

教育科學研究期刊 第五十九卷第一期

2014年，59(1)，1-32

doi:10.6209/JORIES.2014.59(1).01



## 十二年國民基本教育的入學考試與評量變革

宋曜廷

國立臺灣師範大學  
心理與教育測驗研究發展中心

周業太

國立臺灣師範大學  
心理與教育測驗研究發展中心

曾芬蘭

國立臺灣師範大學  
心理與教育測驗研究發展中心

### 摘要

十二年國民基本教育已正式啟動，入學考試和班級評量的變革，是影響十二年國民基本教育成敗的關鍵之一，如何達成所提出的願景，格外引人關注。以標準為本位的評量為近年來的教育評量改革趨勢，許多先進國家已將標準本位評量分別應用在教育會考與班級評量，以達到學力監控、提升教師教學與評量專業，並促進學生學習成果的目標。本研究介紹標準本位評量之意涵，並以國中教育會考與國民中學學生學習成就評量標準為例，說明標準本位評量如何為實施十二年國民基本教育創造有利環境，進而達成多元學習、確保學力、舒緩升學壓力、活化教學等目標。

**關鍵字：**十二年國民基本教育、班級評量、國中教育會考、國民中學學生學習成就評量標準、標準本位評量

## 壹、緒論

歷經多年的倡議與規劃，十二年國民基本教育（以下簡稱十二年國教）自 2011 年起正式啟動，我國的教育工程正式邁向一個新的里程碑。為了順利推動此重大教育政策，主管機關提出七大工作要項與十項方案，在眾多措施中又以入學制度及評量方式的變革最引人關注。依據高中、高職及五專免試入學實施方案，高中、高職及五專多元入學管道將於 103 學年度起整合為「免試入學」與「特色招生」兩種管道，免試入學意指學生不需參加入學考試，經錄取報到後即可進入高中、高職或五專就讀（教育部，2013a），特色招生係指具學科或術科性向的學生參加入學測驗，學校據此招收與其發展特色相符之學生（教育部，2013b）。自 103 學年度起免試入學名額比例將逐年提高且考試入學名額逐漸降低（分別達 75% 以上及 25% 以下），由此可知，未來大部分國中畢業生都將透過免試管道升學。

十二年國教歷經多年規劃，但在眾所矚目的入學與考試制度的變革，卻始終未能獲得共識。主張降低升學壓力的團體與主張維持競爭力的團體，各有論述，各持立場，讓教育政策在天平的兩端擺盪，難以定奪。目前定調以免試為主的升學制度固然有助於緩解升學壓力，卻也不免產生學生學習動機下滑、學習成就品質低落之隱憂。如何制定出能同時平衡降低考試壓力，但又能維持適當競爭力的入學與考試制度，為十二年國教的重要挑戰之一。除了學校以外的外部考試，如何發揮學校內部班級評量，透過低利害（low-stakes）考試評量以帶領學校教學方式的變革，落實培養學生高層次與多元學習能力，以及活化教師教學的目標，是十二年國教的另外一個迫切面臨之課題。

實施以標準為本位之評量方式（standards-based assessment, SBA，以下稱標準本位評量）為許多先進國家與地區（如美國、澳洲、英國與香港等）近年來的教育評量改革趨勢，這些國家（或地區）一方面將標準本位評量應用在外部考試以監控學力，另一方面也將其應用在班級評量，透過促進課程標準、教學與評量三者的對應，提升教師評量專業與學生學習成效（Lyon, 2011; Martone & Sireci, 2009）。以美國為例，2001 年公布的《沒有孩子落後法案》（No Child Left Behind Act, NCLB）要求各州政府確保學生每年都有適度的學習進展，並以每位學生在核心學科上皆能達到優秀（proficient）等級為目標（Linn, Baker, & Betebenner, 2002），為了達到《NCLB 法案》的要求，許多州政府開始訂定適用該州的評量標準且每年施測大型標準化測驗（large-scale standardized assessment）以監測學生整體成就水準。在班級評量部分，美國加州（California State）針對英語能力制定了該州的評量標準，並於部分公立學校實施標準本位班級評量（standards-based classroom assessment, SBCA），希望藉此提升教師教學與評量專業並回饋到學生學習成果（Llosa, 2011）。

本研究之目的在於闡述標準本位評量的理念，以及如何透過標準本位評量促進十二年國教之入學與考試制度變革，進而提升實現十二年國教理想的可行性。

## 貳、標準本位評量

### 一、標準本位評量之意義

標準本位評量是指在評量前制定系統化的評量標準並據此發展評量工具，評量後將結果對照到評量標準以將學生學習表現劃分成不同表現等級，進而解釋與說明該生學習進展之評量歷程。建置評量標準是實施標準本位評量的先決條件，為使評量結果能用來瞭解學生是否達成教師的教學目標，評量工具所測量的知識概念應明確對應到教師的授課內容。在實務上，教師應根據官方制定的課程規範（例如我國的課綱）設立教學目標、設計課程與教材、進行教學活動等，因此，若能同樣根據課綱與能力指標建構評量標準，將有助於串連評量與教學，評量結果更有助於瞭解學生是否具備課綱期望的學習成就（宋曜廷，2012a）。

### 二、標準本位評量之組成要素

評量標準由內容標準（content standards）與表現標準（performance standards）所構成，內容標準意指希望學生具備的知識與展現的技能（Hambleton, 2001），例如，我國數學科課綱針對七年級學生，期望其能具備認識數與數線、理解因數與倍數或一元一次方程式等知識。表現標準則是說明在內容標準所條列的學習內容中，學生能展現相關的知識與技能達到何種程度（Linn & Herman, 1997）。由於不同學生在教學活動後可能會呈現出不同的表現程度，因此須設立適當數量的表現等級（performance levels），用以大致區分學生學習進展的差別，例如，美國「國家教育進展測驗」（National Assessment of Educational Progress, NAEP）將學生學習表現區分為進階（advanced）、優秀（proficient）、基礎（basic）與未達基礎（below basic）等，共四個等級。此外，需針對各個等級撰寫表現等級描述（performance level descriptors, PLDs），讓教師可以進一步瞭解不同等級學生的典型表現或最低門檻水準。

透過將學生表現對應至預先建置的評量標準，教師便可界定出學生在特定學習內容上的表現等級，此種參照到標準以產生成績等級的評量方式，改善了常模參照（norm-referencing）的部分缺失。在常模參照中，透過與群體中其他人的測驗分數進行比較，可以獲得該生在群體中的百分等級（percentile rank, PR）以瞭解相對地位表現，該量化分數雖有利於分發入學，但也被認為助長了對測驗分數斤斤計較的現象。此外，量化分數未能提供學生學習成果的訊息，教師難以根據評量結果進行診斷與補救，例如，我們可以知道 PR 值為 40 的學生其表現低於群體中六成的學生，但無法指出該生缺乏何種知識或哪些技能需要補強。相對地，在標準本位評量中，對該生表現等級的界定不需要與他人比較，評量標準可提供各表現等級的質性表現描述，因而能指出該生目前具備的知識與技能。因此，以標準為本位的評量方式不但可降低競爭壓力、讓學生專注在提升自身學習成果，更能拓展評量結果的解釋與應用，提供

教師與學生更多學習成就的回饋訊息。

### 三、標準本位評量與學校本位評量、標準參照評量之關係

標準本位評量與標準參照 (standards-referencing) 評量非常相似，兩者皆將學生表現參照至事先制定的標準或規範，進而提供學生會做什麼與不會什麼的學習成果訊息。標準本位評量可以是學習階段結束後實施的總結性評量 (summative assessment)，例如國中教育會考，以評估學生 3 年來的整體學習概況，也可以是教師於課堂傳授課程內容後的形成性評量 (formative assessment)，例如隨堂小考，以評估學生對特定知識的吸收程度，因此標準本位評量可以應用在學校外部大型測驗，也能在學校內部班級中實施。

將標準本位評量應用在班級中，與澳洲自 1970 年代推行的學校本位評量 (school-based assessment) 有諸多雷同之處。在理念上，學校本位評量強調促進教師學科與評量專業素養以提升教學品質，提供學生適性與優質之教育機會；在作法上，由學生自己的教師訂定教學目標、選用適當的教材與方法、設計多元化的評量工具，將多次評量結果對照至評量規範以界定學生學習成果 (宋曜廷、周業太、吳佩瓊、林秀珊、曾芬蘭，2010)。然而，兩者最大的差別為：在學校本位評量中各校可建立自己的評量規範，不同學校的教師可能使用不同的評量準則核定學生表現，如此一來，不但不同學校教師對相同表現等級的界定難以溝通，跨校之間的評量結果更需透過分數調整機制 (moderation) 才可進行比較。相較之下，標準本位評量是根據國家或地方政府公布的課程標準來制定評量標準，不同學校的教師使用同一套評量標準核定學生表現，此全國或全區統一的評量標準有助於降低教師對各等級學生典型表現 (或最低門檻表現) 的認知差異，評量結果更適合作為學生學力監控的依據。學校本位評量和標準本位評量之相關特性比較，整理如表 1 所示。整體而言，標準本位評量兼採學校本位評量與標準參照評量之理念，較能廣泛應用在外部考試與班級評量，並有助於改進常模參照需與群體中其他人的考試成績做比較、無法提供學生具體學習成效之不足。

表 1

學校本位評量和標準本位評量之特性比較

測驗別特性	學校本位評量	標準本位評量
性質	總結性評量	兼具形成性與總結性評量
功能	<ul style="list-style-type: none"> <li>● 甄別：作為入學依據</li> <li>● 授證：授予畢業證書</li> <li>● 學習：促進學習發展</li> </ul>	<ul style="list-style-type: none"> <li>● 甄別：作為入學依據</li> <li>● 授證：授予畢業證書</li> <li>● 學習：促進教學相長</li> <li>● 篩選：投入補救資源</li> </ul>
評量規範	各校可自訂評量規範	國家或地方政府制定統一的評量標準

(續)

表 1  
學校本位評量和標準本位評量之特性比較（續）

測驗別特性	學校本位評量	標準本位評量
評量環境	學校內部之班級評量	學校內部之班級評量 學校外部之大型測驗
評量執行者	學生自己的教師	班級評量：學生自己的教師 外部測驗：專業測驗單位
教師的角色	<ul style="list-style-type: none"> <li>• 學習歷程的監督者</li> <li>• 學習結果的仲裁者</li> <li>• 壓力主要來自測驗作業的設計困難與增加的負擔</li> <li>• 需要的評量專業支持高</li> </ul>	<ul style="list-style-type: none"> <li>• 學習結果的監督者</li> <li>• 學習結果的仲裁者</li> <li>• 壓力主要來自測驗作業的設計困難與增加的負擔</li> <li>• 需要的評量專業支持高</li> <li>• 對評量標準的理解程度高</li> </ul>
評量結果的公信力 (public credibility)	低，由於評分標準和作業內容不一致，多數經過調校的過程才具可比性與公信力	高，評量標準全國統一，公信力高，評量結果具可比性

註：修改自「從學校本位評量省思臺灣擴大免試入學方案」，宋曜廷、周業太、吳佩瑛、林秀珊、曾芬蘭，2010，*教育科學研究期刊*，55（2），p. 81。

## 參、臺灣入學與考試制度變革的理想與現實

近五十多年來，臺灣後期中等學校入學與考試方式的發展，大致可分為以下三個階段（宋曜廷等，2010）：第一，傳統高中聯招時期。高中聯考制度自 1956 年開始實施，其目的在於公平地篩選出不同能力水準之學生，全面統一的考題以及依分數高低排序分發學校，是當時大多數人所認同的鑑別工具與入學方式。然實施多年以後，面臨單一次考試機會加大學生壓力、有限的命題委員讓試題品質無法令人信服、缺乏預試讓試題的難度與鑑別力難以掌握等問題與爭議。有鑑於此，教育當局規劃並推行高級中學多元入學方案，於 90 學年度停辦多項聯招並改以實施國民中學學生基本學力測驗（以下簡稱國中基測）。第二，國中基測時期。國中基測自 2001 年開始舉辦，測驗學科包含國文、英語、數學、社會、自然與寫作測驗（自 2007 年起新增）六科，測驗範圍涵蓋國中 3 年的課程。除寫作測驗為標準參照的級分制之外，餘均以量尺分數呈現各科總分，分數計算方式為寫作測驗級分乘以二，再加上各科總分，自 2009 年起滿分為 412 分（宋曜廷、許福元、曾芬蘭、蔣莉蘋、孫維民，2007）。國中基測屬於標準化總結性測驗，旨在評估九年級學生截至該階段所具備的基本能力與學習的成果，不同於傳統聯招，國中基測注重測量學生對學習內容是否融會貫通。由於針對測驗題目進行反覆練習的教學方式難以在基測取得高分，可使教學回歸培養學生應用知識能力的面向，1 年舉辦兩次

考試的設計更可避免學生因參與測驗時失常造成之遺憾。第三，十二年國教時期。十二年國教從 2011 年正式推動至今，為了因應高中、高職與五專多元入學管道，自 103 學年度整合為免試入學與特色招生兩種管道，自 2014 年起停辦國中基測並改以實施國中教育會考以瞭解學生學力品質，並舉辦特色招生考試以達成引導學生多元智能卓越發展之政策目標。

綜觀前述不同時期的入學制度與考試方式，可以發現這些教育改革所欲達成之目的與十二年國教所揭櫫的政策目標相似，不外乎下列四項：第一、提升國民素質、增進國家競爭力。第二、縮小教育落差，促進教育機會均等，實現社會正義。第三、改變教學模式，活化教學方法，降低城鄉差距。第四、舒緩升學壓力，引導學生適性發展。換言之，「低一點的升學壓力、高一點的學習彈性、活一點的教學方式、多一點的社會正義」可以說是變革的主要目標。然而，這四項目標實為如何彼此均衡的問題，而非追求如何同時達到最佳狀態，更進一步來看，這四項目標由兩種張力彼此牽引，以下說明與討論這兩種張力造成的教育困境，藉由瞭解彼此如何牽引拉扯，才能針對現有困境提出可行的解決策略以達成彼此均衡之目標。

## 一、重視多元發展 vs. 強調弱勢權益

多元入學方案強調使用多元化的入學指標（如社團表現、競賽成績、口試等）作為選才依據，雖然多元發展、多元選才的立意良善，相關研究卻指出，教師及一般民眾認為多元入學指標不利社會弱勢族群身分的學生升學，社會大眾認為富裕家庭的子女可以藉由經濟優勢培養多元能力，在多元升學進路中取得更大的優勢，相對剝奪經濟弱勢家庭學生的升學機會（譚光鼎、莊勝義、魯先華、康瀚文、陳怡璇，2007）。另一研究指出，上述問題導因於任何人都想取得更佳的排名進而獲得更多的教育資源，因此任何影響排名的因素其客觀性都會被過度要求，最終導致多元入學方案未能發揮多元選才、適性入學的效果（張道行、陳清誥、徐慧萍、許福元，2007），社會大眾可能因此誤認為多元入學方案加深了城鄉與階級差距，影響入學機會的公平性，這也導致以往的國中基測儼然成為唯一的入學依據（吳清山、高家斌，2007）。在弱勢權益部分，研究指出我國從 2004 年至 2010 年，屬於原住民與非原住民族群的兩組九年級學生之間的學習成就落差有縮減的趨勢，然而，學校學區接近都市與地處偏遠的兩組學生，以及家長社經地位較高與社經地位較低的兩組學生，都觀察到學習成就落差有上升的趨勢（Sung, Tseng, Kuo, Chang, & Chiou, 2014）。雖然實施多元入學方案不見得導致學習成就落差加劇，但此現象提醒主管機關在倡導多元發展的同時，也應提出對應的配套措施扶植學習上的弱勢學生。

## 二、降低升學壓力 vs. 增進國家競爭力

入學考試往往被認為加重了國中學生的升學壓力，其原因不外乎常模參照的方式需將個別學生的測驗分數與群體表現進行比較，以獲得學生的相對地位表現，再加上測驗分數級距較多（例如國中基測的量尺分數從 1 分至 80 分），學生與家長對於考試成績分分計較，希望

獲取更高的表現地位以利升學，在此考試升學風氣盛行之下或許可以解釋為何多元入學方案無法有效紓解國中學生的升學壓力（吳清山、高家斌，2007）。國立臺灣師範大學心理與教育測驗研究發展中心（以下簡稱心測中心）在 2011 年針對九年級學生的考試壓力進行調查研究，結果發現，女生面臨的考試壓力明顯高於男生，也發現壓力最大的族群是 PR 值介於 50 至 59 的學生（宋曜廷、邱佳民、張恬熒、曾芬蘭，2011；宋曜廷、趙子揚等，2013）。適逢十二年國教正式實施之際，以免試為主的入學管道被預期可以降低升學壓力，然而，此舉同時也面臨維持學生學習動機以及增進國家競爭力之問題。針對不同能力的學生採取不同策略可以達到適度提升競爭力之目的，針對程度好的學生應擴大優秀學生的定義，打破 PR 值在 98 或 99 以上的學生才是頂尖的迷思，一旦 PR 值 90 以上的學生也有機會進入明星高中或特色高中，相信可緩解對考試成績分分計較之現象，優秀學生不需耗費時間反覆練習，因而有更多時間與資源培養未來所需的多元能力。再者，針對 PR 值介於 50 至 59 的中等程度學生，由於這些學生的升學落點恰巧或在公、私立學校之間，可能因此承受較大之壓力。這群學生需要的可能不是大班級的課後補習，而是需要更多的診斷與個別式教學，基於他們本身已有的能力基礎，適量的協助就可以產生槓桿作用，提升他們的基礎能力至中上程度，更加保障中段學生的競爭力。針對學習成就落後與不足的學生，可能需要先瞭解其是否對學術領域的學習沒有興趣，進而積極輔助他們在技藝領域的學習，讓學生能真正適才適所的發展，開拓其未來的就業機會（宋曜廷，2012b）。

施行標準本位評量有助於體現上述教育改革所揭櫫的政策目標，因此，以下針對先進國家應用標準本位評量於教育會考與班級評量之理念與作法進行介紹，並對臺灣施行國中教育會考與研發國中學生學習成就評量標準之現況進行說明。

## 肆、標準本位評量在教育會考之應用

### 一、先進國家教育會考之設計與實施

#### （一）香港

##### 1. 沿革與目的

香港政府自 2012 年起針對完成 6 年中學課程（3 年初中及 3 年高中教育）學生實施香港中學文憑考試（Hong Kong Diploma of Secondary Education Examination），評量結果以表現等第區分學生的表現，主要目的是讓學生瞭解自己學習的長處與短處，並提供回饋改善學習狀況，讓教師根據評量的結果檢視與修訂學習目標、教學策略與內容，並提供政府瞭解學生的學業水平（香港考試及評核局，2013）。

## 2. 設計與成績報告

大部分考生報考四個核心科目（中國語文、英國語文、數學及通識教育）及二至三個選修科目，選修科目分為三類：甲類為新高中科目（如：地理、生物、科技與生活等共二十個學科）、乙類為應用學習科目、丙類為其他語言科目（香港考試及評核局，2013）。

香港中學文憑考試的成績報告採用標準參照的方式，以甲類新高中科目為例，考生表現分為五個等級（1-5 級），第 5 級為最高表現等級，針對各等級典型學生之共通等級描述如附錄一所示。由於香港中學文憑考試除了作為學力評核機制也兼具高風險入學考試之角色，考試成績同樣作為錄取學士學位課程與否的入學條件。為了維持入學測驗的區別效果以利分發入學之用，獲第 5 級考生中約 10% 表現最佳者被進一步標示為「5\*\*」，約 30% 表現次佳的第 5 級考生則以「5\*」標示（香港考試及評核局，2011）。

### （二）美國伊利諾州

自從《NCLB 法案》實施以後，美國許多州政府開始訂定適用該州的評量標準且每年施測大型標準化測驗以監測學生整體成就水準，以下舉美國伊利諾州（Illinois State）為例進行說明。

#### 1. 沿革與目的

伊利諾州政府每年舉辦一次伊利諾州標準成就測驗（Illinois Standards Achievement Tests），<sup>1</sup>該測驗由伊利諾州教育委員會邀集教師與課程專家發展而來，測驗結果用以評估達到伊利諾州學習標準之學生比例，進而監控學生學力變化並作為政策調整之依據（Illinois State Board of Education, 2010）。

#### 2. 設計與成績報告

不同年級的學生需應試不同的考試科目，例如三至八年級學生皆須接受閱讀與數學測驗，而科學測驗則僅施測四年級與七年級。考試題型部分，數學測驗包含選擇題、簡答題及申論題（extended-response）；閱讀測驗分為選擇題與申論題；科學則以選擇題為主。學生在伊利諾州標準成就測驗的表現被分為超越標準（exceeds standards）、符合標準（meets standards）、低於標準（below standards）和學習警告（academic warning）四個等級，各等級之等級描述如附錄二所示。

### （三）澳洲

#### 1. 沿革與目的

澳洲政府自 2008 年起每年舉辦一次全國讀寫和計算能力測驗（National Assessment Program-Literacy and Numeracy, NAPLAN），<sup>2</sup>教育主管機關、學校、教師和家長可透過

<sup>1</sup> 請參考 <http://www.isbe.state.il.us/assessment/isat.htm>

<sup>2</sup> 請參考 <http://www.nap.edu.au/naplan/naplan.html>

NAPLAN 瞭解學生在該階段的學習情況，進而評估學生在讀寫和計算方面是否達到國家要求的教育標準，考試結果亦可作為改善課程內容與教育政策調整之參考依據，NAPLAN 成績並不作為學生升學與教師績效之用（Australian Curriculum, Assessment and Reporting Authority [ACARA], 2013）。

## 2. 設計與成績報告

NAPLAN 考試針對 3、5、7 與 9 等四個年級學生進行全國普測，考試科目包含閱讀、寫作、語言習慣（包含拼寫、語法與標點符號）及計算（包含：數字、空間、代數、函數等）四科。NAPLAN 採用常模參照的方式，將學生測驗分數與全國所有學生進行比較，並設立五個不同 PR 值（5、20、50、80 與 95）將學生區分成六個等級區間，以三年級為例，PR 值低於 5 的學生被歸類為第 1 級，依此類推 PR 值 95 以上的學生被歸為第 6 級，若未達第 2 級別代表該生未達國家要求的最低學習進展標準。成績單上除了呈現學生的測驗分數與成績級數之外，還會標示全國考生的平均位置以利家長瞭解學生相對表現，此外，也會針對學生所屬等級說明學生具備何種知識與技能（ACARA, 2013）。

## 二、臺灣國中教育會考之設計與實施

### （一）緣起

教育部於 2012 年公布的「國民小學及國民中學學生成績評量準則修正條文」中，明文規定應由教育部會同直轄市、縣（市）政府辦理國中教育會考，以瞭解並確保國民中學學生學力品質（教育部，2012）。

### （二）設計與成績報告

國中教育會考的目的為評量國中學生 3 年來總結性學習成果，每年定期於 5 月份舉辦，針對所有九年級學生施測國文、英語、數學、社會、自然與寫作測驗等六科。教育部於 2012 年 4 月發布的新聞稿中指出，<sup>3</sup>國中教育會考國文、英語、數學、社會、自然五個科目採標準參照方式呈現學生各科結果，透過標準設定（standard setting），各科評量結果分為「精熟」、「基礎」及「待加強」三個等級，「精熟」表示學生精通熟習該科目在國中階段所要求學習的知識與能力；「基礎」表示學生具備該科目國中階段之基本學力；「待加強」表示學生尚未具備該科目國中教育階段之基本學力（國立臺灣師範大學心理與教育測驗研究發展中心，2013），附錄三為國中教育會考各考科在不同等級之等級描述。

為了區分學生所屬等級，由測驗專家、各考科之大學教授及國、高中教師在會考後召開會議，會中進行三個回合的標準設定程序，與會成員在每個回合中需設立獲得不同表現等級所需的最低答對題數，並在下一回合開始前針對較無共識的試題進行討論以凝聚共識，最終

<sup>3</sup> 請參考 <http://12basic.edu.tw/Detail.php?LevelNo=883>

針對各科目個別設立兩個切截分數 (cutoff score)，以利將學生區分成三個不同表現等級 (宋曜廷、曾芬蘭、邱佳民，2013)。

國中教育會考的原始設計是作為學力監控之用，然而，在 2014 年各學區仍將採用會考的評量結果作為免試入學超額比序項目的參考依據。為了解決升學競爭激烈區域可能的抽籤困境，教育部委託心測中心研擬增加表現等級數量之可行性。心測中心綜觀先進國家與地區之教育會考制度，發現香港中學文憑考試除了作為學力評核機制也扮演高風險入學考試之角色，與國中教育會考初期的雙重角色相似，因此參考其作法有助於解決我國所面臨的入學分發問題。香港中學文憑考試針對第 5 級考生中約 10% 表現最佳者進一步標示為「5\*\*」，約 30% 表現次佳的第 5 級考生則以「5\*」標示，此舉在標準本位評量的架構中融合了常模參照的作法，以顧及入學分發之實務需求。有鑑於此，心測中心在各科仍區分三等級以維持學力監控目的之核心原則下，嘗試多種加註標示之組合並進行分發模擬，並於 2013 年 7 月提出以下建議：<sup>4</sup>針對精熟等級 (A 級) 中分數最高的前 25% 與前 26%~50%，分別加註標示 A++ 及 A+，依此類推，在基礎等級 (B 級) 中加註標示 B++ 及 B+，待加強等級則不加註任何標示，故成績等級分為三等級與四標示。相信此加註標示之作法能兼顧國中教育會考之學力監控目的與免試入學之分發實務需求，適合臺灣重視升學考試的國情，並能適度解決入學超額比序問題。

### (三) 目的與預期效益

國中教育會考成績單除了標示學生表現等級更提供各等級的表現描述，此成績呈現方式有助於達成以下目的：第一，作為學力監控之機制。學生可根據表現等級瞭解自己在各學科上的學習成果，主管機關也可據此評估全體學生在單一年度之學力表現，或追蹤學生在不同年度之學力變化趨勢，達到監控學力之主要目的。第二，縮減學習成就落差、維持社會正義。學習成就落差是許多國家關注的教育議題，在維持社會公平正義的理念下，樂見來自不同族群、不同家長社經地位與不同學區的學生，其學習成就沒有明顯的落差。透過國中教育會考對於待加強等級的界定，將有助於教師找出學習成就低落的學生，進而投入相關補救資源，以消弭不同背景學生之成就落差，待加強學生人數比例的變化也可作為主管機關檢視政策施行成效的參考依據。第三，降低分分計較現象，緩解學生考試壓力。在國中教育會考的成績計算方式中，學生表現與事先制定的評量標準進行比對而不是與他人的成績做比較，同儕間的競爭壓力可望大幅下降。此外，成績等級僅劃分成三個等級 (至多增加四個標示) 而非過去國中基測的八十個分數級距，這些變革皆有助降低分分計較現象，減輕學生升學壓力。

表 2 根據實施對象、考試目的、考試科目及評分等級等項目，綜合歸納各國與地區在中學階段之教育會考制度。可以發現，施行教育會考之主要目的皆為學力監控之用，並提供主管機關作為政策調整之參考依據。不同國家的實施對象與考試科目不盡相同，表現等級的劃

<sup>4</sup> 請參考 <http://12basic.edu.tw/NewDetail.php?Seq=159>

表 2

香港、美國、澳洲與臺灣中學階段教育會考制度綜合比較

國家與地區	考試名稱	實施對象	考試目的	考試科目	表現等級
美國	伊利諾州標準 成就測驗	公立學校三至 八年級學生	學力監控與政 策調整	(1) 三至八年 級需接受 閱讀與數 學測驗  (2) 四、七年級 需接受科 學測驗	分為超越標 準、符合標準、 低於標準、學習 警告四個等級
澳洲	全國讀寫和計 算能力測驗	全國三、五、 七、九年級學 生	學力監控與政 策調整	閱讀、寫作、 語言習慣與計 算	單一年級分為 六個成績等 級，跨年級總計 十個成績等級
香港	香港中學文憑 考試	完成6年中學 教育學生（18 歲）	學力監控、政 策調整與入學 分發	以甲類為例， 包含中國語 文、英國語 文、數學、通 識教育及兩至 三個選修科目	分為 1-5 個等 級，針對第5級 進一步區分5** 與5*
臺灣	國中教育會考	完成國中教育 學生（15歲）	學力監控、政 策調整（2014 年仍採用會考 成績作為入學 分發）	國文（含寫作 測驗）、英語、 數學、社會及 自然	分為精熟 （A）、基礎 （B）、待加強 （C），針對A級 與B級進一步 區分A++與 A+、B++與B+

分，以臺灣國中教育會考僅區分三個等級為最少，澳洲 NAPLAN 由於採用常模參照，跨年級區分為十個等級為最多。值得注意的是，臺灣與香港因考試風氣較為盛行，仍（部分）採計會考成績作為升學依據。

就標準本位評量的實施成效部分，以美國密蘇里州（Missouri State）為例，該州於 1996 年在某三個學區試驗性的使用以標準為本位的數學教材（對應到國家數學課程標準所發展而來）進行至少 2 年的教學。後續研究指出，在控制學生數學能力水平與家庭經濟等因素相似的條件下，發現接受標準本位數學教材的學生，其學業成就表現較佳（Reys, Reys, Lapan, Holliday, & Wasman, 2003）。若以美國整體來看，自 2000 年以來各種族群在 NEAP 的數學表現

呈現出進步的趨勢，例如四年級黑人學生在 2013 年 NAEP 的數學表現被歸類為優秀以上的比例，從 2000 年的 4% 上升至 18%。雖然無法斷定標準本位評量為美國學生整體數學成就提升之主要原因，但標準本位評量之實施成效可見一斑。

## 伍、標準本位評量在班級評量之應用

將標準本位評量應用在班級中，稱為標準本位班級評量（standards-based classroom assessment, SBCA）（Llosa, 2011），意指教師在設定教學目標、設計課程、教材與評量工具、進行評量與給分等歷程中，皆對應到事先制定的評量標準（宋曜廷，2012a）。相較於學校外部考試，實施班級評量因具備以下特點，因此近來益發受到重視（宋曜廷等，2010；Llosa, 2011）。

第一，SBCA 有助於提升課綱、教學與評量三者的對應（alignment）程度，將有助於發揮引導教學與學習的效果。在課程標準部分，我國的課綱與能力指標明確訂出學生在每一個學習階段中應習得的知識、技能與態度。在學習與教學活動部分，教師可以從內容標準與表現標準的內涵清楚知道該傳授哪些學科知識，並據此訂定教學目標、設計與選用適當的教材與方法。此外，評量標準也可扮演學習地圖的角色，讓學生瞭解目前的成就等級與想要達到等級之前的落差，讓學生朝明確的方向努力學習以提升自身的表現。在評量部分，表現描述則有助於教師劃分學生學習成就，瞭解學生已經習得哪些知識概念，判斷其是否達到課前要求的水準，診斷其是否有迷思概念並能立即進行補救教學，進而回饋到調整教材與教法。三者彼此相輔相成，進而有助於達成課綱引導教學、以評量結果促進教與學之理念（宋曜廷，2012a）。SBCA 可以將課程標準、學習與教學活動，以及評量三者緊密連結，如圖 1 所示。

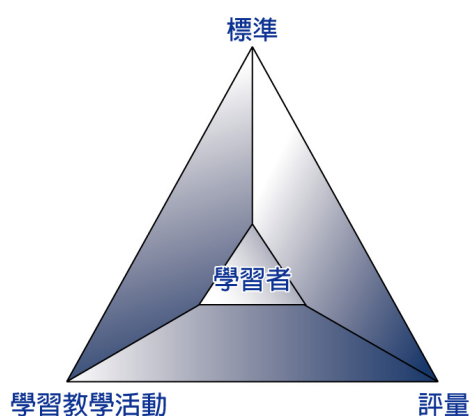


圖1. 課程標準、學習與教學活動以及評量三者關係示意。修改自 *Using standards in your classroom: A teacher resource guide*, Vermont Department of Education, 2012, retrieved from <http://www.wssu.k12.vt.us/teachers/docs/UsingStandards.pdf>

第二，SBCA 能評量學生更多內容面向。有別於外部考試受限於固定的測驗日期與時間、有限的測驗題數等，僅能針對少數學科內容進行測驗，班級評量可在教學活動結束後進行多次施測，因而可評量到較多學習面向。此外，SBCA 可透過不同的形式評量學生各類能力，更能多元蒐集學生表現。例如，可使用紙筆測驗評量學生在實驗程序的知識，可使用實作評量要求學生選取器材、放置儀器並進行實驗，以評估其實驗的技巧。在語文課中可以評量學生表達溝通能力，而非只有閱讀或聽力等接收性能力。這些能力在高風險、高利害的考試中提倡的成本過高，不僅不易達成，且有過度增加學生壓力的疑慮。

第三，SBCA 能體現評量之完整與多元的功能。傳統上多將評量的角色界定在「對學習的評量」(assessment of learning)，指透過評量對於學生的學習歷程或成果進行評鑑或考核，也多著重在學習成效的總結評核功能。隱含著由老師對學生做評量，同時也隱含著評量與學習間的疏離關係。近年來，評量著重另外兩個角色，一為「為學習而評量」(assessment for learning)，意指評量的目的是要給予學生回饋，進而促進其學習，更著重評量的形成性評鑑之功能，同時也隱含著評量與學習更緊密的連結。評量的另一個新角色為「評量即學習」(assessment as learning)，指學生透過評量活動即可學習，隱含著學生由被動的評量者，轉換為主動的評量者。許多具有學習功能的評量活動，例如自我和同儕評量 (self- and peer assessment) 與檔案評量 (portfolio assessment)，都強調學生在評量中的主動角色，以及透過評量活動來學習，因此在班級評量活動中受到重視 (Sung, Chang, Chiou, & Hou, 2005)。SBCA 的實施，能夠體現評量的多重角色和功能。

## 一、先進國家應用標準本位評量於班級評量之現況

### (一) 美國加州

加州很早就發展該州的評量標準，以英語能力為例，該州於 1999 年公布英語能力發展標準 (English-Language Development Standards)，針對聽與說、讀、寫三個向度，將英語能力表現區分為五個不同表現等級，此標準一方面作為教師授課時的教學依據，一方面也作為評量工具的設計依據，加州英語能力發展測驗 (California English Language Development Test) 與一系列的班級評量工具便是根據此標準研發而來。在加州其中一個大型公立學區，教師被要求利用這些與標準對應的評量工具實施班級評量，評量結果一方面用來找出學生的迷思概念，一方面用以區分學生屬於有限的進展、部分進展、平均進展或者明顯進展等何種類別。Llosa (2007, 2008) 將班級評量結果與加州英語能力發展測驗的結果進行比較，研究結果指出，教師實施多次班級評量後對學生整體英語能力的判斷，與加州英語能力發展測驗的結果一致，該結果說明了當評量工具根據評量標準研發而來，不論是外部評量或是班級評量皆能提供有效的總結性學習成就評量結果。此外，能否將標準內涵清楚地以文字加以描述，以及達成該項標準所需運用的能力數量多寡，皆為影響教師班級評量結果的關鍵因素，若是達成該

項標準需要具備多重能力或者對表現標準的文字描述不甚明確，極可能使得教師判斷學生學習進展的一致性降低。

## （二）美國佛蒙特州

為了幫助教師執行 SBCE，佛蒙特州 (Vermont State) 建構出標準本位學習單元 (standards-based unit of study) 的概念。所謂標準本位學習單元，意指教師針對特定學習內容的教學與評量活動，以圖 1 所示課程標準、學習與教學活動及評量三者之關係為核心架構所進行。若以實施教學活動的順序來看，在標準本位學習單元中的第一步驟為選定標準，例如，在佛蒙特州公布的課程標準中選擇希望學生培養的知識技能。在此步驟中，教師要能夠根據學生的年級與程度選擇適合說明與解釋該標準的學科內容，思考需要的教學時數與使用的教材。第二步為擬訂課堂教學計畫，在設計教學計畫時，教師需要思索學生在教學歷程中的學習經驗能否幫助其達成標準中的要求，以促進學習與教學活動及課程標準之對應，或者思考學生於教學活動後的學習成就可以利用何種評量工具進行評估，以促進學習與教學活動及評量之對應。第三步驟為擬訂評量計畫，在此步驟中，教師需要根據欲評量的標準之本質思考希望學生回答或展現的相對應學習成果，進而決定適合的評量工具類型（如利用實作評量評估學生做實驗的技巧），評估結果才有助於學習者瞭解其是否已達到標準之要求，並回饋教師調整教學活動。此外，設計工具的同時也需擬訂對應的評分規準，以利將評量結果與評量標準進行連結，進而界定學習者於該學習單元所達到的表現等級 (Vermont Department of Education, 2012)。

另外，佛蒙特州教育當局研發了一系列的教學資源與實施手冊，<sup>5</sup>教師可參考官方所提供與特定內容標準有關的相關資料（如教材與評量工具）進行教學活動與評量，以降低教師負擔。若教師選擇使用自行研發的教材進行教學活動，那麼便需要思考這些教材是否真的能幫助學生瞭解課堂中的學習內容、是否需要補充其他教材、自行研發的評量工具之評量結果是否能提供學生學習成果之回饋等問題，以利學生能真正習得欲傳授的知識技能。

## （三）美國共同核心課程標準

在《NCLB 法案》簽署後，部分州政府為了避免學生成就未達法案要求將遭受行政處分之窘境，紛紛降低評量標準之內涵，例如密西西比州 (Mississippi State) 在 2005 年宣稱有九成九的四年級生之閱讀能力達到良好等級，但該州卻在同年實施的一項全國測驗中，僅有不到兩成的學生達到及格；此外，2006 年的 PISA (Programme for International Student Assessment) 成績顯示，美國學生的科學素養與數學素養低於參與國家的平均水準。上述現象顯示，建立全國性的評量標準有其必要與迫切性 (詹紹威, 2012)。因此，2010 年由全國州長協會 (National Governors Association, NGA) 及各州教育官員委員會 (Council of Chief State School Officers,

<sup>5</sup> 請參考 <http://www.wssu.k12.vt.us/teachers/docs/UsingStandards.pdf>

CCSSO) 設立並頒布全美國一致的「共同核心課程標準」(Common Core State Standards, CCSS), 其目的在統一美國從幼稚園到十二年級的課程標準以改善前述問題, 協助學生做好升學及就業準備從而提升國際競爭力 (Common Core State Standards Initiative, 2011)。各州對共同核心課程標準之因應方式不盡相同, 以印第安那州 (Indiana State) 為例, 該州首先整理了共同核心標準與該州原有標準的內涵比較, 以促進教師與家長對共同核心標準之瞭解與接受程度。此外, 該州教育當局針對核心科目 (如數學科) 與不同年級發展了一系列的教學引導工具,<sup>6</sup> 教師可從中取得特定年級學生在此核心科目所需習得的知識與技能列表、不同學習內容的教學順序與時數, 並提供於生活中應用相關技能的簡要範例, 上述措施皆有利於該州於共同核心課程標準的架構下, 於班級中實施標準本位評量。

#### (四) 加拿大卑詩省

卑詩省 (British Columbia Province) 針對閱讀素養、寫作素養、數學素養及社會責任 (social responsibility) 等面向, 根據不同年級制定了表現標準, 並區分出未達預期水準、達到部分預期水準、完全達到預期水準及超越預期水準等四個表現等級, 其中, 「達到部分預期水準」為通過門檻。為了協助教師在班級中應用表現標準, 教育當局針對特定學科內容製作了與表現標準對應的評量工具及評分規準, 並提供獲得各個等級學生的真實表現範例, 以利教師、家長與學生透過表現範例加深對評量標準內涵之理解。

由上述數個國家與地區實施 SBCA 的經驗與作法可知, 為了提升教師在班級中應用評量標準之可行性, 除了制定系統化且文字意涵明確易懂的評量標準之外, 針對特定學科內容設計相對應的評量工具與評分規準, 蒐集學生真實作答資料並呈現不同等級的典型表現 (或最低門檻表現), 皆有利於加深對教師評量標準內涵的瞭解, 進而提升 SBCA 的實施成效。

## 二、國民中學學生學習成就評量標準之制定與應用

### (一) 緣起與目的

配合十二年國教的實施, 未來大部分國中畢業生都將透過免試管道進入高中、高職及五專, 如何在沒有升學考試下, 仍維持國中學生的學習動機和學習成效, 並使學生在學科之外的多元智慧與能力獲得充分擴充與展現, 為國中階段教育評量迫切面臨的問題。此外, 過去課綱各學習領域雖訂有能力指標, 卻不易轉換成適合在教學現場使用的評量標準, 加上缺乏評量示例, 導致教師在進行教學時逐漸忽略課綱、評量時缺乏參照準則 (宋曜廷, 2012a)。研發「國民中學學生學習成就評量標準」(以下簡稱評量標準) 旨在建置與課綱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。評量標準能幫助國中教師區分並瞭解學

<sup>6</sup> 請參考 <http://www.doe.in.gov/achievement/curriculum/mathematics-toolboxes>

生的學習表現，進而調整其教學內容與方式，或設計適當的補救教學課程與活動，以協助學生提升其學習品質，達到維繫基本學力、縮減學習成就落差的任務。

## (二) 制定原則

評量標準的制定範圍目前涵蓋國中階段語文、健康與體育、社會、藝術與人文、自然與生活科技、數學及綜合活動等七大學習領域。其中，內容標準的制定係將各學習領域能力指標加以整併、區分為不同「主題」，並根據學科屬性與教學內容在主題下區分「次主題」。在表現等級部分，總計區分 A 到 E 共五個表現等級，C 級代表基礎門檻並向上進一步劃分出優秀 (A 級) 與良好 (B 級) 兩級，此舉除了可避免過多學生皆被歸類為 C 級，更有鼓勵學生精益求精、培養高層次能力之意涵。對於尚未達到基礎門檻的學生，為了能合理與適當地分配補救教學資源，針對其落後程度的差異，區分不足 (D 級) 與落後 (E 級) 兩級。在表現等級描述部分，由於各個內容 (次) 主題皆對應到一個或多個能力指標，在研發過程中便可根據次主題與能力指標的對應關係，撰寫不同等級應具備的表現等級描述。此外，評量標準進一步研發評分規準 (scoring rubric) 與作業示例 (exemplars)，評分規準為協助教師針對特定學習內容區分學生表現等級的具體評分準則，此規準同樣依據評量標準訂定而來，以作為評估學生於特定作業表現的依據。作業示例為可幫助教師瞭解評量標準的評核範例，包含「樣卷」及「樣卷說明」兩個部分，樣卷指的是從作業或評量資料中挑選而出的各等級學生之典型 (或門檻) 表現示例；樣卷說明則為依據評分規準將學生表現狀況予以文字描述之說明。

以語文學習領域國文科為例，國中評量標準針對「閱讀能力」主題的「綜合評鑑」次主題，界定出各個表現等級學生應具備的知識與能力，如表 3 所示，圖 2 則呈現對應此表現描述所設計的評量工具，以評量九年級學生是否能指出文本形式並說出依據與理由。表 3 同時呈現可對應表現描述的評分規準，以利教師於施測後能將學生作答結果對照此評分規準以界定其表現等級。圖 3 以 A 等級為例呈現出兩位學生的真實作答情況，可以發現這兩位學生不但回答正確且闡述合理、完整，此作答表現較符合評分規準中 A 等級的描述，因而推論學生可能已養成高層次的的能力，被歸類為 A 等級的機率較高。

表 3

國文科「閱讀能力」主題中「綜合評鑑」次主題之表現描述與評分規準

表現等級	內容
表現描述 A	<ol style="list-style-type: none"> <li>1. 能整合比較文本或不同文本間的重點與細節，並完整提出個人的觀點。</li> <li>2. 能評鑑文本內容 (如：邏輯、論據或實例等) 的適切性，並完整說出其依據及理由。</li> <li>3. 能指出文本形式，評鑑其適切性，並完整說出依據及理由。</li> <li>4. 能指出文本如何反映文化與社會現象。</li> </ol>

(續)

表 3

國文科「閱讀能力」主題中「綜合評鑑」次主題之表現描述與評分規準（續）

表現等級	內容
B	1. 能整合、比較文本或不同文本間的重點與細節，並提出個人的觀點。
	2. 能評鑑文本內容（如：邏輯、論據或實例等）的適切性，並說出依據及理由。
	3. 能指出文本形式，評鑑其適切性，並說出依據及理由。
	4. 能簡單指出文本如何反映文化或社會現象。
C	1. 大体能整合、比較文本或不同文本間的重點，並簡單提出個人的觀點。
	2. 大体能評鑑文本內容（如：邏輯、論據或實例等）的適切性，並大体能說出依據及理由。
	3. 大体能指出或評鑑文本形式，並大体能說出依據及理由。
D	1. 僅能有限地整合、比較文本或不同文本間的重點，並有限地提出個人的觀點。
	2. 僅能有限地評鑑文本內容（如：邏輯、論據或實例等）的適切性，並有限地說出依據及理由。
	3. 僅能有限地指出或評鑑文本形式，並有限地說出依據及理由。
E	未達D級
評分規準	A 答案正確且闡述合理、完整。
	B 答案正確且闡述合理，但略有瑕疵。
	C 答案大致正確且闡述大致合理，但有部分不恰當。
	D 僅能寫出部分的答案。
	E 答案幾乎完全錯誤或完全錯誤，或未作答。

註：引自國民中學學生學習成就評量標準（試行版）—語文學習領域（國文科），宋曜廷、林世華、曾芬蘭，2013a，取自 <http://140.122.106.29/chinese.pdf>

另以社會學習領域之地理科為例，在綜合歸納相關能力指標後，七年級學生在「環境系統」主題上獲得不同表現等級所應展現的知識與能力如表 4 所示，圖 4 與圖 5 則呈現以等高線為素材進而設計研發的評量作業示例。該評量工具包含五個子題、不同子題用來評量不同等級之表現，以第 5 小題為例，該小題旨在評量學生能否展現「利用地理知識與技能探討環境要素的空間分布與交互作用」之 A 等級表現。表 4 同時呈現可對應表現描述的評分規準，例如，A 等級學生在第 1 小題能夠寫出「等高線」名詞，在第 2 小題能正確選擇等高線的特性（A、C、D 至少選出一個，且不得選出 B、E 選項），在第 3 小題能正確判讀等高線表示的地形類型，進而完整地畫出地形剖面圖，並利用對等高線的認識與瞭解，對自己選擇的路線提出合理的說明。在實務上，教師於實施評量後可將學生作答結果對照此評分規準以界定其表現等級。

請閱讀以下詩作，並回答問題：

煤——寫給一九八四年七月煤山礦災死難的六十七名礦工

杜十三

孩子  
我們生命中的色彩  
是注定要從黑色的地層下面 挖出來的  
家裡飯桌上 綠色的菜  
白色的米  
街頭二輪的彩色電影  
媽媽的紅拖鞋  
姊姊的綠色香皂  
還有你的黃色書包  
都是需要阿爸 流汗  
從黑色的洞裡 挖出來的  
今後阿爸不再陪你了  
因為阿爸要到更深 更黑的地方  
再為你 挖出一條  
有藍色天空的路來

阿爸，你不要再騙我了  
家裡面所有的色彩  
其實，都是假的  
我早就知道  
家裡的飯菜是煤做的  
媽媽的笑容姊姊的衣裳  
還有我的課本和鉛筆……  
統統都是煤做的  
甚至連您啊 我想念的阿爸  
不也是煤做的嗎？  
他們說：煤不再值錢了  
可是 阿爸  
我卻寧願丟掉所有的色彩  
陪著媽媽 姊姊  
守在洞口  
拚命的用眼睛去挖 去挖  
挖出一具  
黑色的  
阿  
爸

題目

詩是精練的文字，下列兩段詩句若刪去「」中的語詞，文意依然完整。請就寫作效果來看，保留「」中語詞的作用是什麼？

- (1)我們生命中的色彩／是「注定」要從黑色的地層下面 挖出來的
- (2)都是需要阿爸 「流汗」／從黑色的洞裡 挖出來的

圖2. 國文科評量作業示例。引自國民中學學生學習成就評量標準(試行版)－語文學習領域(國文科)，宋曜廷、林世華、曾芬蘭，2013a，取自<http://140.122.106.29/chinese.pdf>

圖4與圖5呈現出兩位獲得A等級學生的真實作答情況，可以發現不論其選擇哪一條路線建造鐵路，都能夠歸納該路線之優點，甚至比較不同路線之優缺點，最後清楚地闡述選擇特定路線之理由，最終被歸類為A等級的機率較高。

從上述兩個國文與社會的範例說明可知，評分規準植基於評量標準，進一步具體描寫學生於特定學習內容上的不同等級表現；作業示例可個別呈現出不同等級學生的典型表現，甚至呈現屬於相同等級(如A等級)學生的多種不同作答樣態，這些措施皆有助於促進教師對

表現示例一	<p>(1) 為了要強調在礦區工作非常辛苦,但是已經注定了,我們無奈,也無法改變</p> <p>_____</p> <p>(2) 強調在礦區工作的辛苦,以及因為流了許多汗努力工作才讓一家人得到溫飽,所以珍惜</p> <p>_____</p>
	<p>樣卷說明： 答案正確且闡述合理、完整。</p>
表現示例二	<p>(1) 加強了這個句子的語氣和無耐感。讓人感覺父親好像是很認命的心態。</p> <p>_____</p> <p>(2) 用「流汗」把阿爸在礦坑工作的情形具體化,不只代表阿爸的辛苦,還有阿爸在悶熱的礦坑中挖礦所流下的汗水。</p> <p>_____</p>
	<p>樣卷說明： 答案正確且闡述合理、完整。 註：答案中有錯字，「無耐」應為「無奈」。</p>

圖3. 國文科學生表現示例。引自國民中學學生學習成就評量標準(試行版)－語文學習領域(國文科)，宋曜廷、林世華、曾芬蘭，2013a，取自<http://140.122.106.29/chinese.pdf>

表 4

地理科「環境系統」主題之表現描述與評分規準

內容標準	主題	環境系統
表現描述	A	1. 利用地理知識與技能分析自然環境要素的分布特性與運作。 2. 利用地理知識與技能分析人文環境要素的分布特性與運作。 3. 利用地理知識與技能探討環境要素的空間分布與交互作用。
	B	1. 瞭解自然環境要素的分布特性與運作。 2. 瞭解人文環境要素的分布特性與運作。 3. 利用地理知識與技能瞭解環境要素的空間分布與交互作用。
	C	1. 認識環境系統重要名詞的意義。 2. 認識自然與人文環境要素的分布。 3. 認識用以呈現地表現象位置及分布的地圖、數據及其他資訊。
	D	知道自然環境要素(地形、氣候、水文)與人文環境要素(人口、經濟活動、交通與聚落)的重要名詞。
	E	未達D級

(續)

表 4

地理科「環境系統」主題之表現描述與評分規準（續）

內容標準	主題	環境系統
評分規準	A	寫出「等高線」名詞，選出其特性的正確選項（A、C、D至少選出一個，且不得選出B、E選項），正確判讀等高線表示的地形類型，並完整地畫出地形剖面圖，利用對等高線的認識與瞭解，對自己選擇的路線提出合理的說明。
	B	寫出「等高線」名詞，選出其特性的正確選項（A、C、D至少選出一個，且不得選出B、E選項），正確判讀等高線表示的地形類型，並大致完成地形剖面圖。
	C	寫出「等高線」名詞，並選出其特性的正確選項（A、C、D至少選出一個，且不得選出B、E選項）。
	D	寫出「等高線」名詞。
	E	未達D級

註：引自國民中學學生學習成就評量標準（試行版）－社會學習領域（地理科），宋曜廷、林世華、曾芬蘭，2013b，取自 <http://140.122.106.29/geography.pdf>

於評量標準的理解，提升了評量標準於班級中施行的可行性。整體而言，一旦評量工具的設計以及評分規準的撰寫能有效地對應到評量標準，教師便能於評量後將學生作答結果對應到評分規準去初步界定學生表現等級，最終從評量標準中瞭解該等級學生可能具備的學習成就。

### （三）應用方式與預期效益

將評量標準應用在班級評量中（如圖 6 所示）將具有以下效益：第一，促進教師專業發展。在教學前，教師根據評量標準擬訂教學目標、設計教材與評量工具；教學中，根據評量標準揭示的不同層次教學目標引導學生學習；教學後，利用事前設計的多元評量工具蒐集學生資料，並將學生表現與評分規準加以對應，以決定學生的表現等級。在上述歷程中皆需要不斷地與評量標準進行對照，因此不但可加深教師對評量標準的認識，教師也需要發展多面向的教材（而非僅著重認知面向），以服膺評量標準所規範的多面向學習內容。此外，需設計適當的多元評量工具（而非僅使用紙筆測驗），以利教學與評量能有效呼應，教師教學與評量專業自然提升。第二，實現課綱領導教學之理念。由於評量標準根據能力指標整併而來，因此同樣具備課綱對學習內容的規範作用，教師根據評量標準設立教學目標並設計教學活動，等同於教師根據課綱進行教學。第三，提供診斷與補救之回饋訊息、達成確保學習品質之理念。透過比對學生作業表現與評分規準可以界定出學生於特定學習內容之表現等級，此舉一方面可瞭解學生表現有多好，藉此評估學生高層次能力的培養成效；另一方面幫助瞭解學生表現不佳的程度，有多少學生仍未達到通過基礎門檻的程度，以利教師進行補救教學並調整教學策略，也就是說 D 級與 E 級的評分規準，可以作為學習落後學生的診斷與補救指標，並據以作為教學計畫與教學實務改善的參考依據，達成評量回饋教學之用。

圖(一)是一種用來表現地表高低起伏的地形示意圖，請問：

1.圖(一)中的曲線稱為：等高線。

2.圖(一)中的曲線具有下列哪些特性？(複選)  
 (A)曲線呈閉合狀  
 (B)曲線愈長高度愈高  
 (C)同一條曲線上的高度都相同  
 (D)兩條相鄰曲線間的高度差相等  
 (E)曲線愈密坡度愈緩，愈疏坡度愈陡  
 答：A, C, D。

3.在臺灣，圖(一)表示的地形類型稱為：丘陵。

4.完成圖(一)中甲點至乙點的地形剖面圖。  
 答：(請直接繪製在圖(一)中)

5.若要建造一條由甲點至乙點的鐵路，就地形條件考量，圖(二)中A、B、C三條路線，你會選擇哪一條路線？請說明理由。  
 答：C，因為在平坦的地方蓋金載路較為容易，且路線平直也較好建造。

圖(一)

圖(二)

樣卷說明：  
 Q1.寫出正確名詞「等高線」。  
 Q2.完整選出 A、C、D 選項為等高線的特性。  
 Q3.正確判讀等高線表示的地形類型為「丘陵」。  
 Q4.完整且正確地畫出地形剖面圖。  
 1.標示出甲乙剖面線和等高線的交點。  
 2.將每一個交點對應到地形剖面圖上。  
 3.將對應到的點用圓滑的曲線連結起來。  
 4.合理繪製兩座丘陵連接的谷地(100~200公尺之間)。  
 5.合理繪製丘陵頂部(500~600公尺之間)。  
 Q5.由地形條件說明「平坦」且「路線平直」為「較好建造」鐵路的理由。

圖4. 地理科「環境系統」主題之評量作業示例之一。引自國民中學學生學習成就評量標準(試行版)－社會學習領域(地理科)，宋曜廷、林世華、曾芬蘭，2013b，取自 <http://140.122.106.29/geography.pdf>

## 陸、結論與建議

十二年國民基本教育已正式啟動，入學與考試制度因此產生重大變革。是否能達到「低一點的升學壓力、高一點的學習彈性、活一點的教學方式、多一點的社會正義」，可能是判斷入學與考試制度的變革能否成功的重要依據。標準本位評量是入學與考試制度的一環，如果能有效實施，將能對前述的目標有所貢獻：透過評比級距減少，讓升學壓力低一點；透過升學壓力的降低，增進學生的學習彈性，使其勇於探索並思考教科書以外的世界；透過評量標

圖(一)是一種用來表現地表高低起伏的地形示意圖，請問：

1. 圖(一)中的曲線稱為：等高線。

2. 圖(一)中的曲線具有下列哪些特性？  
(複選)  
(A) 曲線呈閉合狀  
(B) 曲線愈長高度愈高  
(C) 同一條曲線上的高度都相同  
(D) 兩條相鄰曲線間的高度差相等  
(E) 曲線愈密坡度愈緩，愈疏坡度愈陡  
答：ACD。

3. 在臺灣，圖(一)表示的地形類型稱為：丘陵。

4. 完成圖(一)中甲點至乙點的地形剖面圖。  
答：(請直接繪製在圖(一)中)

5. 若要建造一條由甲點至乙點的鐵路，就地形條件考量，圖(二)中 A、B、C 三條路線，你會選擇哪一條路線？請說明理由。  
答：A，因為B點有高低上的落差，而且有一段陡峭的山壁，因此可能必須建造之字形的鐵路，可能有費用上的問題；而A與C雖在同一高度區域之內，但是丘陵的地形中人口大多集中於山谷，所以必有經過山谷的A較佳。

樣卷說明：  
Q1. 寫出正確名詞「等高線」。  
Q2. 完整選出 A、C、D 選項為等高線的特性。  
Q3. 正確判讀等高線表示的地形類型為「丘陵」。  
Q4. 完整且正確地畫出地形剖面圖。  
1. 標示出甲乙剖面線和等高線的交點。  
2. 將每一個交點對應到地形剖面圖上。  
3. 將對應到的點用圓滑的曲線連結起來。  
4. 合理繪製兩座丘陵連接的谷地(100~200公尺之間)。  
5. 合理繪製丘陵頂部(500~600公尺之間)。  
Q5. 由地形條件分析 B 路線較為陡峭，建造鐵路會有費用問題，並將 A 路線通過谷地的人口因素納入考慮，推論尚屬合理。

圖5. 地理科「環境系統」主題之評量作業示例之二。引自國民中學學生學習成就評量標準(試行版)－社會學習領域(地理科)，宋曜廷、林世華、曾芬蘭，2013b，取自<http://140.122.106.29/geography.pdf>

準的理解與實做，讓教師瞭解開拓學生高層次能力的重要性；透過標準參照評量，讓各級政府正視學習成就落差的事實，投入補救的措施，以提升社會正義。

目前標準本位評量已正式進入施行階段，未來還有很多實務上的問題需要克服。若期望入學和考試制度的變革得以成功，以下幾點建議值得特別關注。

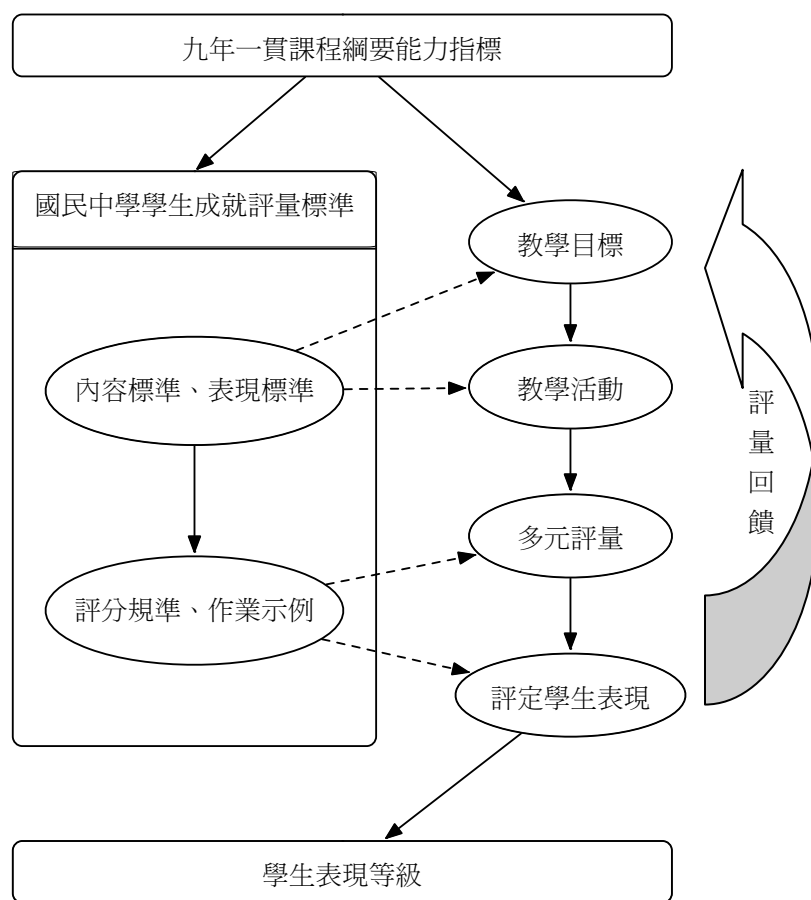


圖6. 在班級評量中應用評量標準示意。國民中學學生學習成就評量標準（試行版）－研發說明與常見問題集，宋曜廷、林世華、曾芬蘭，2013c。臺北市：作者。

### 一、強化觀念的宣導與改變

家長、學生和教師對於升學相關概念的改變，是決定入學與考試制度的變革能否成功的最重要關鍵。家長應該瞭解子女未來的競爭力未必在學科能力的表現，其他能力如獨立思考、創造力、合作溝通等，都是成就事業的重要能力，因此應容許自己的子女在課業外，有時間和精力培養這些能力。傳統明星高中也應有所覺醒，並非在入學考試獲得 RP 值 99 的學生才是理想的招收對象，具有適當的學業表現（如 PR 值 95），且有學科能力之外的適當素養，可能成就更好的未來領袖。有此覺醒，將能帶動國中端的學生和家長不再分分計較，不再日日小考，也讓其他有心適性發展的學生有喘息的空間（宋曜廷，2012b）。這些觀念的改變，有賴持續性的宣導。

## 二、會考成績應用方式可考慮因區因時制宜

目前會考成績的運用，因為若干升學競爭激烈考區的需求，在原有等級加上四個標示。未來教育主管機關可視全國各區實施免試入學的成效，讓現有會考成績加註標示制度可以因區制宜，在升學競爭壓力較低的區域回歸原始設計的三等級而不加註標示，甚至因時制宜，當各區域升學壓力皆降低至相當程度時，可全部回歸三等級，讓成績呈現方式與原本作為學力監控機制之目的相互呼應。

## 三、評量標準的運用應加強與教學的結合

前已述及，評量標準具備有將標準、教學和評量相互鏈結，彼此參考的效用。目前因執行時程的緣故，在將課程標準和評量機制相連結上已有成效，但如何將評量標準和班級教學相結合，真正達到以評量標準引導並回饋教學與學習的目的，是後續重要的努力方向。

## 四、制度性的配套

政府應考慮教師在班級中實施標準本位評量的時間，每位教師所需負責的班級數或授課時數應考慮減少，以減輕教師負擔。另政府也可考慮設立助理教師來協助，並建議由學校成立專業諮詢團隊，定期考評與回饋協助學校。

## 誌謝

本研究獲科技部「跨國頂尖研究中心－學習科學跨國頂尖研究中心」計畫補助（計畫編號：NSC103-2911-I-003-301），特致謝忱。

## 參考文獻

### 一、中文部分

- 吳清山、高家斌 (2007)。臺灣中等教育改革分析：1994-2007 年。教育資料集刊，34，1-24。  
【Wu, C.-S., & Kao, C.-P. (2007). The analysis on the reform of secondary education in Taiwan: 1994-2007. *Bulletin of National Institute of Educational Resources and Research*, 34, 1-24.】
- 宋曜廷 (2012a)。以標準參照的入學考試和班級評量促進科教發展。科學月刊，43(9)，672-678。  
【Sung, Y.-T. (2012a). Standards-referenced entrance examinations and classroom assessment for the promotion of science education. *Science Monthly*, 43(9), 672-678.】
- 宋曜廷 (2012b，9 月 17 日)。十二年國教的壓力與競爭力。中國時報，A12 版。  
【Sung, Y.-T. (2012b, September 17). The pressure and competitiveness of 12-year basic education in Taiwan. *The Chinatimes*, p. A12.】
- 宋曜廷、周業太、吳佩瓊、林秀珊、曾芬蘭 (2010)。從學校本位評量省思臺灣擴大免試入學方案。教育科學研究期刊，55 (2)，73-113。doi:10.3966/2073753X2010065502003  
【Sung, Y.-T., Chou, Y.-T., Wu, P.-Y., Lin, H.-S., & Tseng, F.-L. (2010). A reflection of school-based assessment on the extended open admission program in Taiwan. *Journal of Research in Education Sciences*, 55(2), 73-113. doi:10.3966/2073753X2010065502003】
- 宋曜廷、林世華、曾芬蘭 (2013a)。國民中學學生學習成就評量標準 (試行版) — 語文學習領域 (國文科)。取自 <http://140.122.106.29/chinese.pdf>  
【Sung, Y.-T., Lin, S.-H., & Tseng, F.-L. (2013a). *Standards-based assessment of student achievement handbook in Chinese language arts for junior high school*. Retrieved from <http://140.122.106.29/chinese.pdf>】
- 宋曜廷、林世華、曾芬蘭 (2013b)。國民中學學生學習成就評量標準 (試行版) — 社會學習領域 (地理科)。取自 <http://140.122.106.29/geography.pdf>  
【Sung, Y.-T., Lin, S.-H., & Tseng, F.-L. (2013b). *Standards-based assessment of student achievement handbook in social studies of geography for junior high school*. Retrieved from <http://140.122.106.29/geography.pdf>】
- 宋曜廷、林世華、曾芬蘭 (2013c)。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。臺北市：作者。  
【Sung, Y.-T., Lin, S.-H., & Tseng, F.-L. (2013c). *Standards-based assessment of student achievement handbook for junior high school*. Taipei, Taiwan: Author.】
- 宋曜廷、邱佳民、張恬熒、曾芬蘭 (2011)。以國中基本學力測驗成績探討學習成就落差。教育政策論壇，14 (1)，85-117。  
【Sung, Y.-T., Chiou, J.-M., Chang, T.-Y., & Tseng, F.-L. (2011). Investigating learning achievement gap through students' basic competence test scores. *Educational Policy Forum*, 14(1), 85-117.】
- 宋曜廷、許福元、曾芬蘭、蔣莉蘋、孫維民 (2007)。國民中學學生基本學力測驗的回顧與展望。教育研究與發展期刊，3 (4)，29-50。  
【Sung, Y.-T., Hsu, F.-Y., Tseng, F.-L., Chiang, L.-P., & Sun, W.-M. (2007). Basic competence test: A review and prospect. *Journal of Educational Research and Development*, 3(4), 29-50.】

宋曜廷、曾芬蘭、邱佳民（2013）。國中教育會考各科標準設定說明。載於國立臺灣師範大學心理與教育測驗研究發展中心舉辦之「**2013 第二屆標準本位評量國際研討會**」論文集（pp. 143-152），臺北市。

【Sung, Y.-T., Tseng, F.-L., & Chiou, J.-M. (2013). An explanation of the standard settings for each subject for the comprehensive assessment program for junior high school students. In Research Center for Psychological and Education Testing, National Taiwan Normal University (Ed.), "2013 International Conference of Standards-Based Assessment" symposium (pp. 143-152), Taipei, Taiwan.】

宋曜廷、趙子揚、王雅鈴、黃瓏瑩、陳佳蓉、曾芬蘭（2013）。「中學生考試壓力量表」之編製及其信度與效度之分析。測驗學刊，**60**（2），291-318。

【Sung, Y.-T., Chao, T.-Y., Wang, Y.-L., Huang, L.-Y., Chen, C.-J., & Tseng, F.-L. (2013). The development of examination stress scale for junior high school students. *Psychological Testing*, 60(2), 291-318.】

香港考試及評核局（2011）。香港中學文憑考試－評級程序與水平參照成績匯報。取自 [http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE\\_SRR\\_A4\\_Booklet\\_Jun2011.pdf](http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE_SRR_A4_Booklet_Jun2011.pdf)

【Hong Kong Examinations and Assessment Authority. (2011). *Hong Kong diploma of secondary education Examination—Grading procedures and standards-referenced reporting booklet*. Retrieved from [http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE\\_SRR\\_A4\\_Booklet\\_Jun2011.pdf](http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE_SRR_A4_Booklet_Jun2011.pdf)】

香港考試及評核局（2013）。香港中學文憑考試。取自 [http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE\\_pamphlet\\_Chi\\_2013.pdf](http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE_pamphlet_Chi_2013.pdf)

【Hong Kong Examinations and Assessment Authority. (2013). *Hong Kong diploma of secondary education (HKDSE) examination booklet*. Retrieved from [http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE\\_pamphlet\\_Chi\\_2013.pdf](http://www.hkeaa.edu.hk/DocLibrary/Media/Leaflets/HKDSE_pamphlet_Chi_2013.pdf)】

國立臺灣師範大學心理與教育測驗研究發展中心（2013）。**103 年國中教育會考問與答手冊**。取自 <http://cap.ntnu.edu.tw/documents/103capmanual.pdf>

【Research Center for Psychological and Education Testing, National Taiwan Normal University. (2013). *The manual of the 2014 comprehensive assessment program for junior high school students*. Retrieved from <http://cap.ntnu.edu.tw/documents/103capmanual.pdf>】

張道行、陳清誥、徐慧萍、許福元（2007）。臺灣高中入學指標應用之比較與分析。載於國立臺灣師範大學心理與教育測驗研究發展中心舉辦之「**2007 入學考試與制度國際研討會**」論文集（pp. 79-90），臺北市。

【Chang, T.-H., Chen, C.-K., Hsu, H.-P., & Hsu, F.-Y. (2007). The development and analysis of senior high school admission criteria in Taiwan. In Research Center for Psychological and Education Testing, National Taiwan Normal University (Ed.), "2007 International Conference on Entrance Examination and Admission Policy" symposium (pp. 79-90), Taipei, Taiwan.】

教育部（2012）。國民小學及國民中學學生成績評量準則修正條文。取自 <http://host.lyjh.tyc.edu.tw/~reg/download/1010525.doc>

【Ministry of Education. (2012). *Amendment of guidelines for the grading assessment of elementary and junior high school students*. Retrieved from <http://host.lyjh.tyc.edu.tw/~reg/download/1010525.doc>】

教育部（2013a）。高中高職及五專免試入學實施方案。取自 [http://12basic.edu.tw/File/Levelimg\\_228/1-1.doc](http://12basic.edu.tw/File/Levelimg_228/1-1.doc)

【Ministry of Education. (2013a). *The vocational school, high school, and junior college entrance examination*

- implementation program*. Retrieved from [http://12basic.edu.tw/File/Levelimg\\_228/1-1.doc](http://12basic.edu.tw/File/Levelimg_228/1-1.doc)】
- 教育部 (2013b)。高中高職及五專特色招生實施方案。取自 [http://12basic.edu.tw/File/Levelimg\\_228/1-2.doc](http://12basic.edu.tw/File/Levelimg_228/1-2.doc)
- 【Ministry of Education. (2013b). *Vocational school, high school, and junior college academic ability test implementation program*. Retrieved from [http://12basic.edu.tw/File/Levelimg\\_228/1-2.doc](http://12basic.edu.tw/File/Levelimg_228/1-2.doc)】
- 詹紹威 (2012)。美國共同核心州課程標準倡議之探討。《教育研究與發展期刊》，8(2)，183-202。
- 【Chan, S.-W. (2012). An analysis on common core state standards initiative of the United States of America. *Journal of Educational Research and Development*, 8(2), 183-202.】
- 譚光鼎、莊勝義、魯先華、康瀚文、陳怡璇 (2007)。台灣高級中學多元入學政策之檢討。載於國立臺灣師範大學教育學系舉辦之「公義社會與教育行政革新國際學術研討會」論文集 (pp. 27-64)，臺北市。
- 【Tan, K.-T., Chuang, S.-Y., Lu, H.-H., Kang, H.-W., & Chen, Y.-H. (2007). Reappraising the Policy of Multi-channel School Admission System for Taiwan Senior High Schools. In Department of Education, National Taiwan Normal University (Ed.), *Justice Society and Administrative Initiatives in Education International Conference* symposium (pp. 27-64), Taipei, Taiwan.】

## 二、外文部分

- Australian Curriculum, Assessment and Reporting Authority. (2013). *2013 NAPLAN national report*. Retrieved from <http://www.nap.edu.au/results-and-reports/national-reports.html>
- Common Core State Standards Initiative. (2011). *Frequently asked questions*. Retrieved from <http://www.corestandards.org/frequently-asked-questions>
- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 89-116). Mahwah, NJ: Lawrence Erlbaum Associates.
- Illinois State Board of Education. (2010). *Illinois Standards Achievement Test—2010 Technical Manual*. Retrieved from [http://www.isbe.net/assessment/pdfs/isat\\_tech\\_2010.pdf](http://www.isbe.net/assessment/pdfs/isat_tech_2010.pdf)
- Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3-16. doi:10.3102/0013189X031006003
- Linn, R. L., & Herman, J. L. (1997). *A policymaker's guide to standards-led assessment*. Denver, CO: Education Commission of the States.
- Llosa, L. (2007). Validating a standards-based classroom assessment of English proficiency: A multitrait-multimethod approach. *Language Testing*, 24(4), 489-515. doi:10.1177/0265532207080770
- Llosa, L. (2008). Building and supporting a validity argument for a standards-based classroom assessment of English proficiency based on teacher judgments. *Educational Measurement:*

- Issues and Practice*, 27(3), 32-42. doi:10.1111/j.1745-3992.2008.00126.x
- Llosa, L. (2011). Standards-based classroom assessments of English proficiency: A review of issues, current developments, and future directions for research. *Language Testing*, 28(3), 367-382. doi:10.1177/0265532211404188
- Lyon, E. (2011). Beliefs, practices, and reflection: Exploring a science teacher's classroom assessment through the assessment triangle model. *Journal of Science Teacher Education*, 22(5), 417-435. doi:10.1007/s10972-011-9241-4
- Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79(4), 1332-1361. doi:10.3102/0034654309341375
- Reys, R., Reys, B., Lapan, R., Holliday, G., & Wasman, D. (2003). Assessing the impact of "standards"-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education*, 34(1), 74-95. doi:10.2307/30034700
- Sung, Y.-T., Chang, K.-E., Chiou, S.-K., & Hou, H.-T. (2005). The design and application of a Web-based self- and peer-assessment system. *Computers and Education*, 45(2), 187-202. doi:10.1016/j.compedu.2004.07.002
- Sung, Y.-T., Tseng, F.-L., Kuo, N.-P., Chang, T.-Y., & Chiou, J.-M. (2014). Evaluating the effects of programs for reducing achievement gaps: A case study in Taiwan. *Asia Pacific Education Review*, 15(1), 99-113. doi:10.1007/s12564-013-9304-7
- Vermont Department of Education. (2012). *Using standards in your classroom: A teacher resource guide*. Retrieved from <http://www.wssu.k12.vt.us/teachers/docs/UsingStandards.pdf>

## 附錄一 香港中學文憑考試之共通等級描述

表現等級	共通等級描述
第5級	<ul style="list-style-type: none"> <li>對課程內容有廣泛的認識和透徹的理解，能把概念和技巧有效地應用到多元和複雜的不熟悉情境，並顯示深入的見解</li> <li>能分析、綜合和評價廣泛的資料</li> <li>能精簡及邏輯地傳達意念和見解</li> </ul>
第4級	<ul style="list-style-type: none"> <li>對課程內容有良好的認識和理解，能把概念和技巧有效地應用到不熟悉的情境，並顯示深入的見解</li> <li>能分析、綜合和詮釋各種資料</li> <li>能邏輯地傳達意念和見解</li> </ul>
第3級	<ul style="list-style-type: none"> <li>對課程內容有足夠的認識和理解，能把概念和技巧適當地應用到不同的熟悉情境</li> <li>能分析和詮釋各種資料</li> <li>能恰當地傳達意念和見解</li> </ul>
第2級	<ul style="list-style-type: none"> <li>對課程內容有基本的認識和理解，能把概念和技巧應用到熟悉的情境</li> <li>能辨識和詮釋直接的資料</li> <li>能平衡地傳達簡單意念</li> </ul>
第1級	<ul style="list-style-type: none"> <li>對課程內容有初步的認識和理解，在協助下，能把概念和技巧應用到簡單熟悉的情境</li> <li>在指導下，能辨識和詮釋簡單的資料</li> <li>能粗略地傳達簡單意念</li> </ul>

## 附錄二 美國伊利諾州標準成就測驗之等級描述

表現等級	等級描述
超越標準	學生作品展現學科上優異的知識與技能，學生能創造性地應用知識和技能、解決問題和評估結果。
符合標準	學生作品展現學科上熟練的知識和技能，學生能透過運用所學知識和技能，有效地解決問題。
低於標準	學生作品展現學科上基本的知識和技能，但由於在學習上的落差，學生僅能有限地運用知識和技能。
學習警告	學生作品展現學科上有限的知識和技能，但由於在學習上的巨大落差，學生無法有效地運用知識和技能。

### 附錄三 國中教育會考各考科之表現等級描述

科目	等級		
	精熟	基礎	待加強
國文	能具備與教材相關的語文知識，並能深入的理解文本內容、評鑑文本的內容與形式。	大致能具備與教材相關的語文知識，並能大致理解文本內容、評鑑文本的內容與形式。	僅能具備部分與教材相關的語文知識，並有限的理解文本內容、評鑑文本的內容與形式。
聽力	能聽懂主題熟悉、訊息稍為複雜、段落較長的言談，指出言談的主旨與結論等重要訊息，並從言談中言語及其他如語調與節奏等線索做出推論；能理解短片及廣播節目的大意。	能聽懂日常生活主題、訊息單純的短篇言談，指出言談的主旨與結論等重要訊息，並從言談中明顯的言語及其他如語調與節奏等線索做出簡易推論。	僅能聽懂單句及簡易問答；僅能有限的理解短篇言談。
英語 閱讀	能整合應用字詞、語法結構及語用慣例等多項語言知識；能理解主題較為抽象、訊息或情境多元複雜、語句結構長且複雜的文本，指出各類文本的主旨、結論與作者立場等重要訊息，並從文本結構、解釋或例子等做進一步的推論或評論。	能理解字詞基本語意及語法概念；能理解主題具體或貼近日常生活、訊息或情境略為複雜、語句結構略長的文本，指出文本主旨、結論與作者立場等重要訊息，並從文本的解釋或例子做出推論。	僅能理解主題貼近日常生活、訊息或情境單純且明顯、語句結構簡單的文本或語句，僅能指出文本明白陳述的主旨、結論與作者立場等重要訊息，僅能藉文本明顯的線索做出簡易的推論。
數學	能作數學概念間的連結，建立恰當的數學方法或模式解題，並能論證。	理解基本的數學概念、能操作算則或程序，並應用所學解題。	認識基本的數學概念，僅能操作簡易算則或程序。
社會	能廣泛且深入的認識及瞭解社會科學習內容，並具有運用多元的社會科知識之能力。	能大致認識及瞭解社會科學習內容，並具有運用基礎的社會科知識之能力。	能約略的認識及瞭解社會科學習內容。
自然	能融會貫通學習內容，並能運用所培養的能力來解決需要多層次思考的問題。	能知道及理解學習內容，並能運用所培養的能力來解決基本的問題。	能部分知道及理解學習內容。

Journal of Research in Education Sciences

2014, 59(1), 1-32

doi:10.6209/JORIES.2014.59(1).01

# Standards-Based Assessments for 12-Year Basic Education in Taiwan

Yao-Ting Sung

Research Center for Psychological and Educational Testing, National Taiwan Normal University

Yeh-Tai Chou

Research Center for Psychological and Educational Testing, National Taiwan Normal University

Feng-Lan Tseng

Research Center for Psychological and Educational Testing, National Taiwan Normal University

## Abstract

Taiwan officially implemented 12-Year Basic Education. Entrance examinations and classroom assessments are one of the keys to the success of this new program. Because thousands of Taiwanese children will be affected, how to help the program succeed deserves scrutiny. In recent years, many developed nations have used standards-based assessment to enhance their educational programs. Standards-based assessment, therefore, could potentially be advantageous for Taiwan as well. In this study, the implications of broadly applying standards-based assessment were investigated by exploring its effect on knowledge diversity, scholastic ability, and academic pressure among junior high school students.

**Keywords:** 12-Year Basic Education, classroom assessment, comprehensive assessment program for junior high school students, standards-based assessment of student achievement, standards-based assessment