量改革

新制評量自 2016 年起啟動，小學的評量安排與結果報告說明如後 (DfE, 2014a)。

1. 入學基準線評量

學生在進入小學就讀後的幾週內，必須接受入學基準線評量 (reception baseline assessment)，其主要目的是，在 KS2 結束時，作為計算學校進步指標 (progress measures) 的起點。基準線評量施測由教師負責，教育部自 2015 年二月起陸續公布一些已經核准的評量工具供學校選用 (Standards and Testing Agency & Department for Education, 2015)。自 2023 年起，KS2 學校進步指數計算的起點一律採入學基準線評量結果，在中間這段過渡期，KS2 進步指數的計算用 KS1 的評量結果為起點 (DfE, 2014a)。

2. 第一關鍵期 (KS1) 評量

(1) 全國測驗

測驗科別包括數學、英文的閱讀，及英文的文法、標點與拼字。試題由標準和測驗局 (Standards and Testing Agency [STA]) 研發，在每年的 5 月舉行，施測時間由學校自行安排，結果用量尺分數呈現，以 100 分為期望標準 (expected standard)。教育部在 6 月上旬提供原始分數與量尺分數對照表給學校，協助教師判斷學生是否達到期

望標準 (STA, 2016a)。之前提過，KS1 學生國定課程評量結果以教師評量為據。

(2) 教師評量

科別包括數學、科學、英文的閱讀和寫作。由於新課程不再使用成就層級，STA 在 2015 年 9 月公布教師評量架構，提供表現描述 (performance descriptor) (STA, 2015a) 協助教師做判斷，不過此標準只適用 2016 年，未來將視情況修正。數學、閱讀與寫作皆訂定三個標準：邁向期望標準 (working towards expected standard)、達到期望標準 (working at expected standard)，以及達到期望且更深化的標準 (working at greater depth within the expected standard)，學生表現分為四類，除上述三類外，尚有「低於邁向預期標準」一類。科學只訂「達到期望標準」一標，學生表現分為達到標準和未達標準兩類。教師評量結果規定在 6 月底送到地方當局，地方當局確認資料品質後，在 7 月底前按規定格式送到教育部 (STA, 2016a)。

3. 第二關鍵期 (KS2) 評量

(1) 全國測驗

測驗科別與 KS1 相同，但施測時間全國統一，例如 2016 年測驗時間安排在 5 月 9 日至 12 日，每天考一個科別，由外部機構計分；科學則採抽樣施測，每兩年舉辦一次。測驗結果皆以量尺分數報告，以 100 分為期望標準 (STA, 2016b)。

(2) 教師評量

科別包括數學、科學，及英文的閱讀和寫作。根據暫定的教師評量架構 (STA, 2015b) 的表現描述，閱讀、數學和科學都只訂「達到期望標準」一標；寫作則訂定三個標準：邁向期望標準、達到期望標準、達到期望且更深化的標準。教師評量結果在 6 月底前上傳的國定課程評量資料庫網站，寫作評量結果在上傳前，須要經過外部校正 (external moderation) (STA, 2016b)。

(六) 小學學校績效指標的改革

配合評量改革，KS2 學校表現榜資訊與樓地板標準說明如下。

1. 學校表現榜資訊

學校表現榜公佈四項資訊：(1)閱讀、數學和寫作三科皆達到期望標準的學生百分比；(2)閱讀、數學和寫作三科皆達到高標準 (higher standard) 的百分比；(3)學生在全國測驗 (閱讀和數學) 的平均量尺分數；(4)學生在閱讀、數學和寫作的平均進步分數。

上述資訊的說明如後，閱讀和數學的期望標準指全國測驗之量尺分數達 100 分以上，寫作指教師評量結果達期望標準以上；閱讀和數學的高標準視 2016 年考後實際考生分數分佈才界定，寫作則指教師評量結果在達到期望且深化的標準。進步分數計算採增值分數 (value-added score) 取向，2016 年學校進步分數計算方法，扼要說明如後 (DfE, 2017)。以學生在 KS1 (2012 年) 閱讀、寫作和數學之教師評量結果為基準，計算學生各科的進步分數，個別學生的進步分數不會告知學生和家長，只是用來計算學校的進步指數，學校學生進步分數的平均數就是學校的進步指數。全國進步分數的期望值 (平均數) 為 0，因此，學校進步指數若為 0，表示與 KS1 成就相似者相比，該校學生在 KS2 的進步是合乎期望的，若指數為正值，表示進步情形優於全國，若進步指數為負值，不表示學生沒有進步，只是進步情形不如全國 (DfE, 2017)。

2. 樓地板標準

學校可從成就指標 (attainment measure) 或進步指標 (progress measure) 兩者當中，選擇一種。

成就指標指學生在數學、閱讀和寫作三科皆達到期望標準的百分比，雖然 2014 年三月公布擬以 85% 為標準 (DfE, 2014a)，但 2016 年 1 月公布 2016 年以 65% 為標準 (DfE, 2017)，可見改革不是一次到位。

進步指標係建構學校各科進步指數的 95% 信賴區間，如果進步指數為負值，但區間仍包含 0，表示學校的進步仍然符合期望，若區間上限低於 0 分，表示未達充分進步 (sufficient progress)。學校必須三科 (數學、閱讀和作文) 的進步指標都達到充分進步，才算通過樓地板標準 (DfE, 2017)。

(七) 中學學校績效指標的改革

為減輕學校的負擔，自 2014 年起，終止 KS3 法定教師評量，也就是說，KS3 當前已無法定的國定課程評量。自 2016 年起，中學學校績效指標以 KS4 學生在八個科目證書考試的表現為據，所謂八個科目係由三群科目組成，包括：(1) 兩科核心科目：英文和數學；(2) 三科英格蘭文憑 (EBacc) 科目，若學生修課超過三個科目，此群採計成績最好的三科；(3) 三科其他科目，包括第二群未採計的 EBacc 科目，及教育部認可的藝術、學術或職業證書，採計成績最好的三科 (DfE, 2013c)。學校表現榜資訊與樓地板標準說明如下。

1. 學校表現榜資訊

學校表現榜公布下列四項資訊：(1)八科目的成就指標 (Attainment 8 measure)；(2)八個科目的進步指標 (Progress 8 measure)；(3)英文和數學兩科皆得到 C 等級以上的學生百分比；(4)各英格蘭文憑科目得到 C 等級以上的學生百分比。

八科目成就指標的計算方式說明如後 (DfE, 2016)，先將學生的證書等級轉換成分數，例如：GCSE 的 A* 轉為 8 分，A 為 7 分，以此類推，G 為 1 分。計算個別學生八科目證書的加權平均數，學校的成就指標就是學生八科目的加權平均數的平均數。值得一提的是，從 2017 年起，GCSE 證書成績改成 1-9 分制，不再使用 A* 到 G 級，如此，計算成就指標時，無須作等級與分數的轉換，直接用分數計算即可，屆時將使用新的計算方式 (DfE, 2016)。八科目進步指標的計算原理和前述小學的進步指標相同，以學生當年 KS2 數學和英文兩科測驗結果為基準，學校的進步指數即是該校學生進步分數的平均數。全國進步分數的期望值為 0，若學校的進步指數為正值，表示該校學生進步情形優於全國，如為負值，則表示進步不如全國。如果接近 0，表示該校的學生進步合乎期望 (DfE, 2016)。

2. 樓地板標準

KS4 的樓地板標準只採用進步指標。建構各校的進步指數 95% 的信賴區間，如果學校的進步指數低於 -0.5，而且區間的上限低於 0，就未達樓地板標準 (DfE, 2016)。

三、小結

聯合政府上台執政後，立即對國定課程與評量進行多項改革，最重大的改革之一，是國定課程不再用層級來架構 KS1 到 KS3 的成就目標，連帶影響到評量結果報告的方式。學生關鍵期內的學習成就報告與進步監控方式，交由學校自行發展，不再全國統一規定，小學階段的法定教師評量另訂表現標準和期望標準，全國測驗以量尺分數報告。此外，KS4 的 GCSE 成績自 2017 年起也將由等級制改成分數制，這些改革的目的是使學校績效指標更容易進行量化管理。

聯合政府與當前的保守黨政府依舊延續測驗本位的學校績效責任，實施樓地板標準政策，小學可從成就指標與進步指標兩者之中選擇一種，中學則單單採用進步指標。採用進步指標對弱勢學校較為公平，舉例來說，有些學校達到期望標準學生的比率雖不理想，但學生們卻有顯著的進步，此種情況下，成就指標無法肯定學校的努力，進步指標則能激勵學校關注到每位學生的進步和發展。學校課責雖然重視進步指標，但全國測驗的性質仍然屬於高風險測驗，是否因此降低全國測驗對學校課程與教學負

面的回浪效應，仍有待觀察。

伍、我國學生學習成就評量實施現況

國內雖然沒有法定課程評量制度存在，但配合十二年國教實施，教育部推出一些評量配套與措施，以下概述國內當前較重大評量措施。

一、臺灣學生學習成就評量資料庫

教育部（2011）在《十二年國民基本教育實施計畫》中將「臺灣學生學習成就評量資料庫」（Taiwan Assessment of Student Achievement [TASA]）列為提升學生學習成就的配套措施。TASA 建置始於 2004 年，國家教育研究院（當時仍為籌備處）承接教育部的委託，針對國民中小學學生學習成就建立常態性之資料庫，以做為教育部研訂課程與教學政策之重要參據。TASA 施測對象為國小四年級、國小六年級、國中二年級、高中二年級、及高職二年級的學生，評量科目為國語（文）、英語（文）、數學、自然和社會五個考科，採抽樣施測，受試學生接受評量題本施測外，亦填寫背景問卷，被抽為樣本的學校，校長或學校代表也填寫學校問卷，以利更了解與學生學習成就有關的因素（國家教育研究院，2017a）。

在 2005 年，由小六國語、英語和數學三科率先登場，2006 年與 2007 年五種年段的五個科目（惟小四社會科例外）全面施測，2008 年之後，國小、國中、高中（職）的施測週期改為三年一次，小四考科也改成只施測國語、數學和自然三科（國家教育研究院，2017a）。TASA 結果的報告方式有量尺分數（平均數定為 500，標準差定為 50）與成就水準（分未達基礎、基礎、精熟、進階四級）。筆者從 TASA 的一些報告，早期的如 TASA 2005（陳清溪、林淑敏，2005）、TASA 2006（曾建銘、陳清溪，2008）、TASA 2007（曾建銘、陳清溪，2009），與近期報告如 TASA 2013（國家教育研究院，2014），發現研究報告內容以學生在課程綱要之內容層面、能力指標、評量細目的平均答對率為主，建議的重點放在教學的意涵上，雖偶見部分科目年度的學生整體表現統計，但未見歷年來檢測結果的趨勢分析。換言之，TASA 臚列的第一項目的：「建立國民中小學、高中及高職學生學習成就長期資料庫，以追蹤、分析學生在學習上變遷之趨勢，進而檢視目前課程與教學實施成效。」並未落實。

此外，教育部在 2009 年委請 TASA 團隊協助縣市辦理學生學習能力檢測計畫，由縣市自行提出申請，採用普測方式。2010 至 2012 年檢測小四的國語和數學兩科，自 2013 年起，改成小五的國語和數學兩科。為了及早發現待加強學生，盡早啟動積極性教學介入，自 2016 年起，檢測對象向下延伸至小二和小三，小五增加英語科的施測。此外，為回應部分縣市對於國中學生學習能力的檢測需求，2016 年小規模試辦國中七、八年級國文、數學及英語三科，未來將依實施成效，評估擴大辦理可能性（國家教育研究院，2017b）。縣市檢測分析結果報告的主要內容為試題解析與其對教學輔導的建議。

二、國中教育會考

國中教育會考（以下簡稱會考）於 103 年登場，會考的法源依據是〈國民小學及國民中學學生成績評量準則〉第 13 條，「為瞭解並確保國民中學學生學力品質，應由教育部會同直轄市、縣（市）政府辦理國中教育會考」（教育部，2012a）。會考五個考科將成績區分「精熟」（A）、「基礎」（B）和「待加強」（C）三個等級，就瞭解國中學生學力來說，成績區分「精熟」、「基礎」和「待加強」三個等級，已可滿足此需求。然而目前會考成績的主要用途是，做為免試入學比序的採計項目之一，這種情況下，三等級的區別力稍嫌不足，於是 A 和 B 級各再增加標示（++/+），以補不足。各免試就學區通常把各考科的等級與標示再轉換成分數後加總，這一來，又違背會考成績採用等級制的初衷，就是要避免對分數錙銖必較的現象（教育部，2012b）。

此外，根據教育部（2012b），會考的目的有四項：(1)降低考試壓力，活化學生學習；(2)檢視學生學力，確保學習品質；(3)回饋學習成果，強化適性輔導；(4)提供學力資訊，俾利因材施教。由於會考成績作為免試入學同分比序用，會考仍帶有高風險測驗的性質，學生的考試壓力並未獲得紓解。再者，用會考成績作為國中端的教學回饋，確保學習品質，顯然為時已晚。若要作為高中（職）篩選補救教學對象，以強化適性輔導與俾利因材施教，會考成績提供的訊息，實質助益極其有限，因為會考的等級描述係因應考試而生，非為課程與教學設計。總之，會考的目的與定位，實有再釐清與確立的必要。

三、國民中小學生學習成就標準本位評量

「國民中學學生學習成就評量標準」是《十二年國民基本教育實施計畫》之「5-1 落實國中教學正常化、適性輔導及品質提升」工作要項下的配套之一，教育部在 100 年 8 月委託國立臺灣師範大學心理與教育測驗研究發展中心研發、推廣和試辦之。此配套的主要目的在建置與國民中小學九年一貫課程綱要能力指標相對應的評量標準，以補足目前課綱各學習領域雖訂有能力指標，卻未提供評量檢核指標之不足，並作為全國教師在進行教學評量時的統一參照依據。其臚列的具體目的包括：(1)提供與課綱相對應的評量參照依據，轉型為標準參照的評量；(2)即時提供教學回饋；(3)促進教師專業發展；(4)建立學力監控的機制；(5)評量概念與國際接軌。評量的表現標準分為 A~E 五級，A 表「優秀」、B 表「良好」、C 表「基礎」、D 表「不足」、E 表「落後」，試辦領域和科目計七大領域共 16 科（國立臺灣師範大學心理與教育測驗研究發展中心，2015）。在 104 年 9 月，教育部又增加委託國小階段評量標準研發、推廣和試辦，評量標準的介紹與說明，以及國中的試辦成果皆公布在「國民中小學生學習成就標準本位評量」官網（教育部，2017b）。持平而論，國中小學生學習成就評量標準的建置與試辦計畫頗有英國教師評量的理路。

四、國民小學及國民中學補救教學科技化評量

補救教學是十二年國教的重要措施，教育部委託財團法人技職專校入學測驗中心負責建置與執行「國民小學與國民中學補救教學科技化評量」，題庫命題係根據教育部頒定的基本學習內容，自 103 年起，試題委請國家教育研究院協助研發。實施方式為，在 105 年之前，學校在每年 9~10 月初步提報規定比率的學生（約 25%~35%，提報率年有所改變）參加國語（文）、數學和英語的篩選測驗，對象為 2 到 9 年級，其中 2 和 3 年級學生採紙筆測驗，4 到 9 年級學生採用電腦線上測驗，不及格者（未達 60 分）將列為受輔對象，並於隔年的 2~3 月與 6~7 月實施第一次與第二次的學習成長測驗，了解其進步情形，並計算各校進步率。自 105 年起，篩選測驗改在每年 5 月實施（第二學期末），施測對象為 1~8 年級學生，成長測驗於 12~1 月（第一學期末）實施，106 年起篩選和成長的通過標準訂為：四年級以下答對率為 80%、五、六年級為答對率 72%、七年級以上為 60%。相關的實施資訊或政策可於「國民小學及國民中學補救教學科技化評量」網站檢索（教育部，2017a）。

五、小結

檢析國內當前的評量方案與措施，發現外部評量（如 TASA 和會考）與內部評量（如評量標準和補救教學評量）均有，只是獨立發展，不免有疊床架屋之嫌，評量結果也難以整合運用，舉例來說，國中教育會考成績分三級、TASA 分四級、國民中小學生學習成就評量標準分五級，都有「基礎」一級，但其意義是否相通，再者，補救教學科技評量的通過標準是否與所謂「基礎」層級相容，皆令人存疑。

陸、反思與借鏡

英國國定課程評量是一法制化的評量，特別將教師評量也納入法定評量的一環。國內雖然沒有法定課程評量制度存在，但綜觀十二年國教重大的評量配套與措施，涵蓋面向已算充分，只是缺乏共同的詮釋架構與有效的整合。英國國定課程評量實施已超過二十餘載，在變革之中仍有延續一貫之處，筆者從中擷取可資借鏡與反思之處，並以國內既有的努力成果為基礎，對十二國教課程評量之規劃與實施，提出建議。

一、英國國定課程評量的借鏡

（一）法制化的課程評量制度

英國教育大臣 Keeneth Baker 在 1987 年指出，單有課程改革仍不足以提升學校水準，還需要國家評量定期測量學生的進步，並提供客觀評量資訊給學生、家長和教師（House of Commons: Children, Schools and Families Committee, 2008）。1988 年教育改革法案通過國定課程與評量的實施，國定課程評量採外部全國評量和內部教師評量雙軌並行。在 1994 年首次課程修訂後，確立法定評量以核心科目（英文、數學與科學）為主。國定課程評量一路走來，雖然批評與爭議不斷，評量安排也經數次大小調整，但充當國家教育水準趨勢的監控機制，卻是始終如一。

（二）明訂學生的期望表現與重視學生學習進展的監控

國定課程在以成就層級與層級描述為目標的時期，針對各關鍵期訂定期望層級（見表 1），成就層級也構成學生學進展的評估與監控架構，全國測驗和教師量結果

都以成就層級報告。具體來說，政府藉由各校、各地區與全國之各年度，全國測驗達到期望層級以上的學生百分比等統計資料，了解並監管學校、地區和全國的教育水準的發展趨勢。另一方面，在關鍵期間，藉由教師評量來評估與監控學生核心科目的學習進展情形，以能對需要幫助的學生提供適時的介入，協助其在關鍵期評量達到期望層級，政府提供一些輔助與措施（如 OP 測驗與 APP 套件）協助教師判斷。自 2016 年起，配合課程改革實施的新制評量，全國測驗與教師評量結果的報告方式不再一致，但各自依然訂有期望的標準；在各個關鍵期間，學生學習進步的監控仍是教師評量的重點，只是把監控的架構與方式，交由學校規畫，不再全國統一規定。

（三）測驗本位的學校課責政策

英國將全國測驗做為評估學校績效責任與提升全國教育標準的主要機制，以至於全國測驗成為高風險測驗，這種測驗本位的課責（test-based accountability）政策引發許多的爭議和討論。全國測驗為學校與師生帶來壓力，為考試而教學是相當普遍的現象，特別是接受評量的年級，師生花不成比例的時間準備考試，反覆練習考試材料，非測驗科目被忽略或排擠，課程學習內容被窄化。甚至，國會調查報告書都對此提出改革建議，對全國測驗加以改革，降低其多重目的性，讓學校從不計代價追求測驗結果的窘境中脫困；對於學校表現的監控，則改成抽樣調查（House of Commons: Children, Schools and Families Committee, 2008）。

全國教育水準提升了嗎？Torrance（2011）整理歷年 KS2 全國達到期望層級以上學生的百分比，筆者在此僅舉出 1996、2000、2005 與 2010 四個年度為例，英文科分別是 58%、75%、79%與 81%，數學為 54%、72%、75%與 80%，科學為 62%、85%、86%、與 85%，整體看來，大幅提升的情況只出現 2000 年之前，2000 年之後的成長趨緩，甚至平滯。之前提過，政府為突破發展平滯現象，曾提出大躍進（MGP）試驗方案（DfES, 2007），不過試驗期滿後，皆未再延續。顯然，以外部高風險測驗來鞭策學校提高表現並非有效之途，而且還引發一些負面的回滾效應。

（四）運用教師評量來增進、監控與報告學生學習

TGAT 將國定課程評量設計成兼顧多重目的的系統，採全國測驗與教師評量雙軌並行，但建議以形成性目的與教師評量為優先（DES/WO, 1988）。然而國定課程評量正式登場後，即偏離 TGAT 的設計，學校表現榜獨尊全國測驗，民間學術團體（ARG）因而發起並推廣促進學習的評量。教師評量涵蓋面包括日常的、定期的與關鍵期的法

定評量，舉例來說，在日常的課堂評量之外，在關鍵期間，監控學生往期望層級（標準）的學習進展，在關鍵期末的法定評量，教師就學生綜合表現判斷其達到的成就層級或標準為何，評量結果呈送到教育部。雖然政府與學術團體對教師評量強調的面向不同，前者重視法定評量結果的品質，後者重視與教學密切結合並能即時回饋學習的課堂評量，但對協助教師提升相關的專業能力，卻是有志一同的。英國將教師評量制度化，賦予教師評量增進、監控與報告學生學習的角色，有層次與系統的運用教師評量，值得我們參考。

二、對十二年國教課程評量規劃與實施的建議

之前提過，國內當前重要的評量方案與措施已涵蓋外部評量與內部評量，但缺乏共同的評定與詮釋架構，評量結果難以整合運用。以下以國內既有的努力成果為基礎，提出十二年國教課程評量規劃與實施的建議。

（一）根據課程綱要訂定各學習階段的期望表現或標準

國內現有的評量方案所定之標準層級或水準的數目不一，彼此意義也難以相通，若要整合與運用各評量結果，需要有一共同的詮釋架構，因此，可借鏡英國的國定課程與評量，制定關鍵期的表現標準並訂定期望標準（至少針對核心科目）。之前提過，國內國民中小學生學習成就評量標準的建置雖有類似的意圖，然而其並不具任何法制地位。建議教育部針對十二年國教各領域或科目之課程綱要各學習階段的學習重點，制定官方的表現標準並訂定期望標準，為不同的評量方案奠立共同的詮釋架構，進而建置監控學生學力品質的評量網絡。建議先檢視與檢討國中學生學習成就評量標準歷年來的試辦成果，及遇到的問題與困難，做為經驗傳承。此外，持平而論，此一工程龐大複雜，不妨排定優先學習領域與學習階段，循序漸進的研發建立。

（二）強化常態性 TASA 的學力監控功能

TASA 施測對象從小學到高中（職）共 5 種年段，施測科目有 5 個考科，採用抽樣施測，屬於低風險的評量，作為全國或地區學力品質監控機制來說，TASA 的實施設計算是周全。比較可惜的是，已公開之 TASA 的報告，分析重點與建議著重在對教學的回饋，然因其採取抽測，並沒有明確特定的回饋對象，成效如何呢？有待商榷。值得一提的，TASA 臚列的目的共有六項（國家教育研究院，2017a），持平而論，TASA 提供的評量資訊無法確實顧及所有目的，因此，建議常態性 TASA 以全國學力監控為

優先，呼籲教育部定期對外公布 TASA 施測學生整體學習表現，以及歷來趨勢變化的統計數據。我國雖沒有如英國般的法定課程評量制度，但 TASA 的建置實有與之相似的意涵。

（三）整合 TASA 協助縣市辦理學力檢測計畫與補救教學科技評量

TASA 協助縣市辦理學力檢測擴及的年級越來越多，補救教學科技評量施測的時程自 105 年也調整，換言之，兩個方案重疊的部分越來越多，而且兩者目的皆包括篩選出需要接受補救教學的學生。因此，建議將兩個方案做整合並加以擴展，建構 1～9 年級的國語（文）、數學、英語的題庫平台，可仿效英國的自選測驗（OP），由學校自行施測，但官方提供一些輔助工具，如計分指引、成就表現切點資料與登記的程式等等，若學生未達期望標準，則列為補救的對象。之前提過，若課程綱要有明定的表現標準與期望標準，那麼這個平台可以充當輔助教師與學校評估學生成就及監控其學習進展的機制，包括接受補救教學學生的進步情形。

（四）重新思考國中教育會考的目的與定位

國中教育會考係為高中職免試入學政策實施後，確保國中學生學力品質而設立。筆者之前已對會考的性質和目的做了一番評述，在此，再對其用途做一些討論。會考定位為標準參照測驗，將學生成績區分為三級，然而在會考成績的使用上，各免試就學區皆將各科的等級（含標示）轉成分數後加成總分後，再做比序，此種情況下，會考不折不扣是一個常模參照測驗，常模參照測驗的目的通常將受試者程度做最大的鑑別，顯然目前會考的定位和成績的使用有自相矛盾之處。實施以來，國中生的壓力未減，反而爭議不斷，一點都不意外。會考實施已有四年，這四年來，學生各科在精熟、基本和待加強的分布百分比，其實變化不大，而且數學、英語被評為待加強等級者皆有 3 成以上，足見藉由外部考試改革來提高學生學習成就之路是不通的。當前的會考仍具高風險測驗的性質，因此，會考在十二年國教中扮演的角色與功能，有必要再釐清與再確定。

（五）規劃與推動全方位的教師評量

外部評量通常在特定時間點實施，故有「拍快照」評量（snapshot assessment）之稱，內部評量可對學生的學習成果作多面向且較長期、深入的觀察與評量，且諸多研究證據顯示，優質的課室評量才是提高學生學習與成就的王道（Black & William, 1998a,

1998b)。值得一提的是，十二年國民基本教育課程綱要以「核心素養」為課程發展的主軸，新課綱能否落實，「素養導向的教學與評量」為其關鍵（高鴻怡，2017）。《十二年國民基本教育課程發展建議書》（潘文忠主編，2014）指出，核心素養是指知識、能力和態度的綜合表現；核心素養之培養，有賴課程、教學與評量的有效連結。就素養評量的實踐而言，外部評量仍然有力未及之處，教師評量才是最可行的管道。持平而論，國內並不缺教師評量的配套，例如推動多年的多元評量、當前的國民中小學生學習成就評量標準配套措施，只是缺乏全面性或系統性的政策。教育部可參考英國的教師評量制度與實施，研訂全方位的教師評量政策與配套，並發展教師相關的專業知能，有層次的與系統的運用教師評量，來提高學生學習，及監控與報告其成就與進步情形。

柒、結語

英國在 1988 年通過教育改革法案，實施國定課程與評量，本文梗概的回顧與闡析國定課程評量從規劃到當前的重大發展與變革及其脈絡與影響。國定課程與評量雖漸次鬆綁，但核心科目根據課程綱要定訂成就層級（或表現標準），並定期望層級或標準，讓法定評量結果的詮釋與應用，有共同的架構，卻是延續一貫的。此外，國定課程評量也一貫維持全國測驗與教師評量雙軌並行，全國測驗充當學校課責與管控全國教育水準發展的機制；教師評量的功能則定位在提高學生學習，監控與評估學生關鍵期間的學習進展，以幫助其關鍵期末達到期望層級或標準。

從英國國定課程評量的實施，可以觀察到不同評量利害關係人（如學生、教師、家長、政治人物等）對評量目的與資料運用的期待，不盡然相同，甚至彼此之間有張力衝突存在。TGAT 的主席 Black（1994）在國家評量設計經驗談的結尾，拋出這個問題：「我們如何使形成性評量與授證績效考試並行，且彼此相互支援、增添豐富，而不扭曲優質的課堂教學呢？」二十餘載已過去，這仍然是個待解的問題，足見要運作一個滿足所有厲害關係人的期待與需求的評量系統，是一件極不容易的任務。國內雖沒有實施學校績效責任的測驗，然而學校也深受高風險測驗（如升學考試）的衝擊，這也是歷次教改的焦點之一。如果所有評量利害關係人都回到學校教育的初心，學校的課程、教學與評量都回到支持學生學習與進步的角色，而不是在服務外部測驗或政

策，問題是否就變得簡單些？國內尚未有法定課程評量制度，本文檢視當前十二年國教重大的評量配套與措施，包括：臺灣學生學習成就評量資料庫、國中教育會考、國民中小學生學習成就標準本位評量、國民小學與國民中學補救教學科技化評量等四種方案，發現兼具外部與內部評量，涵蓋面已算充分，只是缺乏共同的詮釋架構與有效的整合運用。若要如英國國定課程評量可以做系統的整合與運用，並建立起評估與監控學生學習進展的架構，根據新的課程綱要建立官方的表現標準，應該是建置十二年國教課程評量系統的優先要務。此外，為符應新課程的特色，落實素養導向的教學與評量，制定全方位教師評量政策與發展教師相關專業知能，也應是當前急務之一。

誌 謝

本文的完成，感謝科技部計畫（NSC 100-2410-H-003 -074）的經費補助，也感謝審查委員與期刊編輯委員的寶貴回饋與建議。

參考文獻

- 吳麗君（2000）。交互損益見真章—談英國國家課程評量的來時路。*課程與教學季刊*，**3**（1），79-94。
- [Wu, L. -C. (2000). Trade-offs: The historical development of national curriculum assessment in England. *Curriculum & Instruction Quarterly*, 3(1), 79-94.]
- 吳麗君（2001）。蘇格蘭與英格蘭小學階段之國家考試的比較。*國立臺北師範學院學報*，**14**，35-56。
- [Wu, L. -C. (2001). The comparison of national examination between England and Scotland. *Journal of National Taipei Teachers College*, 14, 35-56.]
- 高鴻怡（2017）。素養導向教學與評量的轉化、簡化及流通。載於蔡清華（主編），*課程協作與實踐*（頁 31-41）。取自
<http://ws.moe.edu.tw/001/Upload/23/refile/8006/54442/650a914c-867a-493c-96a1-4c63fa51e7.pdf>

[Gao, H. -Y. (2017). The transformation, simplification and circulation of competence-based teaching and assessment. In C. -H. Tasi (Ed.), *Curriculum collaboration and implementation* (pp. 31-41). Retrieved from <http://ws.moe.edu.tw/001/Upload/23/refile/8006/54442/650a914c-867a-493c-96a1-4c63fa51e7.pdf>]

教育部 (2011)。十二年國民基本教育實施計畫核定本。台北：作者。

[Ministry of Education. (2011). *Masterplan for 12-year basic education*. Taipei: Author.]

教育部 (2012a)。國民小學及國民中學學生成績評量準則修正條文。台北：作者。

[Ministry of Education. (2012a). *Amendment of guidelines for assessment and grading for elementary and junior high school students*. Taipei: Author.]

教育部 (2012b)。十二年國民基本教育落實國中教學正常化、適性輔導及品質提升方案－國中教育會考。取自 <http://cap.ntnu.edu.tw/documents/101049.pdf>

[Ministry of Education. (2012b). *The project for realizing normalized teaching, adaptive guidance, and quality enhancement in 12-year basic education: Comprehensive assessment program for junior high school students*. Retrieved from <http://cap.ntnu.edu.tw/documents/101049.pdf>]

教育部 (2017a)。國民小學及國民中學補救教學科技化評量。取自 https://exam.tcte.edu.tw/tbt_html/index.php

Ministry of Education (2017a). *Project for implementation of remedial instruction-technology-based testing programs*. Retrieved from https://exam.tcte.edu.tw/tbt_html/index.php

教育部 (2017b)。國民中小學生學習成就標準本位評量。取自 <https://www.sbasa.ntnu.edu.tw/SBASA/HomePage/index.aspx>

[Ministry of Education. (2017b). *Standards-based assessment of student achievement for elementary and junior high school students*. Retrieved from <https://www.sbasa.ntnu.edu.tw/SBASA/HomePage/index.aspx>

國立臺灣師範大學心理與教育測驗研究發展中心 (2015)。國民中學學生學習成就評量標準 (試行版)：研發說明與常見問題集。台北：作者。

[National Taiwan Normal University, Research Center for Psychological and Education Testing. (2015). *Standards-based assessment of student achievement handbook for junior high school*. Taipei: Author.]

- 國家教育研究院 (2014)。2013 年臺灣學生學習成就評量資料庫國中二年級學生學習成就評量結果報告。取自
http://web.lyhs.ptc.edu.tw/tw/upload/01_academic_affair/teaching_affair/2013_taiwan_8th_grade_students_Learning_report.pdf
- [National Academy for Educational Research. (2014). *The report for grade 8 students on 2013 assessment of Taiwan student achievement*. Retrieved from
http://web.lyhs.ptc.edu.tw/tw/upload/01_academic_affair/teaching_affair/2013_taiwan_8th_grade_students_Learning_report.pdf]
- 國家教育研究院 (2017a)。 「臺灣學生學習成就評量資料庫」建置計畫。取自
<http://www.naer.edu.tw/files/11-1000-1408.php?Lang=zh-tw>
- [National Academy for Educational Research. (2017a). *Establishment and management of Taiwan assessment of student achievement (TASA)*. Retrieved from
<http://www.naer.edu.tw/files/11-1000-1408.php?Lang=zh-tw>]
- 國家教育研究院 (2017b)。縣市學力檢測公布系統。取自 <http://rap.naer.edu.tw/>
- [National Academy for Educational Research. (2017b). *The implementation of regional Taiwan assessment of student achievement*. Retrieved from <http://rap.naer.edu.tw/>]
- 陳明印 (2004)。英國國定課程學生成就評量制度。研習資訊, 21 (6), 90-109。
[Chen, M. -Y. (2004). National curriculum assessment system in England. *Inservice Education Bulletin*, 21(6), 90-109.]
- 陳清溪、林淑敏 (2005)。2005 年台灣學生學習成就評量簡介。教育研究與發展期刊, 1 (3), 241-254。
[Chen, C. -H., & Lin, S. -M. (2005). Introduction to 2005 assessment of Taiwan student achievement. *Journal of Educational Research and Development*, 1(3), 241-254.]
- 黃光雄、周淑卿 (1992)。英國國定課程評析。國立臺灣師範大學教育研究所集刊, 34, 181-201。
[Huang, K. -H., & Chou, S. -C. (1992). The analysis and comment for the national curriculum of Britain. *Bulletin of Graduate Institute of Education Taiwan Normal University*, 34, 181-201.]
- 曾建銘、陳清溪 (2008)。2006 年臺灣學生學習成就評量結果之分析。教育研究與發展期刊, 4 (4), 41-86。
[Cheng, C. -M., & Chen, C. -H. (2008). The analysis of Taiwan assessment of student achievement 2006. *Journal of Educational Research and Development*, 4(4), 41-86.]

- 曾建銘、陳清溪 (2009)。2007 年臺灣學生學習成就評量結果之分析。 *教育研究與發展期刊*，5 (4)，1-38。
- [Cheng, C. -M, & Chen, C. -H. (2009). The analysis of Taiwan assessment of student achievement 2007. *Journal of Educational Research and Development*, 5(4), 1-38.]
- 潘文忠 (2014) (主編)。十二年國民基本教育課程發展建議書。取自
<http://12cur.naer.edu.tw/upload/files/133525b450597ebad78b62473aa59941.pdf>
- [Pan, W. -C. (2014) (Ed.). *A handbook for developing the 12-year Basic Education Curriculum*. Retrieved from
<http://12cur.naer.edu.tw/upload/files/133525b450597ebad78b62473aa59941.pdf>]
- Assessment Reform Group [ARG]. (2002). *Assessment for Learning: 10 principles*. Retrieved from
<https://www.aiaa.org.uk/content/uploads/2010/06/Assessment-for-Learning-10-principles.pdf>
- Black, P. J. (1994). Performance assessment and accountability: The experience in England and Wales. *Educational Evaluation and Policy Analysis*, 16(2), 191-203.
- Black, P. J. (1997). Whatever happened to TGAT?. In C. Cullingford (Ed.), *Assessment versus evaluation* (pp. 24-50). London, UK: Cassell.
- Black, P. J. (1998). Learning, league tables and national curriculum: Opportunity lost or hope deferred?. *Oxford Review of Education*, 24(1), 57-68.
- Black, P. J., & William, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P. J., & William, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. London, UK: King's College School of Education.
- Commission on Assessment without Levels. (2015). *Final report of the Commission on Assessment without Levels*. Retrieved from
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/483058/Commission_on_Assessment_Without_Levels_-_report.pdf
- Daugherty, R. (1995). *National curriculum assessment: A review of policy 1987-1994*. London, UK: Falmer Press.
- Daugherty, R. (1996). In search of teacher assessment - its place in the national curriculum assessment system of England and Wales. *The Curriculum Journal*, 7(2), 137-152.

- Daugherty, R. (1997). National curriculum assessment: The experience of England and Wales. *Educational Administration Quarterly*, 33(2), 198-218.
- Daugherty, R. (2007). Mediating academic research: The Assessment Reform Group experience. *Research Papers in Education*, 22(2), 139-153.
- Daugherty, R. (2009). National curriculum assessment in Wales: Adaptations and divergence. *Educational Research*, 51(2), 247- 250.
- Dearing, R. (1993). *The national curriculum and its assessment: Final report*. London, UK: School Examinations and Assessment Council.
- Department for Children, Schools and Families [DCSF]. (2008). *The assessment for learning strategy*. London, UK: Author.
- Department for Children, Schools and Families [DCSF]. (2009). *Report of the expert group on assessment*. Retrieved from <http://dera.ioe.ac.uk/10585/1/Expert-Group-REPORT.PDF>
- Department for Children, Schools and Families [DCSF]. (2010). *Assessing pupil's progression: A teachers' handbook*. London, UK: Author.
- Department for Education [DfE]. (2010). *The importance of teaching: The schools white paper 2010*. Retrieved from <http://www.educationengland.org.uk/documents/pdfs/2010-white-paper-teaching.pdf>
- Department for Education [DfE]. (2011). *Independent review of key stage 2 testing, assessment and accountability*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/176180/Review-KS2-Testing_final-report.pdf
- Department for Education [DfE]. (2013a). *Reform of the national curriculum in England*. Retrieved from <https://ico.org.uk/media/about-the-ico/consultation-responses/2013/2134/consultation-document-reform-of-the-national-curriculum-20130207.pdf>
- Department for Education [DfE]. (2013b). *Phonics screening check for Year 1 pupils*. Retrieved from <https://www.education.gov.uk/schools/teachingandlearning/pedagogy/a00197709/phonics-screening-year-1>

- Department for Education [DfE]. (2013c). *Reforming the accountability system for secondary schools: Government response to the February to May 2013 consultation on secondary school accountability*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/249893/Consultation_response_Secondary_School_Accountability_Consultation_14-Oct-13_v3.pdf
- Department for Education [DfE]. (2014a). *Reforming assessment and accountability for primary schools: Government response to consultation on primary school assessment and accountability*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/297595/Primary_Accountability_and_Assessment_Consultation_Response.pdf
- Department for Education [DfE]. (2014b). *The national curriculum in England: Framework document*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/381344/Master_final_national_curriculum_28_Nov.pdf
- Department for Education [DfE]. (2015). *Consultation on implementing the English Baccalaureate*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/473455/Consultation_on_implementing_the_English_Baccalaureate.pdf
- Department for Education [DfE]. (2016). *Progress 8 measure in 2016, 2017, and 2018: Guide for maintained secondary schools, academies and free schools*. Retrieved from http://dera.ioe.ac.uk/27619/1/Progress_8_school_performance_measure_13_Oct_update.pdf
- Department for Education [DfE]. (2017). *Primary school accountability in 2016: A technical guide for primary maintained schools, academies and free schools*. Retrieved from http://dera.ioe.ac.uk/28118/1/primary_school_accountability_technical_guidance_January_17.pdf
- Department for Education and Employment [DfEE]. (1997). *Excellence in schools*. London, UK: Author.
- Department for Education and Skills [DfES]. (2002). *14-19: Extending opportunities, raising standards*. London, UK: Author.

- Department for Education and Skills [DfES]. (2003). *Excellence and enjoyment: A strategy for primary schools*. London, UK: Author.
- Department for Education and Skills [DfES]. (2007). *Making good progress: How can we help every pupil to make good progress at school?*. Nottingham, UK: Author.
- Department of Education and Science and Welsh Office [DES/WO]. (1987). *The national curriculum 5-16: A consultation document*. London, UK: Author.
- Department of Education and Science and Welsh Office [DES/WO]. (1988). *National curriculum task group on assessment and testing: A report*. London, UK: Author.
- House of Commons: Children, Schools and Families Committee. (2008). *Testing and assessment. Third report of session 2007-08* (Vols. 1). London, UK: The Stationery Office.
- House of Commons: Children, Schools and Families Committee. (2009). *National curriculum. Fourth report of session 2008-09* (Vols. 1). London, UK: The Stationery Office.
- Isaacs, T. (2010). Educational assessment in England. *Assessment in Education: Principles, Policy & Practice*, 17(3), 315-334.
- James, M. (2011). Assessment for learning: Research and policy in the (dis) United Kingdom. In R. Berry & B. Adamson (Eds.), *Assessment reform in education* (pp. 15-32). New York, NY: Springer.
- McCallum B., McAlister S., Brown M., & Gipps, C. (1993). Teacher assessment at key stage one. *Research Papers in Education*, 8(3), 305-327.
- Nuttall, D. L., & Stobart, G. (1994). National curriculum assessment in the U.K. *Educational Measurement: Issues and Practice*, 13(2), 24-27, 39.
- Office of Qualification and Examination Regulation. (2014). *Completing GCSE, AS and A level reform*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/377804/2014-06-24-completing-gcse-as-and-a-level-reform.pdf
- Qualifications and Curriculum Authority. (2008). *Optional tests for years 3, 4 and 5. Supporting you in your teacher assessment*. Retrieved from http://webarchive.nationalarchives.gov.uk/20081230200534tf_/http://www.qca.org.uk/qca_13503.aspx

- Reed, J., & Hallgarten, J. (2003). *Time to say goodbye? The future of school performance tables*. Retrieved from <http://www.leeds.ac.uk/educol/documents/00003500.htm>
- Roberts, N. (2014). *National curriculum review*. Retrieved from <http://dera.ioe.ac.uk/22754/1/SN06798.pdf>
- Standards and Testing Agency [STA]. (2015a). *Interim teacher assessment frameworks at the end of key stage 1*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/461547/Interim_teacher_assessment_frameworks_at_the_end_of_key_stage_1_PDFA.pdf
- Standards and Testing Agency [STA]. (2015b). *Interim teacher assessment frameworks at the end of key stage 2*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/473675/Interim_teacher_assessment_frameworks_at_the_end_of_key_stage_2_PDFA_V3.pdf
- Standards and Testing Agency [STA]. (2016a). *2016 key stage 1 assessment and reporting arrangements (ARA)*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/518970/2016_KS1_Assessment_and_reporting_arrangements__ARA__April_PDFA.pdf
- Standards and Testing Agency [STA]. (2016b). *2016 key stage 2 assessment and reporting arrangements (ARA)*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/512097/2016_KS2_Assessmentandreportingarrangements__ARA__PDFA.pdf
- Standards and Testing Agency, & Department for Education. (2015). *Reception baseline assessment: Guide to signing up your school*. Retrieved from <https://www.gov.uk/guidance/reception-baseline-assessment-guide-to-signing-up-your-school>
- Stanley, G., MacCann, R., Gardner, J., Reynolds, L., & Wild, I. (2009). *Review of teacher assessment: Evidence of what works best and issues for development*. Retrieved from http://oucea.education.ox.ac.uk/wordpress/wp-content/uploads/2011/01/2009_03-Review_of_teacher_assessment-QCA.pdf
- Torrance, H. (1993). Combining measurement driven instruction with authentic assessment: Some initial observation of national assessment in England and Wales. *Educational Evaluation and Policy Analysis*, 15(1), 81-90.

- Torrance, H. (2011). Using assessment to drive the reform of schooling: Time to stop pursuing the Chimera?. *British Journal of Educational Studies*, 59(4), 459-485.
- Whetton, C. (2009). A brief history of a testing time: National curriculum assessment in England 1989-2008. *Educational Research*, 51(2), 137-159.
- Wolf, A. (2011) *Review of vocational education: The Wolf report*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/180504/DFE-00031-2011.pdf

投稿收件日：2017年4月7日

接受日：2017年11月22日

