

基本能力評量跨國 發展經驗之比較研究

*Comparing the Experiences of
Developing Basic Student Performance
Assessment in Different countries*



基本能力評量跨國發展經驗之比較研究

計畫主持人：王世英（國立教育資料館館長）

研究主持人：張鈿富（國立暨南國際大學教育政策與行政學系教授）

協同主持人：吳慧子（國立暨南國際大學教育政策與行政學系助理教授）

研 究 員：梅瑤芳（國立教育資料館秘書）

王秉倫（國立教育資料館視聽組主任）

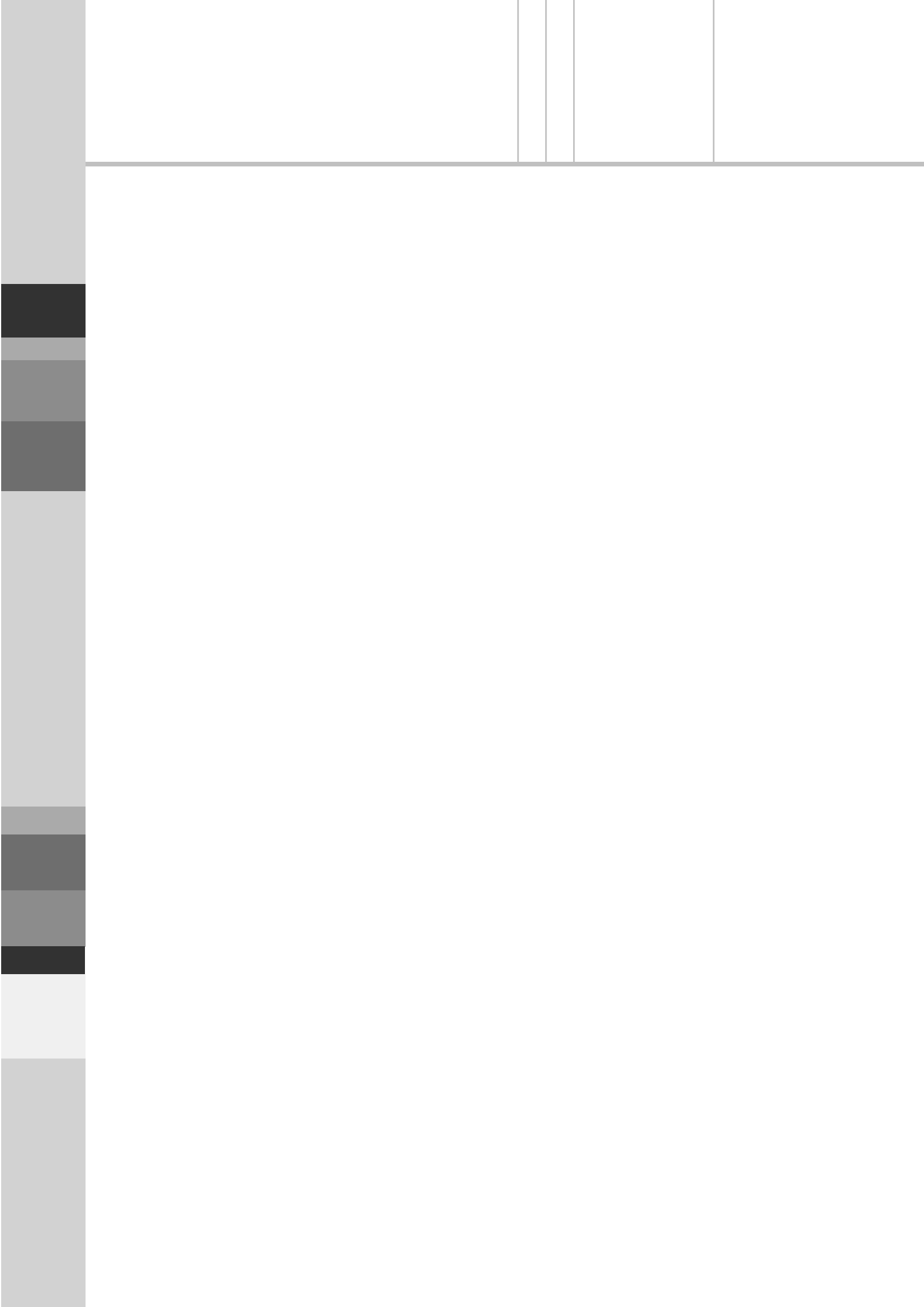
張雲龍（國立教育資料館編輯）

周文菁（國立暨南國際大學教育政策與行政學系碩士）

研 究 助 理：黃奕通（國立暨南國際大學教育政策與行政學系研究生）

潘佳汝（國立暨南國際大學教育政策與行政學系研究生）

中華民國九十五年五月



摘要

本研究透過文獻分析，以比較研究為主。首先，蒐集各主要國家有關學生基本能力評量的文獻，「各主要國家」以英、美、OECD 等主要國家或組織為主，就目前已發展的學生基本能力評量為文獻蒐集的範圍。其次，就蒐集的文獻進行比較，比較各主要國家學生基本能力評量的內涵，並分析這些內涵的代表意義，以做為我國發展類似能力評量的參考。初步研究成果提出後並進行專家的諮詢，以確保研究結果的正確性。最後，提出幾點建議，以做為教育部相關政策訂定之參考。

第一，評量機構方面：1.慎選發展評量的機構，做長期性、系統性的研究發展；2.採取以校際或是機構分工的設計模式，結合學者專家的專長，可以快速達到分工整合的功效；第二，評量內涵方面：1.審慎參考國際的評量架構與評量內涵，做為發展國內的評量系統之參考；2.打破以教科書本為範疇的評量型態，以學習技能及生活技能為評量學生表現的重點；3.建立長期性評量題庫的建置、發展單位；第三，評量實施方面：1.初期先以抽樣的方式進行試驗，以建立基本的評量內涵；2.最終以建立全國性的學生評量資料庫為目標；第四，評量結果的呈現：1.可以顯現全國性學生表現的資訊；2.可以呈現區域比較及校際比較的資訊；3.評量結果的呈現具有時間序列的意義。

關鍵詞彙：基本能力評量、美國學生基本能力評量、英國學生基本能力評量、OECD 基本能力評量、NAEP、GCSE、PISA

Comparing the Experiences of Developing Basic Student Performance Assessment in Different countries

ABSTRACT

This research focused on the assessment of students' performance from cross-country experiences by analyzing literature and relying on comparison. First of all, this research collected relevant documents that related to assessment of students' performance in several major nations, such as United Kingdom, America, and other OECD countries. Secondly, the pertinent literatures and documents were compared, and it was to find out what are major nations' practice and experiences of the assessment on students' basic abilities. In addition, educational experts had reviewed the findings of this research to ensure the accuracy of results. Finally, this research provided some suggestions to serve as the reference for Ministry of Education when making educational policies.

First, assessment institution: 1. to select the institution carefully and the institution will be responsible for developing the assessment of students' performance and for studying critical long-term, systematic research; 2. to adopt the cooperative model between schools or organizations, and will efficiently combine various experts' professional specialties. Second, assessment content: 1. to consider deliberately the framework and content of the international assessment which may be served as the reference of developing students' assessment system in Taiwan; 2. to focus on learning basic skills for living instead of the traditional belief that only focused on textbook; 3. to develop the long-term assessment database. Third, assessment practice: 1. to randomly select sample receiving the assessment in initial stage in order to develop the valid assessment; 2. to build the national database of the assessment ultimately. Fourth, assessment results: 1. to demonstrate the information and performance of nationwide students; 2. to present the result of the comparison between the regions or schools; 3. to illustrate if the performance changes over time.

Key Words: student assessment, NAEP, GCES, OECD/PISA

目次

第一章 緒論	01
第一節 研究緣起	01
第二節 研究設計	02
第三節 研究進度與人力	03
第二章 美國學生基本能力評量	05
第一節 NAEP 規劃	05
第二節 NAEP 閱讀能力評量	14
第三節 NAEP 數學能力評量	17
第四節 NAEP 科學能力評量	21
第三章 英國學生基本能力評量	25
第一節 英國之國定課程與評量	25
第二節 英國第一關鍵期評量	31
第三節 英國第二關鍵期評量	34
第四節 英國第三關鍵期評量	37
第四章 OECD/PISA 基本能力評量	43
第一節 OECD/PISA 內涵	43
第二節 PISA 閱讀能力評量	45
第三節 PISA 數學能力評量	50
第四節 PISA 科學能力評量	57

第五章 實施策略分析與比較	63
第一節 美國學生評量的實施策略	63
第二節 英國學生評量的實施策略	64
第三節 OECD/PISA 的實施策略	66
第六章 結論與建議	69
第一節 結 論	69
第二節 建 議	73
參考文獻	75

第一章 緒論

學習品質的確保是近年來各先進國家教育發展的重點，爲了確保學校教育的品質，各國紛紛進行了「基本能力」或是「基本素養」的評量。

爲因應國際教育發展的趨勢，未來我國的學生基本能力評量應何去何從？因此，進行跨國性學生基本能力評量之比較研究有其必要性。如何展開系統性的研究、掌握最新的發展趨勢，以做爲我國發展類似評量的參考，爲本研究發展的重點。

第一節 研究緣起

爲了教育品質，各地區或是各國都有因應的策略。例如，美國評量學生成就的代表爲 NAEP 所發展，是由教育測驗服務社（ETS）所發展的聯邦補助計畫。NAEP 採用大量的、有代表性的學生樣本去建立學生學習成就的趨勢。此一評量的結構和目標經由課程專家、教師、家長、商業領導者和政策制定者的共識發展而成，反映了學生應該知道和可以做的廣泛能力。成就評量的焦點放在 9、13、17 歲的學生。

英國亦把學習評量納入國家課程發展的一環。這些學校外部的評量設計都以建立標準（standard）或是標竿（benchmark）爲主，以確保學校教育系統品質。

歐盟推動教育「品質保證」（quality assurance）機制，其中教育結果的部分以「經濟與合作發展組織」（OECD）所發展的 PISA（Program for International Student Assessment）評量結果爲主，來了解各參與國家 15 歲學生的學習表現。

一、研究目的

由於學生素質確保的重要性，而學習評量是維繫教育品質的一個重要手段。尤其是學生基本能力的評量，與未來公民的素質、從事經濟發展所需的人力素質息息相關。在此一前提下，本研究以國際發展這類評量的趨勢為主，參考各國的經驗與特點，以收他山之石之效。基於此一前提，本研究探討之主要目的如下：

- 1.比較各主要國家學生基本能力評量的內涵。
- 2.探討各主要國家推動學生基本能力評量的策略。
- 3.分析各主要國家學生基本能力評量的呈現方式。
- 4.分析我國推動此類學習評量的配套措施。

本研究的「各主要國家」界定為美、英、OECD 等先進國家或組織為主，以他們發展的學生評量為比較的基礎。

二、研究待答問題

根據以上研究目的，本研究主要的待答問題如下：

- 1.各主要國家學生基本能力評量的內涵與優缺點為何？
- 2.各主要國家推動學生基本能力評量的策略為何？
- 3.各主要國家學生基本能力評量的呈現方式有何借鑑之處？
- 4.我國推動學生基本能力評量的可能配套措施為何？

第二節 研究設計

本研究透過文獻分析，以比較研究為主。首先，蒐集各主要國家有關學生基本能力評量的文獻，「各主要國家」以英、美、OECD 等主要國家或組織為主，就目前已發展的學生基本能力評量為文獻蒐集的範圍。其次，就蒐集的文獻進行比較，比較各主要國家學生基本能力評量的內涵，並分析這些內涵的代表意義，以做為我國發展類似能力評量的參考。初步研究成果提出後並進行專家的諮詢，以確保研究結果的正確性。最後，提出發展我國學生基本能力評量的

配套措施，以做為教育部相關政策訂定之參考。具體而言，本研究的執行流程如圖 1.1：

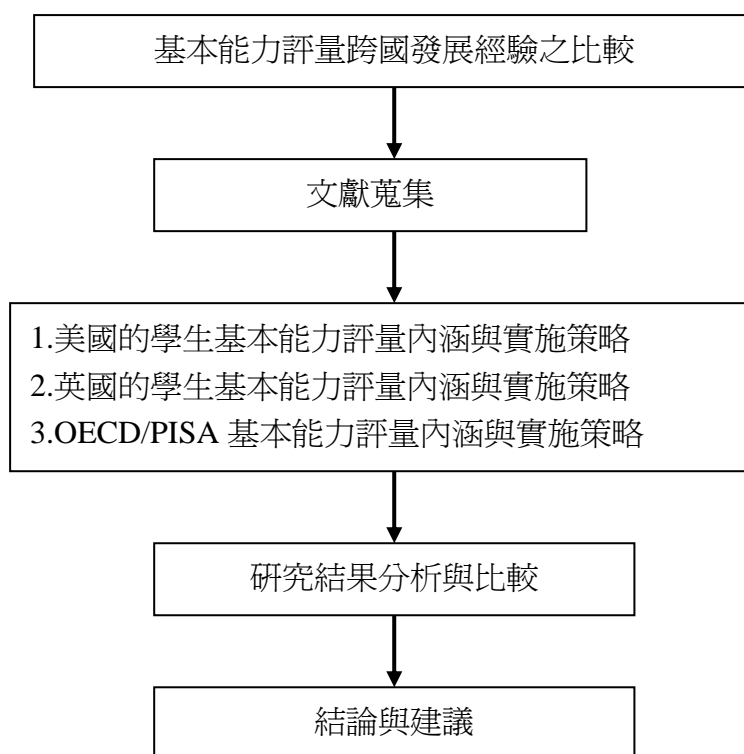


圖 1.1 研究架構

第三節 研究進度與人力

本研究的期程自 94 年 8 月 1 日至 94 年 12 月 31 日止，主要的工作與進度如表 1.1 所示。

表 1.1 研究主要工作期程與進度

日期 工作項目	民國 94 年				
	8/1 8/31	9/1 9/30	10/1 10/31	11/1 11/30	12/1 12/31
一、研究設計與架構確立	■				
二、文獻蒐集與分析	■	■	■		
三、發展比較架構		■			
四、進行比較分析			■	■	
五、分析結果與討論				■	■
六、撰寫研究報告					■

本研究依業務需要，主要的人力配置如下：

計畫主持人：王世英

（國立教育資料館館長）

研究主持人：張鈿富

（國立暨南國際大學教育政策與行政學系教授）

協同主持人：吳慧子

（國立暨南國際大學教育政策與行政學系助理教授）

研 究 員：梅瑤芳（國立教育資料館秘書）

王秉倫（國立教育資料館視聽組主任）

張雲龍（國立教育資料館編輯）

周文菁（國立暨南國際大學教育政策與行政學系碩士）

研 究 助 理：黃奕通

（國立暨南國際大學教育政策與行政學系研究生）

潘佳汝

（國立暨南國際大學教育政策與行政學系研究生）

第二章 美國學生基本能力評量

「國家教育發展評量」The National Assessment of Educational Progress (NAEP, 以下簡稱 NAEP) 是美國評量學生成就的代表，本章介紹此一評量的內涵，並以 NAEP 為例，分別說明四年級、八年級以及十二年級實施的閱讀能力評量 (reading)、數學能力評量 (mathematics) 與科學能力評量 (science)。

第一節 NAEP 規劃

依據美國法律規範 (U.S. Department of Education, 2005)，全國教育統計中心 (National Center for Education Statistics, 以下簡稱 NCES) 的最高行政長官應負責執行 NAEP 這個政策。自 1988 年以來，NAEP 在全國評量管理委員會 (National Assessment Governing Board, 以下簡稱 NAGB) 所制定的政策指導之下，平穩的執行其功能。NAGB 是由 NCES 所指定組成的一個管理委員會，成員包括教育專家學者、推選出的官方代表、家長、以及一般民眾代表。平時美國民眾所聽見熟悉的名詞 The Nation's Report Card，其實即是 NAEP 的一種親切用語。The Nation's Report Card 採用一種非技術性 (即易懂的用詞) 方法來介紹 NAEP 主要內容。The Nation's Report Card 的目的是為了老師們、家長們和一般社會大眾，了解關於全國評量的重要資訊。

一、組織

當談到與 NAEP 相關的組織時，NCES 與 NAGB 是兩個關鍵的單位。全國教育統計中心 (NCES) 是美國教育部 (U.S. Department of Education) 直屬教育科學部門 (Institute of Education Sciences, 以下簡稱 IES) 的四個所屬研究單位之一。NCES 負責執行此項由上而下的計畫 (mandated project): 「國家教育

發展評量」(NAEP)。

全國評量管理委員會(NAGB)於1988年經國會通過,是一個獨立於美國教育部的機構,目的在於制定及領導NAEP之相關政策(The Nation's Report Card, 2005a)。NAGB的成員包含聯邦的(federal)、州政府的(state)以及地方的官員(local officials),此外還有教育學者專家、家長、商業領導者與社會一般民眾的成員代表。

二、目的

NAEP由實施評量的地方範圍可簡單分成:全國的(National NAEP)、各州的(State NAEP)、以及地方的(NAEP Trial Urban District Assessment)評量(The Nation's Report Card, 2005a)。NAEP主要目的是:

- 1.反映美國學生在主要課程領域上應該知道和可以做的廣泛能力(to discover what American students know and do in key subject areas);
- 2.測量長時間範圍內的教育發展情形(to measure educational progress over long periods of time)。

三、類型

為了達成這些目的,NAEP的計畫包括兩種重要評量類型(The Nation's Report Card, 2005a):(1)主要評量(Main NAEP)與(2)長期發展趨勢評量(Long-term Trend NAEP)。

主要評量(Main NAEP)注重學生現在學習的內容是什麼(what students are learning today)。一般NAEP所呈現的教育評量結果大多屬於主要評量(Main NAEP)。Main NAEP評量的內容每十年都會重新檢視與更新。每兩年,全國的4年級、8年級與12年級學生需接受閱讀能力(reading)與數學能力(mathematics)評量;同樣的,每兩年,各州的4年級與8年級學生需接受閱讀能力與數學能力評量。每四年,全國的4年級、8年級與12年級學生需接受科學(science)與寫作(writing)評量;同樣的,每四年,各州的4年級與8年級學生需接受科學能力與寫作能力評量。

長期發展趨勢評量(Long-term Trend NAEP)主要為全國的施測,屬於全

國性評量。主要評量閱讀（reading）能力與數學（mathematics）能力的長期發展趨勢。每四年，NAEP 針對 9、13 和 17 歲的學生實施數學（mathematics）與閱讀（reading）評量，採用的評量內容與架構是修正自 1969 年以來的評量測驗題目。因此，長期發展趨勢評量（Long-term Trend NAEP）的每次結果可以拿來比較。

除了 Main NAEP 與 Long-term Trend NAEP 兩種主要評量類型之外，還有一些特別議題的 NAEP 評量研究，例如目前正在進行中的高中學習成績研究（High School Transcript Study）及科學基礎的評量（Technology-Based Assessment）等（The Nation's Report Card, 2005a）。

表 2.1 長期發展趨勢評量（Long-Term Trend NAEP）與主要評量（Main NAEP）的比較

	長期發展趨勢評量 Long-Term Trend Assessment	主要評量 Main NAEP Assessment
目的 Purpose	每四年評量學生在數學與閱讀上面的表現，2004 年公布目前最新的評量，2008 年再公布下一次的評量結果。	每兩年評量學生在數學與閱讀上面的表現，2005 年公布目前最新的評量，其他的科目領域學生也會接受評量。
評量的內容 Content Assessed	評量的基本內容從第一次的評量（1971 年第一次閱讀評量，1973 年第一次數學評量）開始就維持保留不變，雖然自 2004 年有一些修正。	評量每十年會修正改變，來反應在全國學校課程上的調整改變。新的評量架構可以因應這些修正。
學生的抽樣 Students Sampled	全國性代表即選取 9 歲、13 歲與 17 歲的學生。樣本人數比主要評量的樣本人數小。2004 年的結果，僅限於公立學校的白人學生、非裔學生與西班牙裔學生。	選取 4 年級 8 年級與 12 年級的學生。偶數年評量全國性代表的學生，奇數年評量各州及城市地區的學生。抽取較大量的樣本來代表少數族群學生，例如亞裔及太平洋學生、美國印地安學生。
評量實施 Administration	每四年實施評量。	每兩年實施數學與閱讀能力評量。

評量結果呈現方式 Results Reported	結果呈現學生成就表現與長期的趨勢發展。成就表現的呈現方式利用評量分數，但並不是使用學生成就的好壞來呈現結果。結果只報告全國性的資料。	評量結果的呈現方式為評量分數，並且呈現成就表現結果為基礎、熟練或是精熟。自 1990 年以來，結果的呈現包含全國性的、參與的各州以及其他的參與行政地區。
------------------------------	--	--

資料來源："What are the differences between long-term trend NAEP and main NAEP, "by The Nation's Report Card, 2005b, Retrieved October 5, 2005, from http://nces.ed.gov/nationsreportcard/about/ltr_main_diff.asp

四、實施評量時間

NAEP (The Nation's Report Card, 2005a) 並不呈現學生個人或是個別學校的評量結果，NAEP 呈現的結果是特定領域的學習成就 (NAEP does not provide scores for individual students or schools ; nstead, it offers results regarding subject-matter achievement)。NAEP 的評量是由一群經抽樣選出的學生來接受評量所代表的結果。實施 NAEP 可區分為全國性的評量 (National NAEP)，州層級的評量 (State NAEP)，以及都市地區的評量試驗 (NAEP Trial Urban District Assessment)。全國性的評量 (National NAEP) 呈現的評量結果是代表全國性的評量結果，包括公立學校與非公立的學校，評量學生於 4 年級、8 年級與 12 年級的成就。自 1990 年起，加入州層級的評量 (State NAEP)，呈現各州的評量結果，目前評量學生以 4 年級與 8 年級的學習成就為主。都市地區的評量試驗 (NAEP Trial Urban District Assessment) 主要是因應 No Child Left Behind Act，針對都市地區的 (urban districts) 學生實施 NAEP 評量。

NAEP 實施評量時程表，以下分成兩個階段以表格來呈現。表 2.2 是由 1969 到 2004 年 NAEP 實施評量的時間表。表 2.3 是預計由 2005 年到 2017 年的評量計畫。

表 2.2 從 1969 年到 2004 年 NAEP 實施評量順序時間表 Chronology of National Assessment of Educational Progress (NAEP) Assessments from 1969 until 2004

時間 Year	全國的 National	各州的 State	長期發展趨勢 Long-term Trend
1969-70	公民 citizenship 科學 science 寫作 writing	1990 年州評量才正式開始 State assessments began in 1990	科學 science
1970-71	文學 literature 閱讀 reading		閱讀 reading
1971-72	音樂 music 社會 social studies		
1972-73	數學 mathematics 科學 science		數學 mathematics 科學 science
1973-74	職業發展 career/ occupational development 寫作 writing		
1974-75	藝術 art 基本技巧 index of basic skills 閱讀 reading		閱讀 reading
1975-76	公民 citizenship/ 社會 social studies 數學 mathematics		公民 citizenship/ 社會 social studies
1976-77	生活技巧 basic life skills 科學 science		科學 science
1977-78	消費技巧 consumer skills 數學 mathematics		公民 citizenship/ 社會 social studies
1978-79	藝術 art 音樂 music 寫作 writing		

1979-80	閱讀 reading 文學 literature 藝術 art		閱讀 reading
1981-82	數學 mathematics 科學 science 公民 citizenship 社會 social studies		數學 mathematics 科學 science
1984	閱讀 reading 寫作 writing		閱讀 reading 寫作 writing
1986	電腦 computer competence 美國歷史 US history 文學 literature 數學 mathematics 科學 science 閱讀 reading		數學 mathematics 科學 science 閱讀 reading
1988	公民 civics 文學 document literacy 地理 geography 美國歷史 U.S. history 閱讀 reading 寫作 writing		公民 civics 科學 science 閱讀 reading 寫作 writing
1990	數學 mathematics 科學 science 閱讀 reading	數學 mathematics (8)	數學 mathematics 科學 science 閱讀 reading 寫作 writing
1992	數學 mathematics 閱讀 reading 寫作 writing	數學 mathematics (4,8) 閱讀 reading (4)	數學 mathematics 科學 science 閱讀 reading 寫作 writing
1994	地理 geography 美國歷史 U.S. history 閱讀 reading	閱讀 reading (4)	數學 mathematics 科學 science 閱讀 reading 寫作 writing

1996	數學 mathematics 科學 science	數學 mathematics (4, 8) 科學 science (8)	閱讀 reading 寫作 writing 數學 mathematics 科學 science
1997	藝術 arts (8)		
1998	閱讀 reading 寫作 writing 公民 civics	閱讀 reading (4, 8) 寫作 writing (8)	
1999			閱讀 reading 數學 mathematics 科學 science
2000	數學 mathematics 科學 science 閱讀 reading (4)	數學 mathematics (4, 8) 科學 science (4, 8)	
2001	美國歷史 U.S.history 地理 geography		
2002	閱讀 reading 寫作 writing	閱讀 reading (4, 8) 寫作 writing (4, 8)	
2003	閱讀 reading (4, 8) 數學 mathematics (4, 8)	閱讀 reading (4, 8) 數學 mathematics (4, 8)	
2004	外國語文 foreign language (12) (postponed) 延期實施		閱讀 reading 數學 mathematics

資料來源："Chronology of National Assessment of Educational Progress (NAEP) assessments from 1969 until 2004," by The Nation's Report Card, 2005c, Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/about/assesshistory.asp>

表 2.3 從 2005 年到 2017 年 NAEP 預計實施評量順序時間表 Schedule for the State and National Assessment of Educational Progress (NAEP) from 2005–2017

年	主要評量 Main NAEP		長期發展趨勢 Long-term Trend
	全國的 National	各州的 State	
2005	閱讀 reading 數學 mathematics 科學 science 高中成績研究 high school transcript study	閱讀 reading (4, 8) 數學 mathematics (4, 8) 科學 science (4, 8)	
2006	美國歷史 U.S. history 公民 civics 經濟 economics (12)		
2007	閱讀 reading (4, 8) 數學 mathematics (4, 8) 寫作 writing (8, 12)	閱讀 reading (4, 8) 數學 mathematics (4, 8) 寫作 writing (8)	
2008	藝術 arts (8)		閱讀 reading 數學 mathematics
2009	閱讀 reading 數學 mathematics 科學 science 高中成績研究 high school transcript study	閱讀 reading (4, 8) 數學 mathematics (4, 8) 科學 science (4, 8)	
2010	美國歷史 U.S. history 公民 civics 地理 geography		
2011	閱讀 reading (4, 8) 數學 mathematics (4, 8) 寫作 writing	閱讀 reading (4, 8) 數學 mathematics (4, 8) 寫作 writing (4, 8)	

2012	世界歷史 world history (12) 外國語文 foreign language (12) probe : technological literacy (special study)		閱讀 reading 數學 mathematics
2013	閱讀 reading 數學 mathematics 科學 science 高中成績研究 high school transcript study	閱讀 reading (4, 8) 數學 mathematics (4, 8) 科學 science (4, 8)	
2014	美國歷史 U.S. history 公民 civics 地理 geography		
2015	閱讀 reading (4, 8) 數學 mathematics (4, 8) 寫作 writing	閱讀 reading (4, 8) 數學 mathematics (4, 8) 寫作 writing (4, 8)	
2016	藝術 arts (8)		閱讀 reading 數學 mathematics
2017	閱讀 reading 數學 mathematics 科學 science 高中成績研究 high school transcript study	閱讀 reading (4, 8) 數學 mathematics (4, 8) 科學 science (4, 8)	

資料來源：“Schedule for the State and National Assessment of Educational Progress (NAEP) from 2005–2017,” by The Nation’s Report Card, 2005d, Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/about/assessmentsched.asp>

NAEP 實施評量的課程領域有 11 個 (administers assessments in 11 subject areas)：藝術 (The Arts)、公民 (Civics)、經濟 (Economics)、外國語文 (Foreign Language)、地理 (Geography)、數學 (Mathematics)、閱讀 (Reading)、科學 (Science)、美國歷史 (U.S. History)、世界歷史 (World History) 以及寫作 (Writing)。較常見的評量領域為其中的 8 個能力：閱讀、數學、科學、寫作、美國歷史、公民、地理、和藝術。以下列舉閱讀、數學與科學領域為例，說明 NAEP 的評量內涵以及 4 年級、8 年級與 12 年級的評量方式。

第二節 NAEP 閱讀能力評量

NAEP 的閱讀能力評量強調的是全國學生在閱讀領域上所表現的學習成就 (The NAEP reading assessment measures the achievement of the nation's students in reading.)。每兩年，針對全國的 (National NAEP) 4 年級、8 年級與 12 年級學生實施閱讀能力評量，同時每兩年，還有各州的 (State NAEP) 以 4 年級與 8 年級學生實施閱讀能力評量。閱讀能力乃是評量學生在閱讀各式各樣的文章內容上表現如何，評量的題型採用選擇題 (multiple-choice) 與結構式問答題 (constructed-response)。自 1992 年起，NAEP 的閱讀評量架構一直擔任閱讀評量題庫建立過程中的參考依據。NAEP 閱讀能力評量架構 (NAEP Reading Framework) 反應出閱讀能力是一個多元動態、互動的完整過程。

一、閱讀能力評量架構

NAEP 閱讀能力評量學生在各種不同文章內容的理解、詮釋、與批判性思考的學習成就。例如閱讀能力評量架構，即由閱讀目的及閱讀理解層次兩者來做區分。閱讀目的可分為 (The Nation's Report Card, 2005e)：為文藝學識而閱讀；為獲取資訊而閱讀；為執行任務而閱讀 (reading for literary experience, reading to gain information, and reading to perform a task)。另一方面，閱讀理解層次可區分為 (The Nation's Report Card, 2005e)：

第一層級：形成一般性的理解 (Forming a general understanding)，要求學生將整篇閱讀文章做整體一般性的理解。

第二層級：發展解釋 (Developing an interpretation)，要求學生理解文章內容的前後關係。

第三層級：讀者與文章之間的連結 (Making reader/text connection)，要求學生將文章內容與其先備知識做一個連結。

第四層級：檢視文章內容與架構 (Examining content and structure)，要求學生對文章做出批判、評量與統整。學生需展現他們對於文章閱讀的了解，透過回答一個全面性的問題，反應出四個不同層級的閱讀成就表現。表 2.5 說明 NAEP 的閱讀評量架構。

表 2.5 NAEP 的閱讀評量架構。

The Framework of the NAEP Reading Assessment

閱讀能力評量三個不同脈絡 Three different contexts for reading were assessed :	閱讀能力評量四個不同層級 Students were assessed on four different aspects of reading :
為文藝學識而閱讀 Reading for literary experience：閱讀者藉由閱讀小說、短篇故事、詩、劇本、傳奇、傳記、神話及科幻故事來探索事件、特徵、場景、背景、情節、表演及文學創作的語言。	形成一般性的理解 Forming a general understanding：形成一般性的理解，要求學生將整篇閱讀文章做整體的理解、發展解釋 developing interpretation：要求學生理解文章內容的前後關係
為獲取資訊而閱讀 Reading for information：閱讀者可以透過閱讀雜誌、報紙、教科書、短篇文及演講稿以獲得資訊，進而了解世界。	讀者與文章之間的連結 Making reader/text connections：讀者與文章之間的連結，要求學生將文章內容與其先備知識做一個連結
為執行任務而閱讀 Reading to perform a task：閱讀來表現在工作上：閱讀者使用他們從公車或是火車時刻表、裝修指南、遊戲、報稅表格（十二年級）、地圖等等，學到的東西。	檢視文章內容與架構 Examining content and structure：檢視文章內容與架構，要求學生對文章做出批判、評量與統整

資料來源：“2005 Reading Assessment：What does the NAEP reading assessment measure?” by U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005a, Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/reading/whatmeasure.asp>

二、閱讀能力評量題型

NAEP 的閱讀能力評量實施對象是全國的 (National NAEP) 4 年級、8 年級與 12 年級學生，2006 年春季將會公布 12 年級學生的閱讀評量結果，就全國性 (National NAEP) 評量計算，有超過 165,000 位 4 年級學生，159,000 位 8 年級學生，與 12,000 位 12 年級學生接受過閱讀能力的評量 (The Nation's Report Card, 2005e)。NAEP 的閱讀能力評量題型包含選擇題以及短的結構性問答題與長的結構性問答題。(The NAEP reading assessment contains multiple-choice questions, as well as short and extended constructed-response questions.) 表 2.6 與表 2.7 列出 4 年級、8 年級與 12 年級閱讀評量，在目標與實際閱讀問題上的分布比率。

表 2.6 4 年級、8 年級與 12 年級閱讀評量，在目標與實際閱讀問題分布比率
Target and actual percentage distribution of questions, by context for reading, grades 4, 8, and 12 : 2005

	4 年級 Grade 4		8 年級 Grade 8		12 年級 Grade 12	
	Target	Actual	Target	Actual	Target	Actual
為文藝學識而閱讀 Literary experience	55	51	40	29	35	23
為獲取資訊而閱讀 Information	45	49	40	40	45	50
為執行任務而閱讀 Perform a task			20	31	20	27

資料來源：“2005 Reading Assessment : Distribution of reading question 2005,” by U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) , 2005b, Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/reading/distributequest.asp>

表 2.7 四、八、十二年級生，目標與實際學生閱讀時間分布比率

Target and actual percentage distribution of student time, by aspect of reading, grades 4, 8, and 12 : 2005

	4 年級 Grade 4		8 年級 Grade 8		12 年級 Grade 12	
	Target	Act.	Target	Act.	Target	Act.
組成一個廣泛式的了解／發展、說明 Forming a general understanding／ developing interpretation	60	68	55	59	50	56
閱讀者與文字的連結 Making reader／text connections	15	14	15	17	15	14
測驗內容與架構 Examining content and structure	25	17	30	24	35	29

資料來源：“2005 Reading Assessment : Distribution of reading question 2005,” by U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) , 2005b, Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/reading/distributequest.asp>

第三節 NAEP 數學能力評量

每兩年，全國的 4 年級、8 年級與 12 年級學生需接受閱讀能力 (reading) 與數學能力 (mathematics) 評量；同樣的，每兩年，各州的 4 年級與 8 年級學生需接受閱讀能力與數學能力評量。根據 NAEP 的 2005 年數學評量架構中建議增修的項目，目標預期能夠 (1) 反應現今課程重點與目標 (to reflect recent curricular emphases and objectives)；(2) 包含很多政策制訂者、教育專家學者、實務執行者及感興趣的市民們認為應評量的部分 (to include what various policymakers, scholars, practitioners, and interested citizens believe should be in the

assessment)；(3) 維持短期趨勢可觀察出學生成就表現的變化，例如從 1990 年開始的 4 年級及 8 年級的數學能力評量的；(3) 維持短期趨勢可觀察出學生成就表現的變化，例如從 1990 年開始的 4 年級及 8 年級的數學能力評量的 (to maintain the short-term trend lines in grades 4 and 8 that began with the 1990 mathematics assessment to permit the reporting of changes in student achievement over time)；(4) 每年級包含更清楚及更多特定的目標 (to include clearer and more specific objectives for each grade level) (The Nation's Report Card, 2005)。自 1992 年的評量開始，評量題型包含選擇題、短結構問答題與延伸型結構性問答題 (multiple choice, short constructed response, and extended constructed response)。評量時間平均分配給選擇題以及兩類問答題的部分，各占一半的比率。

一、數學能力評量架構

類似自 1990 年到 2003 年所使用的數學能力評量架構，2005 年數學能力評量架構包含五個數學內涵的標準 (The Nation's Report Card, 2005f)：

1. 數字概念與運算 (number properties and operations)
2. 測量 (measurement)
3. 幾何的概念 (geometry)
4. 資料分析與機率 (data analysis and probability)
5. 代數 (algebra)

五個數學內涵標準是數學能力評量架構中的第一個框架，另一個接續的框架是數學運算過程中的複雜性程度，可區分為低階複雜、中等複雜以及高階複雜的題型 (low complexity, moderate complexity, and high complexity)。例如：低階複雜的題型要求學生記憶特性，中等複雜的題型認為學生應該能連結特性之間的關係，以及高階複雜的題型可以評量學生在一個數學模式上分析假設的能力 (The Nation's Report Card, 2005f)。

數學領域評量題目 2005 年修訂的新架構採用的是複雜性程度，此方式既類似又不同於 1996-2000 年所使用的架構：數學能力的分級 (概念理解、知識建立與解決問題) (The Nation's Report Card, 2005f)。表 2.8 進一步比較 2005 年與 1990-2003 年的數學能力評量架構的異同處。

表 2.8 2005 年數學能力架構

比較 2005 年與 1990–2003 年的數學能力評量架構	
評量內容 Content Areas	評量內容部分是相同的。2005 年架構提供更清楚的各年級目標，4 年級的評量內容比重分布情形並未改變，只有 8 年級與 12 年級的部分稍做修訂。
認知部分 Cognitive Dimension	在 2005 年的架構中，數學能力的部分強調其複雜性，指的是數學結果的正確性取得的難易程度。The 2005 framework replaces the dimensions of mathematical ability and power (which require inferences about the student responding to the item) with the dimension of mathematical complexity (which describes the mathematical expectations of an item) .
評量成就結果 呈現方式 Achievement Levels	評量成就結果呈現方式未改變，仍採用基本的、精熟的與進階的三等級。The achievement levels—Basic, Proficient, and Advanced—have remained the same.
題目種類 Item Types	題目種類有選擇題、簡答題、問答題，各種類的回答時間比例仍維持一樣。The item types—multiple choice, short-constructed response, and extended-constructed response—and the percentage of time students spend on items of each type have remained the same.
計算過程 Computation	計算過程整個脈絡的前因後果在題型中皆會評量，在 2005 年的架構中，尤以導出結果的部分為主要評量關鍵。 Computation is assessed in and out of context and across content areas in both frameworks ; in the 2005 framework, the emphasis on computation not in context has increased.
運算過程與工具 Use of Manipulatives and Tools	運算過程與工具的需求維持一樣。The use of manipulatives and tools has remained the same.
計算機策略 Calculator Policy	計算機策略對 4 年級與 8 年級 2005 仍維持一樣。2005 年對於 12 年級學生允許在測驗時攜帶課堂上使用的計算機，沒帶者，將提供計算機。

資料來源：“2005 Mathematics Assessment,” by National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005a, Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/mathematics/frameworkcomparison.asp>

二、數學能力評量題型

2005 年 NAEP 的數學能力評量架構包含五個數學內涵的標準：數字概念與運算、測量、幾何的概念、資料分析與機率、代數。從 1990 年到 2003 年所使用的數學能力評量架構加以修訂而成。數學領域的評量題目分成三個等級：低階複雜、中等複雜以及高階複雜的題型（Within each of these five content strands, students are asked questions that involve low, moderate, and high mathematical complexity）。數學能力評量題型包含選擇題、短結構問答題與延伸型結構性問答題。延伸性問答題讓學生在回答問題的過程中展現推理的能力以及融入自我思想。短的結構問答題與延伸性問答題共占學生回答時間的 50%。因此，數學領域評量時間平均分配給選擇題以及兩類問答題的部分，各占一半的比率。表 2.9 詳細列出 2005 年數學能力評量在內容上，目標與實際的時間分配情形。

表 2.9 2005 年數學能力評量在內容上，目標與實際的時間分配情形

Target and actual percentage distribution of items, by content area and grades 4, 8, and 12 : 2005

	4 年級 Grade 4		8 年級 Grade 8		12 年級 Grade 12	
	Target	Actual	Target	Actual	Target	Actual
數字概念與運算 Number properties and operations	40	42	20	24	10	18
測量 Measurement	20	19	15	16	301	331
幾何的概念 Geometry	15	16	20	21		
資料分析與機率 Data analysis and probability	10	11	15	14	25	19
代數 Algebra	15	12	30	25	35	30

資料來源：1 At grade 12, measurement and geometry are combined.

“2005 Mathematics Assessment,” by National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005b, Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/mathematics/distributequest2005.asp>

第四節 NAEP 科學能力評量

科學能力評量可提供的資訊有（1）全國學生對於知識與科學的了解（the knowledge and scientific understanding of the nation's youth），（2）科學教育課程中能提高學生的學習成就的部分（the features of science education programs that relate to high levels of student achievement）（The Nation's Report Card, 2005g）。通常 NAEP 的科學能力評量設計過程中都包含著此兩項功能。NAEP 的科學能力評量架構目標是在於測驗出學生在科學領域上的基本知識、概念以及處理科學性質的問題，同時強調學生在科學領域上所知道的與可以做的能力。每四年，全國的 4 年級、8 年級與 12 年級學生需接受科學（science）與寫作（writing）評量；同樣的，每四年，各州的 4 年級與 8 年級學生需接受科學能力與寫作能力評量。

一、科學能力評量架構

美國 NAEP 對 4 年級、8 年級及 12 年級學生進行科學知識跟技巧的評量，根據 NAGB 所制定的 NAEP 科學能力評量架構，針對以下三個重要科學範疇對學生進行評量（The Nation's Report Card, 2005 g）：

- 1.事實的知識（knowledge of facts）
- 2.整合知識成爲大架構（an ability to integrate this knowledge into larger constructs）
- 3.使用科學的能力如經由工具、程序或是推理的方式，發展對自然世界的進一步了解（the capacity to use the tools, procedures, and reasoning processes of science to develop an increased understanding of the natural world）

依據這個科學能力評量架構，NAEP 預設的三個重要的科學領域，分別是：

- 1.地球科學（Earth）
- 2.自然科學（physical）
- 3.生命科學（life）

同時，科學能力評量也測驗學生「需要知道」且「可以做」的三個科學關鍵能

力：

1. 概念理解 (conceptual understanding)
2. 科學探究 (scientific investigation)
3. 實際推理 (practical reasoning)

由圖 2.1 呈現科學領域能力評量架構 (The Nation's Report Card, 2005h)。

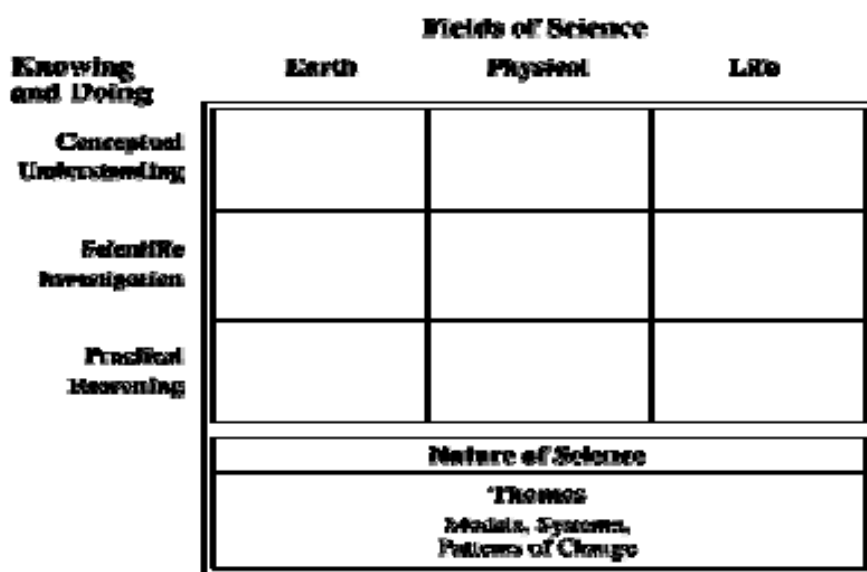


圖 2.1 2005 年 NAEP 科學能力評量架構

二、科學能力評量題型

美國的 NAEP 科學能力評量架構主要是由一個矩陣 (包含內容與過程) 所建立 (The NAEP science assessment is organized according to a content-by-process matrix)。內容指的是地球科學、自然科學、與生命科學，過程指的是對於概念理解、科學探究與實際推理。評量三個類別中知道的與可以做的能力。此外還有兩個支配的主軸：科學的本質以及科學的主題。2005 年 NAEP 科學能力評量測驗題型包括選擇題、短的結構性問答題以及延伸性的問答題。結構性問答題占有最少 50% 的測驗時間。表 2.10 與 2.11 詳細列出 NAEP 科學能力評量在內

容與概念過程中，時間分配的原則。

表 2.10 科學能力評量內容時間分配情形 Distribution of Assessment Time : 1996 and 2000

	地球科學 Earth			自然科學 Physical			生命科學 Life		
	Target	Actual	Actual	Target	Actual	Actual	Target	Actual	Actual
		1996	2000		1996	2000		1996	2000
Grade 4	33%	33%	33%	33%	34%	33%	33%	33%	33%
Grade 8	30%	30%	31%	30%	30%	34%	40%	40%	35%
Grade 12	33%	33%	33%	33%	33%	31%	33%	34%	37%

資料來源：“1996 and 2000 Science Assessments,” by National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005c, Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/science/timedevoted.asp>

表 2.11 科學能力評量概念過程分配情形 Knowing and Doing Science

	概念理解 Conceptual Understanding			科學探究 Scientific Investigation			實際推理 Practical Reasoning		
	Target	Actual	Actual	Target	Actual	Actual	Target	Actual	Actual
		1996	2000		1996	2000		1996	2000
Grade 4	45%	45%	56%	45%	38%	27%	10%	17%	17%
Grade 8	45%	45%	59%	30%	29%	18%	25%	26%	24%
Grade 12	45%	44%	56%	30%	28%	24%	25%	28%	20%

資料來源：“1996 and 2000 Science Assessments.,” by National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005c, Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/science/timedevoted.asp>

第三章 英國學生基本能力評量

教育是培育人材、提升國力的基礎。面對即將來臨的二十一世紀資訊社會，以及愈趨增強的國際經濟競爭，世界各國莫不致力於教育的反省與革新。英國在 1988 年 7 月通過教育改革法案，確立了國定課程及評量測驗的實施。而國內近年來社會輿論與民間團體倡導教育改革的呼聲幾已成爲全民的共識，行政院於 83 年成立「教育改革審議委員會」，提供今後教育改革的理念與方向，以上這些改革策畫行動，都說明了時代變遷與教育改革之間的密切關係。

由於近年來英國在世界舞台上倍感國際競爭的壓力，不但國力無法與日本、美國相提並論，即在歐洲共同市場的地位也受到德、法兩國的威脅。因此國定課程的實施是英國教育史上相當重要的里程碑，具有提升國力、重振往日聲威的重責大任。

本章先對英國的國定課程做一簡略的描述，並針對國定課程中測驗與評量內涵做一說明。

第一節 英國之國定課程與評量

本節就英國的國家課程設立目的、課程目標與科目以及國家課程與評量發展等內涵做一探討。

一、國定課程設立目的

英國的國定課程設立基於下列的目的：

1. 減少各校之間的歧異

在國定課程實施之前，英國的義務教育在各校之間存有很大的歧異。比如，在一個地區，某個學校可能有完整的科學課程，另一個學校可能完全沒有。即

使兩校均有的科學課程，也不能保證都以同樣的方式教學，甲校可能非常重視基礎科學，而乙校則偏重物理及地球科學，其他科目亦是如此。國定課程的實施可以確保所有學生學習廣博而均衡的學科內容。

2. 排除偏頗選擇

在英國的中學裡，常可看到偏頗的課程選擇，比如女生通常會選生物而放棄物理，男生則多選科技而非家政，不少男生也常放棄外國語或音樂。國定課程的實施將使學生接受同樣良好而適切的課程內容。

3. 提升學習品質

評量的實施將有效監控學生在各階段的表現，與各項目標上的進步，沒有通過的學生要加強補救，可以保障學生獲得最起碼的知能。同時，國定課程所定的標準將引導年輕人有較高的預期及自信，成為較有能力的人。

4. 建立績效責任

教育是公共政策事務的一環，所以依據社會大眾的期望所設計的共同課程，將使家長得以判斷孩子的進步情形及學校績效，地方教育當局更能評估其轄區學校的得失，甚至全國性的評鑑亦得以進行，如此教育投資將不至浪費，績效責任得以建立。

5. 導引課程的連繫

國定課程設立目的之一是改善小學和中學之間的連繫。在此之前，由於各校因地方當局的教育政策或各校的課程安排互異，教師面對來自不同學校的學生，常抱怨無法掌握課程內容，特別是一些有次序性的科目（如數學、科學、音樂等）更是困難。其次，小學的課程設計多不以學科為單位，但中學面對十六歲證書考試，必須以學科為主，且選修課程增多，課程難以連貫。國定課程的實施將可使義務教育的課程結構形成一個連續而一貫的整體，便於銜接。

6. 改善評鑑標準

國定課程的評量結合了老師評量與標準評量，兼具了「形成性」、「診斷性」、「綜合性」、「評鑑性」四種方式，是一種比較完整且符合人性化的設計。

二、課程目標與科目

1. 國定課程的目標

使個體具備成人生活所需的知識及技能，並具備謀生能力。

使社會發展共通的文化價值，能尊重不同，並導引和諧。

使國家造就出世界一流的人力，創造國家繁榮。

2.課程科目

核心基礎科目：英文、數學、科學。

基礎科目：藝術、地理、歷史、現代外語、音樂、體育、科技。

以上十個科目是義務教育中必須教導的內容，所有五至十六歲的兒童都必須學習（除了現代外語是在十一歲以上才是必修的）。宗教教育雖未列在十科之內，一般學校也都會教導，但其課程內容是由地方教育當局籌畫，並非屬於國家層次。

3.課程結構

國定課程的結構包括以下幾個重要部分：

4.學習計畫

這是一份非常清楚的說明書，告訴老師到底該教些什麼，所有在此科目中必須教導的內容、技巧、方法、策略等都包含在內。

5.成就目標

代表學習者在每一個科目中必須獲得的知識、技能與理解，每個目標都有評鑑的工具來記錄學生應該知道什麼，以及能夠做些什麼。

成就水準

每一個成就目標都可以再分為十個水準，代表由五至十六歲的進步情形。比如 level 1 相當於一年級可以到達的程度，level 10 則相當於中學以上的程度。

6.集合檔案

每個兒童的成就目標都會記錄下來，當要報告家長時，這些分數都會集合起來如同一份檔案，清楚說明每個兒童的表現。

7.成就說明

這是針對每一個成就目標及成就水準所做更詳細的說明。

8. 課程實施

以往由於英國學校制度相當不統一，即使同一地區也常有不同的系統，比如一個十歲的兒童，可能就讀 middle school 的五年級，也可能是 secondary 或 upper school 的三年級，常易混淆。國定課程因此採取關鍵階段的畫定，將五至十六歲的義務教育學習分為四個階段，每一個階段包含二至三年的學習。表

3.1 說明國定課程實施的四個關鍵時期。

表 3.1 國定課程實施的關鍵階段

第一關鍵階段	5 歲-7 歲
第二關鍵階段	7 歲-11 歲
第三關鍵階段	11 歲-14 歲
第四關鍵階段	14 歲-16 歲

三、國定課程與評量發展

英國的正式國名為「大不列顛與北愛爾蘭聯合王國」(the United Kingdom of Great Britain and Northern Ireland, UK)，全境由靠近歐洲大陸西北部海岸的不列顛群島的大部分島所組成，包括英格蘭 (England)、威爾斯 (Wales)、蘇格蘭 (Scotland) 與北愛爾蘭 (Northern Ireland) 四個地區，隔北海、多佛爾海峽和英吉利海峽同歐洲大陸相望，是一個島國。其各區教育制度差異極大，一般所稱英國教育，系指英格蘭和威爾斯兩區之教育。

英國自 1988 年教育改革法案通過之後，首次提出「國定課程」(National curriculum)，明定義務教育階段的基本科目、學習方案、成就目標和評鑑設計，其屬於強制性的施行，適用於 5-16 歲學生，共計十一年，分為四階段來施行。而其中測驗與評量是國定課程最受重視的部分，每一關鍵階段完成後，需舉行全國性的評量，稱為「中等教育普通證書考試」(General Certificate of Secondary Education, GCSE)。所有國定課程之施行由兩個獨立的行政機關監督分別是英格蘭的學校課程與評量局 (the School Curriculum and Assessment Authority, SCAA) 及威爾斯的課程與評量局 (the Curriculum and Assessment Authority)。

2004 年 4 月由國家評量機構 (National Assessment Agency, NAA) 全權負責國定課程與 GCSE 的評量與監督。其主要執掌工作如下：

1. 確保高品質的國定課程內容。
2. 公開 GCSE 與相關成就水準的評量具體項目與程序。

- 3.採用現代化的評量與嚴謹的審查系統。
- 4.確保評量過程中的公平、公正、公開，並提供一個可供諮詢的場域。
- 5.進行整體評量計畫之評估，並做一改進。
- 6.建立教育評量人員的專業知能。

測驗與評量方式是國定課程中最受重視的部分，目前的規定是在每一個關鍵階段完成後，舉行全國性的評量。換言之，全英國 7、11、14、16 歲的兒童都必須接受測驗。和以往最大的不同是全國性的評量提早進行，以前兒童除接受閱讀測驗外，都是進入高中後才接觸全國性的評量。此種改變的主要動力，來自家長及民意的支持。大多數人都希望儘早得到有關子女學習的回饋 (Cutles, 1992)。

全國性評量是一個相當複雜的問題，回顧英國評量發展的歷史，醫學及法律的資格考試開始於十九世紀的早期，只有通過入學考試者才有機會進入這些學校就讀。而 1944 年通過的教育法案雖然在標榜福利國家的政策下，欲使全體國民皆有接受中學教育的機會，不因社經地位的不同而有差別，但法案中認為人的心智能力分為三等，所以應有三種程度不同的學校，依據學生的年齡、能力、性向分類入學。因此，小學畢業後接受十一歲的考試，決定學生進入文法、技術或現代中學。基本上，文法與技術中學，其課程偏重學術性內容且水準甚高，而現代中學則側重特定科目的職業教育，因此雙軌的精神依然存在。許多學者認為這種十一歲的分流測驗，過早區分兒童為 academic 與 less academic，失之武斷，可能造成難以彌補的錯誤。

1951 年，英國採行十六歲的「一般教育證書考試」(General Certificate of Education, GCE)，考試分普通級 (O level) 與高級 (A level)，一方面代表大學或其他高等學術機構的入學鑑定，一方面也是進入政府機關或其他企業的資格考試。考試科目包括學術性及技藝性的學科，往往多達七、八十種。由於 GCE 普通考試的通過率限在應考學生的前百分之二十，競爭相當激烈。

1965 年英國實施「中等教育證書考試」(Certificate of Secondary Education, CSE)，提供中等以下學生通過考試的機會。但是一般家長仍希望子女參加 GCE 考試，因為其聲望較高 (吳知賢，1996)。

保守黨政府上台後，希望建立全國的教育標準，稱為「一般中等教育證書考試」(General Certificate of Secondary Education, GCSE)。依照國定課程的實施構想，GCSE 與目前國定課程的四階段考試可以取得密切關連，四個關鍵期

考試如同 GCSE 的準備工作。1944 年第一次的第四關鍵期評量，學生參加 GCSE 的英文、數學、科學三科考試，即代替核心科目的評量，其他科目也將有類似安排，如此構成一套嚴密的考試系統。

依據 1988 年教育改革法對國定課程之規定，國定課程採取關鍵學習階段的劃定，將五至十六歲的義務教育學習分為四個關鍵階段，每一個階段包含二至四年的學習，如表 3.2 所示。為確保每一位學生在義務教育階段的各個學習關鍵期皆能達到成就目標，每一位學生在每一個關鍵學習期結束前必須接受測驗，及接受國定考試（National Tests），經由國家考試來了解學生學習成就，以及學校教學表現。

英國義務教育階段共分為四個關鍵學習期，依據英國教育及技能部設計，在第一關鍵期的二年級結束前，必須接受測驗，考試科目為數學與英文；第二、三關鍵期結束前必須再度接受測驗，考試科目是數學、英文與科學；第四關鍵期畢業之前，則可參加屬於資格考試之「中學教育普通證書」（General Certificate of Secondary Education ,GCSE）考試，取得中學學歷文憑，一方面取得繼續升學資格，另一方面可作為應徵工作之依據。

表 3.2 英國義務教育階段關鍵學習期之考試類別與科目

關鍵期	年級	年齡	考試類別	考試科目
第一關鍵期 (key stage1)	一 二	5-6 6-7	SATs (Standard Assessment Task)	數學、英文
第二關鍵期 (key stage2)	三 四 五 六	7-8 8-9 9-10 10-11	SATs	數學、英文、科學
第三關鍵期 (key stage3)	七 八 九	11-12 12-13 13-14	SATs	數學、英文、科學
第四關鍵期 (key stage4)	十 十一	14-15 15-16	GCSEs	五至九門 GCSE 學科

英國學生成就評量，大可從法定強制性評量與非法定的選擇評量來區分。而法定強制性評量，又可區分為教師班級評量與全國性測驗（國家考試）二種。茲就法定強制性評量敘述如下（陳明印，2004）：

一、教師班級評量

被視為教室裡教和學的一部分，評量部分包括教學方案的所有內容範圍，必須考慮到教室情境中所有學生表現的證據，如經由討論和觀察等評量所獲知；在每一階段的國定課程中，為學生設定了每一科目的成就標準，教師會根據此些標準的規定，透過課堂上對學生的了解，來判斷其學業進步情況；對其進行評鑑，評定每一科目的成就，以便提供下一步學習的建議。

二、全國統一評量

從考試設計來看，英國在義務教育階段的統一考試相當密集，在前三階段以數學、英文、科學為核心考試科目，目的在檢視學生的基本能力，以便作為更進一步學習的基礎；第四階段的考試則在確保義務教育的成效，以及評估學生升學的能力。總言之，國家考試的設計在確保義務教育階段，學生學習成果的品質（National Assessment Agency，2005）。

第二節 英國第一關鍵期評量

在關鍵第一階段末期，通常在 7 歲，你的孩子將會做英語和數學的測驗和任務。當這些將是你的孩子第一個強制被實施的測驗或任務時，它們也將形成你的孩子最後評分的其中一部分。關鍵第一階段的評價主要是基於教師本身的評估準則，及教導你的孩子如何在整整兩年內做出優秀的專業判斷。教師本身的評估準則中，教師將會運用並進行教學過程中的評價，但是，你的孩子此外將還需參與國家課程在第 2 年的英語和數學測驗。

一、第一關鍵期評量的任務

在整個第 2 年，你的孩子將參加一些考試並且完成一些任務。而他們的教師將決定你的孩子應該完成哪一次的測驗。關鍵第 1 階段的測驗和任務，從一月開始就隨時可能開始進行。這也表明了教師能保證你的孩子已經準備好接受這些測驗和任務，使你的孩子能順利通過沒在以前的兩年被國家課程所包括的任何事情上的測驗。而有時候測驗將是平常課堂作業的一部分，你的孩子甚至可能不知道他們正被評價。這些全部的測驗和任務一開始將會對孩子做一個十分鐘的完整介紹，並對孩子加以解釋好讓他們知道未來將會發生的事。這不僅僅是給你的孩子檢查並詢問問題的一次機會，同時也讓他們理解到未來他們需要做什麼。而班級教師也將注意你的孩子的測試和任務。

二、主要測驗科目及內容

(一) 英文

小孩將在他們的關鍵第一階段的英語測驗中被評價（分別是閱讀、書寫、拼寫測驗）。

1. 閱讀 (Reading)：此項測驗主要是教師採取說故事給孩子聽的方式，並在反覆的問答中來評價你的孩子理解故事情節的能力。測驗中包含評價你的孩子理解故事（約在 45 分鐘左右）的一個能力，並且能順利回答書面問題。閱讀測驗的形式可能採分組、一組或一種種類，而形式將交由教師來決定哪種是非常合適的。
2. 寫作 (Writing)：此項測驗主要是評價兩項的書寫測驗，而這兩項的書寫測驗分別在不同日期來進行並完成（較長時間的測驗約 45 分鐘左右，而較短時間的測驗約 30 分鐘左右）。同時，你的孩子的書寫測驗也將會是被評價的一部分。
3. 拼寫 (Spelling)：此項測驗主要是評價你的孩子拼寫 20 個單詞（在不到 30 分鐘內）的一次單獨拼寫測驗。

試題範例

這是在關鍵第一階段的閱讀測驗中的一個問題。

After reading a story titled Grandfather's pencil, children were asked the following question.

Find and copy 3 words about how things moved in the wind

(二) 數學

小孩將做一數學測驗，並且被要求他們能運用數學的技能。這些技能包括：

1. 數、讀和寫整個數字直到 100，並且將他們安排好。
2. 將數字進行平分、加倍。
3. 解釋他們怎樣解決一個問題。
4. 能說出數字是古怪異常的。

試題範例

這是在關鍵第一階段的數學測驗中一個典型的問題。

在一個袋子裡有 60 顆甜食。20 顆甜食是紅的。16 顆甜食是黃的。其餘是綠色的。

多少甜食是綠色的？這會顯示你將如何設法安排的能力。

三、等級的闡述

在你的孩子已經進行關鍵第一階段的閱讀、書寫、拼寫及數學測驗之後，他們的成績標記之後會被拿去和水準門檻做一比較（最小成績標記會被要求增加每一個水準），以知曉他們已經取得什麼水準。而你的孩子的教師將運用他們最後的評價來做為測驗和任務的結果。水準是 W（努力達到第 1 水準），1，2C，2B，2A，3 和 4 或者以上。這些個別的測驗結果將不會在過程中報告，但

是會用來作為教師本身的評估準則。大多數的孩子在關鍵第一階段末期將會到達第 2 水準（2C，2B，2A）（National Assessment Agency，2005a）。

第三節 英國第二關鍵期評量

在關鍵第二階段結束時，通常在 11 歲，孩子將會做英文、數學和科學測驗。教師也將監控並且評價你的孩子在英文、數學和科學的能力和弱點，並進行基礎上的教師本身評估準則報告（運用他們的專業判斷，告知你的孩子他們正處於什麼水準），及你的孩子在此階段的測驗結果。

一、第二關鍵期評量的任務

與每次測驗一個小時相比較，此階段的測驗將通常進行到 5 月初，並在許多天中進行，而時間更長且並不持續。你的孩子將有豐富的機會來為這些測驗作準備，並且通過在任何事情上的測試，包括沒在以前的四年被國家課程所包括的任何事情上的測驗。而你的孩子的試卷將被寄到被評閱的一個外部記分員那裡。

孩子的學校可以選擇使用 QCA 在英文部分的選定測驗及其在第 3、4 和 5 年的數學測驗。這些測驗可能在此年內被選定實施，並被你的孩子的學校作為選定測驗，但其實施結果並不會報告給全國周知。此選定測驗將會幫助孩子的學校看見孩子已經取得什麼進步。如果你想知道你的孩子的學校是否使用這些測驗，你必須和你的孩子的老師進行交流和對話。

二、主要測驗科目及內容

茲就英文、數學、科學敘述如下：

（一）英文

你的孩子將在他們的關鍵第二階段的英語測驗中被評價（分別是閱讀、書寫、拼寫測驗）。

1. 閱讀（Reading）：在此過程的測驗中，你的孩子將被測試關於他們在閱讀中，字裡行間寓意的理解程度、想法、性格和分析事件的能力。他們將被問及一系列閱讀的技能並做一能力的評價，例如解釋訊息、評論作者所使用的語言。這次的測驗大約持續 45 分鐘，另外的 15 分鐘為閱讀時間。
2. 寫作（Writing）：此項測驗主要是評價兩項的書寫測驗，一為較長時間的測驗約 45 分鐘左右，一為較短時間的測驗約 20 分鐘左右。在這次測驗過程中，記分員將會注意你的孩子在書寫和閱讀上的能力、標點語法及詞彙的運用上之使用是否得宜。
3. 拼寫（Spelling）：此項測驗主要是評價你的孩子拼寫 20 個單詞（在 10 分鐘內）的一次單獨拼寫測驗。

試題範例

這是在關鍵第二階段的英文測驗中較長時間的書寫測驗的例子。

<p>Time for a change Your school is thinking about making some changes to the school day. This is an outline of the new timetable that might be introduced :</p> <p>7.00am Registration and assembly 7.20am Exercise session 8.00am Breakfast 8.30am Lessons 12.30pm Sport or homework club 1.30pm School closes</p>	<p>Here are some reactions to the new timetable.</p> <p>'I don't like the idea of having all the lessons in the morning'</p> <p>'I think it would be healthy to start the day with some exercise'</p> <p>'It would mean getting up very early'</p> <p>Pupils have been asked to give their views in assembly.</p> <p>Your task is to write down what you will say about this new timetable in the assembly.</p>
--	---

(二) 數學

在此測驗中，數學有 3 次單獨的測驗必須被完成。一次 45 分鐘的測驗（不能使用電算機）；一次 45 分鐘測驗（能使用電算機）；一次 20 分鐘測驗（精神數學測試）。而全部的測驗主要是要測出孩子在使用數學方面的技能問題。這些技能包括：

- 1.知道乘法表，並且能靈活運用。例如：乘、除
- 2.能使用一個量角器測量角度
- 3.計算可能被分成長方形的形狀的周長和區塊
- 4.透過收集和使用表格、圖和圖表訊息解決問題
- 5.解決與比率、比例有關的問題。

試題範例

這是在關鍵第 2 階段的數學測驗中一個典型的例子。

這是 5 張數字卡片



使用這全部 5 張數字卡片，並使之正確。

$$\square \square \times 2 = \square \square \square$$

(三) 科學

在此測驗中，科學有 2 次單獨的測驗必須被完成，而每一個測驗時間為 45 分鐘。此測驗主要是評價你的孩子理解科學發展和使用科學證據的想法。孩子將在科學的很多方面進行測驗且有下列的要求：

- 1.了解身體主要的器官並知道它們在那裡。

- 2.透過它們不同的特性來加以分類材料。
- 3.對一個簡單的實驗設計作評論。
- 4.對簡單的電氣電路進行繪圖並且修改。

試題範例

這是在關鍵第二階段的科學測驗中一個典型的問題。

Meena 在一本書裡讀到人的心臟在休息時的速度大約每小時跳動 4300 次。

Meena 說想要檢查這訊息，但我不能花一個小時來測量我的心跳。

Meena 該怎樣做才能迅速查明她的心臟是否大約每小時跳動 4300 次？

三、等級的闡述

在你的孩子已經進行關鍵第二階段測驗之後，他們每一部分考試的成績會加總起來，然後會被拿去和水準門檻做一比較（最小成績標記會被要求增加每一個水準），以知曉他們已經取得什麼水準。可能被授予的整套教師自身評估準則是從 W 以來（努力達到第 1 水準）到第 8 水準。大多數的孩子在關鍵第二階段末期將會到達第 4 水準。你的孩子將會收到各個測驗的單獨水準，例如：書寫（這包括拼寫）、閱讀和英語所有測驗加總的單獨水準（National Assessment Agency, 2005b）。

第四節 英國第三關鍵期評量

在第三關鍵期即將結束時，通常在 14 歲，小孩需接受英文、數學、科學的測驗。此測驗被設計來評量小孩在接受中等教育三年來的發展情形，以及確認是否已預備好接受 GCSEs 的挑戰和取得職業文憑。老師將對英文、數學、科學，做持續的觀察記錄與評量小孩的強項與弱點，以及宣布老師所評價的等級，並附上小孩測驗的結果。老師也需評量小孩在其它國定課程中的科目，並且

提供老師所評定的等級。學校因此將能夠提供更多小孩學習進展情形的資訊給家長。

一、第三關鍵期的任務

評量將舉行為期四天，通常在五月初，每次測驗將不超過 75 分鐘為限。測量範圍不涉及到非國定課程的部分。學校也可能選擇使用 QCA 非強迫性測驗，來評量七年級和八年級小孩的英文與數學程度，這些測驗可在每年的任何時段舉行，成績由學校來評定，但結果不將公告。

學校也可能針對七年級學生的進步情形作評量，測驗科目包括英文和數學，針對在第二關鍵期結束時，沒有達到等級四的學生，他們是由外部主考人來評定成績。非強迫性與進展的測驗，將幫助學校了解小孩學習進展情形。如果家長想了解學校是否使用這些測驗，將可詢問小孩的教師。

二、主要測驗科目及內容

(一) 英文

1. 閱讀 (Reading)：關於閱讀試卷，學生需根據小冊子所提供的內容來回答問題。此測驗將包括文學與非文學，小說與散文，然而不是每種形式將出現在每次測驗中。測驗持續 75 分鐘，包含 15 分鐘的閱讀時間。
2. 寫作 (Writing)：寫作試卷將進行 75 分鐘，學生需完成一篇短篇寫作 (30 分鐘)，以及一篇長篇寫作 (45 分鐘)。
3. 莎士比亞 (Shakespeare)：學生需回答關於課堂上所討論過的關於莎士比亞戲劇中的場景、角色，此測驗持續 45 分鐘。

試題範例

這是在第三關鍵期英文測驗中，一篇短篇寫作之例題：

Going wild

You have suggested turning an overgrown part of the school grounds into a wildlife area, with a pond, seating and plants.

The headteacher sends you this note :

This is a good idea but you will need to get parents involved.

The pond has to be dug, seating made and all the planting done.

Can you write a section for the school newsletter persuading parents to help?

Thank you.

Write the section of the school newsletter persuading parents to help make the wildlife area.

(二) 數學

數學評量方面有四個不同層級，每一個層級評量不同的等級。不同層級評量的等級分別為：等級 3-5、等級 4-6、等級 5-7、等級 6-8。老師將依最適合學生程度的層級來予以施測。

每一個測驗層級分成三部分，有一小時測驗需在不用計算機的情況下完成評量；一小時的測驗是可用計算機的情況下；以及 20 分鐘關於智力的數學測驗。小孩需接受一系列的數學概念測驗，內容包括：

1. 分數、小數、百分比，以及代數問題。
2. 測量角度與計算面積。
3. 機率、計算平均數、範圍與頻率，了解圖表、座標圖的廣泛範圍。

試題範例

這是在關鍵第三階段的數學測驗中一個典型的問題。

I think of a number, then I carry out these operations on my number.

Multiply by 5

Add 8

When I carry out the operations in one order the answer is 105.

When I carry out the operations in the other order the answer is 73.

What is my number?

Show your working out.

(三) 科學

科學評量方面有二個不同層級，每一個層級評量不同的等級。不同層級評量的等級分別為：等級 3-6、等級 5-7。如同數學評量，老師將選擇最適合學生程度的層級來予以施測。

每一個測驗層級可分成二部分，每部分持續一小時，小孩將接受科學課程的幾個不同方面的評量，內容包括：

1. 單細胞與多細胞功能，營養學、植物呼吸，以及健康狀況。
2. 固體、液體、氣體，元素、化合物、混合物，化學的反應，地質的變化與金屬作用。
3. 物質的變化過程，如電流、電學、以及磁性作用，風雨等自然現象與動態性，光與聲音，以及太陽系系統。

試題範例

這是在關鍵第三階段的科學測驗中一個典型的問題。

Samantha opened a tin of white paint. The paint consisted of a liquid and particles of titanium dioxide that are insoluble in the liquid. The paint had separated into layers.

(i) What type of substance is the paint? Tick the correct box

a compound an element a mixture

(ii) What type of substance is titanium dioxide? Tick the correct box.

a compound an element a mixture

(iii) Why did the particles of insoluble titanium dioxide sink to the bottom?

三、等級的闡述

在小孩完成第三關鍵期的測驗後，每部分的測驗將加總起來，給該科目一個最後分數以及等級。然後，與該科等級門檻比較後，看是否該科以達到第三關鍵期所應達到的等級。老師評量給予的等級範圍可從等級一至等級八，大部分學生在完成第三關鍵期的學習後應該達到至少等級五或等級六的程度（National Assessment Agency，2005c）。

下表為你的小孩在學校完成每關鍵學習階段時，所應完成的測驗、任務與評量摘要表（National Assessment Agency，2005d）：

End of key stage 1 (age 7)	End of key stage 2 (age 11)	End of key stage 3 (age 14)
此測驗包含什麼？		
閱讀、寫作（包含書寫）、拼寫、數學。第一關鍵期強調教師的評量。	閱讀、寫作（包含書寫）、拼寫、數學、智力的數學與科學。	英文（包括閱讀、寫作和研讀莎士比亞戲劇）、數學、智力數學與科學。
測驗時數？		
總共三小時內。	總共在五～六小時之間。	總共在七～八小時之間。
何時測驗？		
此測驗可在一月後的任何時間舉行。由課堂老師決定小孩何時準備好可接受測驗。	2005 年在 5 月 9 日～13 日舉行；2006 年將在 5 月 8 日～12 日舉行。	2005 年在 5 月 3 日～6 日舉行；2006 年將在 5 月 2 日～5 日舉行。
教師評量包含什麼？		
英文、數學、科學。	英文、數學、科學。	英文、數學、科學、歷史、地理、現代外語、設計、科技、美術、音樂、公民、資訊與網路科技、宗教教育。

如何獲知小孩評量結果？		
學校將寄送成績單，告知家長關於小孩在測驗中已達到的等級，以及教師對小孩評量的等級。		
大部分小孩需達成等級？		
大部分小孩應達到等級二的程度（等級二又被分為三部分：二 A、二 B、二 C）。	大部分小孩應達到等級四的程度。	大部分小孩應達到等級五的程度，很多小孩將被預期達到等級六的程度。

第四章 OECD/PISA 基本能力評量

1999 年「國際經濟合作與發展組織」(The Organization for Economic and Cooperation Development, 簡稱 OECD) 發展跨國的評量學生計畫, 此一計畫稱之為「國際性學生評量計畫」(Program for International Student Assessment, 簡稱 PISA)。此一計畫推動數年之後, 在跨國性的研究架構支持下, 在國際已具有相當的影響力。

第一節 OECD/PISA 內涵

OECD/PISA 主要的功能在了解個人參與社會活動的能力。評量的內涵包括閱讀能力、數學能力與科學能力三者。

「閱讀能力」在解釋、使用與反映文字的內涵, 以便達成個人的目標、發展個人的知識與潛能以及社會的參與。閱讀能力的組成與領域面向包括: 閱讀不同種類的文字內涵, 例如連續散文、文告及結構不同文體。表現不同種類的閱讀任務, 例如回憶特定的資訊、發展一種解釋或是反應文字的內涵。是否能閱讀因不同情境而寫作的文字內涵。

「數學能力」在確認、了解與從事數學以及透過數學在世界扮演的角色能作成良好基礎的判斷, 並協助個人就目前與未來生活所需, 做一個建設性、關懷心與反省性的市民。數學能力的組成與領域面向包括: 數學內涵——主要數學的「大概念」、模式與問題解決、以及數學在不同情境中的應用。

「科學能力」在結合科學知識與證據本位結論的形成, 以及發展相關假設以了解與協助作成自然世界有關的決定, 並且透過人類活動來促成改變。科學能力的組成與領域面向包括: 科學概念、科學過程技巧以及不同情境的科學應用。整個評量架構的項目如表 4.1 (OECD, 1999):

表 4.1 OECD/PISA 學力評量的領域

	閱讀能力	數學能力	科學能力
定義	了解、使用與反映文字的內容，以便達成個人的目標、發展個人的知識與潛能以及社會的參與。	確認、了解與從事數學以及透過數學扮演的角色能作成良好基礎的判斷，並協助個人就目前與未來生活所需，當一個建設性、關懷心與反省性的市民。	結合科學知識與證據本位結論的形成，以及發展相關假設以了解與協助作成自然世界有關的決定。並且透過人類活動來促成改變。
成分及領域的面向	<p>閱讀不同種類的文字內容，例如不同形態的連續散文（敘述性）、文告及結構不同文體。</p> <p>表現不同種類的閱讀任務，例如回憶特定的資訊，發展一種解釋或是反應文字的內容或文字的形式。</p> <p>閱讀因不同情境而寫作的文字內容，例如為個人的興趣，或為符合工作的要求。</p>	<p>數學內容包括主要數學的「大概念」，第一測量循環中這些「大概念」以改變與成長以及空間與形狀為主。在未來的測驗循環中，機會、數量的推理、不確定性與依賴關係會應用上去。</p> <p>數學能力即模式、問題解決；這些能力區分為三類：執行程序、連結能力以及數學思考與規約能力。</p> <p>數學在不同情境中的應用，即問題影響個人、社群或全世界。</p>	<p>科學概念包括能源保護、調適、解組分化——從主要的物理、生物、化學等領域選擇、應用這些知識於能源保護、物種保育與物質使用。</p> <p>過程技巧——確認證據、形成、評鑑及溝通結論。這些不依賴先前設定的科學知識，但也不能用在沒有的科學內容。</p> <p>不同情境的科學應用</p>

資料來源：“Measuring student knowledge and skills” by OECD, Paris:

Author, p. 12.

第二節 PISA 閱讀能力評量

在 OECD/PISA 的閱讀能力一詞是一個比較廣義的了解。相對的少數年輕人在社會中沒有閱讀技巧，此一計畫所需測量的 15 歲學生並不強調他們在閱讀方面技術性的意義，而主要在反映當前的閱讀觀點。此一觀點要學生離開中學時，應該能建構、拓展與反應校內外不同情境所閱讀的連續或不連續的內容。

一、閱讀能力定義

閱讀與閱讀能力的定義是隨著社會、經濟與文化之改變而不同。學習的概念，尤其是終身學習的特殊性，已經拓展了閱讀能力的知覺與必要性。閱讀能力不在只考慮童年時在學校所獲得的能力，更重要的是把閱讀能力看成是一套知識、技能與策略的進步，個人從生活的不同脈絡與透過同儕的互動中建立這些能力。

認知觀點的閱讀能力強調互動性的閱讀本質與建構性的理解本質（Bruner, 1990; Binkley and Linndkyla, 1997; OECD, 1999）。讀者利用先前的知識和社會與文化共享的文字及情境線索對閱讀的內容產生意義。建構意義時，讀者使用不同的程序、技能與策略去強化、監督與維繫理解。這些過程與策略隨著不同的情境與目的以及不同的閱讀內容而不同。

目前有兩個國際閱讀能力的評量，一個是 IEA 主持的閱讀能力評鑑（The International Association for the Evaluation of Educational Achievement's Reading Literacy Study, 簡稱 IEA/RLS）；另一個是國際成人識字調查（The International Adult Literacy Survey, 簡稱 IALS）。目前是由加拿大統計局與 OECD 共同主持，並強調功能性本質的閱讀。

IEA/RLS 對閱讀能力的界定是：“The ability to understand and use those written language forms required by society and/or valued by the individual.”

IALS 對閱讀能力的界定是：“Using printed and written information to function in society, to achieve one's goals, and to develop one's knowledge and

potential.”

OECD/PISA 閱讀能力的定義是閱讀能力在解釋、使用與反映文字的內涵，以便達成個人的目標、發展個人的知識與潛能以及社會的參與。

二、閱讀能力內涵

OECD/PISA 閱讀能力的評量包含：情境、文字內容（文章）與測驗題目。

情境

情境涉及使用作者所構思的文章內容之地方或環境，OECD/PISA 強調閱讀情境不只在學校，還包含學校以外的地方。因此，情境可以被了解成基於使用的目的而予以任務歸類。基本上，將閱讀分成「對於個人使用之閱讀」、「對於公眾使用之閱讀」、「對於職業之閱讀」、以及「對於教育之閱讀」四類。例如閱讀一本教科書可以看成是教育情境，使用方面，在教育這個歸類上的意義是獲得資訊；它與他人的關係是由教師或其他的教導者所指定。在內容方面主要是導向教導與學習。OECD/PISA 閱讀情境的歸類如表 4.2。情境方面的閱讀任務所占比率如表 4.3（OECD, 1999）。

表 4.2 OECD/PISA 閱讀的情境歸類

	個人使用	公眾使用	工作上使用	教育上使用
週遭環境	自己相關的朋友	匿名者	對象、同事、上司	教導者
使用上	好奇、接觸	資訊	做	學
內 容	文件、小說、傳記、說明手冊和雜誌、地圖	告示、規則、節目、小冊子、表格	說明、手冊、時間表、報告、表格、圖	課文、地圖、概要、表格、圖

表 4.3 OECD/PISA 各種閱讀情境任務的分配比率

情 境	任務比率 (%)
個人使用的	28
教育使用的	28
工作使用的	16
公眾使用的	28
合 計	100

1. 文字內容型態

OECD/PISA 將文字內容的型態區分為連續性的文字內容型態與非連續性文字內容型態（結構與型式）。連續性的文字內容型態包括：描述（description）、故事（narration）、解說（exposition）、立論（argumentation）、說明（instruction）或指令（injunction）與超媒體文字（hypertext）。OECD/PISA 建議的各種連續性文字內容型態的比率如表 4.4。

表 4.4 OECD/PISA 各種連續性文字內容型態的分配比率

課文類型	連續性文章 (%)	測驗 (%)
故 事	20	13
解 說	33	22
敘 述	20	13
立論／說服	20	13
說明／指令	7	5
合 計	100	66

非連續性文字內容組成方式與連續性文字內容組成不同，需要不同的閱讀方法。比較簡單的可以將非連續性文字內容劃分為結構與型式兩類。結構性的非連續性文字內容包括：簡單的陳述、聯合的陳述、有趣的陳述、巢狀的陳述以及組合的陳述。型式的非連續性文字內容包括：表格、資訊表、有價的證明、證件、電化與廣告、圖表、應用圖示、表與矩陣、條列與地圖。OECD/PISA 建議的非連續性文字內容的比率以及測驗比重如表 4.5。

表 4.5 OECD/PISA 建議的非連續性文字內容比率以及測驗比重

文字內容類型	非連續性文字內容比率 (%)	測驗所占比率 (%)
圖	33	11
表	33	11
概要	10	3
地圖	10	3
表格	8	3
廣告	6	2
合計	100	33

三、測驗內容與任務特徵

OECD/PISA 的閱讀能力測驗由三種題目所組成：1.問題或指導，設定任務給學生；2.反應型態，設計要求學生展示對任務的精熟；3.計分的規則，訂定學生相關反應的評分。

在問題或指導方面，可分為巨觀與微觀來看。巨觀面包括：1.形成較廣的一般性了解；2.回憶資訊；3.發展一種解釋；4.反思文字內容；5.反思文字型式。這些測驗焦點的比率如表 4.6。

表 4.6 文字內容型態的任務區分

領域	考試應占比率 (%)	連續文字內容比率 (%)	非連續文字內容比率 (%)
回憶資訊	20	13	7
廣泛的了解	20	13	7
發展一種解釋	30	20	10
反思內容	15	10	5
反思型式	15	10	5
合計	100	66	34

微觀方面，考慮資訊要求的型態、資訊提供與要求之配對型態、以及可能的干擾資訊。OECD/PISA 所建議的閱讀五個主要領域有關的測驗反應結構與多重選擇的分配比率如表 4.7。而閱讀能力此一概念架構如圖 4.1（OECD, 1999）。

表 4.7 OECD/PISA 閱讀能力主要領域測驗反應結構與多重選擇的分配比率

領域	測驗所占比率 (%)	結構反應要求的任務比率 (%)	結構反應要求的測驗題目比率 (%)	要求多重選擇的測驗題目比率 (%)
回憶資訊	20	35	7	13
廣泛的了解	20	35	7	13
發展一種解釋	30	35	11	19
反思內容	15	65	10	5
反思型式	15	65	10	5
合計	100		45	55

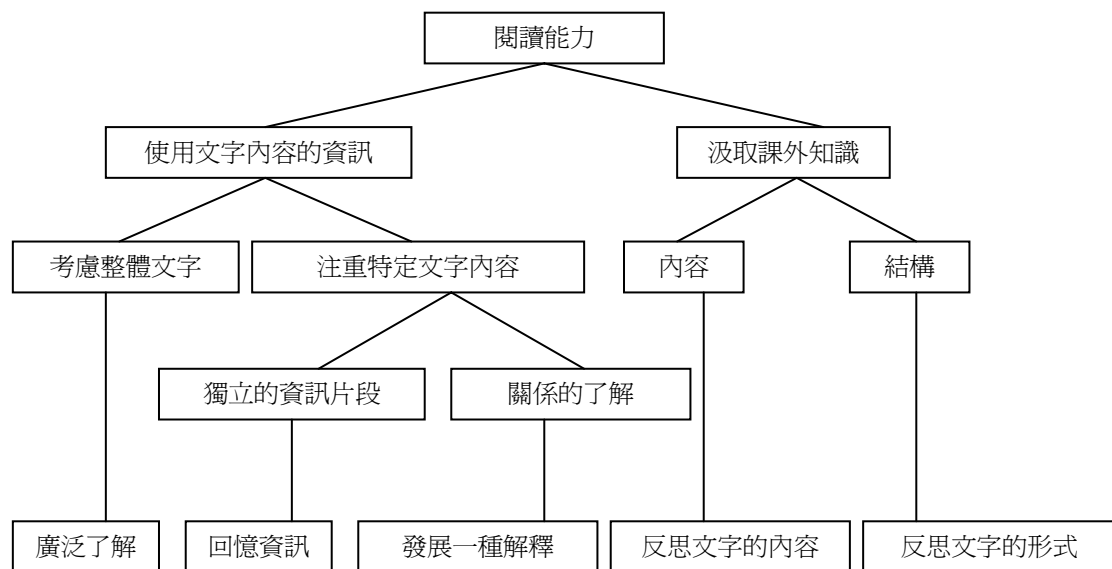


圖 4.1 閱讀能力的主要結構特徵

第三節 PISA 數學能力評量

數學能力 (mathematic literacy) 這個領域關注的學生能力是培養他們的數學能力中去面對未來的挑戰。這些能力包括分析、推理以及有效的溝通概念，透過形式、公式解決不同領域與情境的數學問題。

一、數學能力的界定

OECD/PISA 對數學能力的定義是指「在確認、了解與從事數學以及透過數學在『世界』扮演的角色能作成良好基礎的判斷，並協助個人就目前與未來生活所需，做一個建設性、關懷心與反省性的市民。」

在 OECD/PISA 的數學能力界定中，「能力」(literacy) 一詞，在傳統學校的數學課程中所選擇並強調的數學知識與技能，並不是 OECD/PISA 所強調的重點。OECD/PISA 強調的數學能力重點是數學知識應用在不同的脈絡與不同的方面所需要的反思與洞察能力。「世界」(world) 意謂著個人所居住的自然、社會與文化的環境。「從事」(engage in) 並不是狹隘的侷限於物理或社會行為。「從事」也意味著溝通、採取方式、關聯、評估甚至是欣賞數學。定義中並不限定數學的功能性應用，數學在美學與休閒方面的要素亦包括在此一定義的精神裡。此外數學能力的定義中「目前與未來的生活」包括個人生活、職業生活與社會生活，與同儕、親人及社區公民的生活。

二、數學能力的內涵

數學能力的組成方面，OECD/PISA 就上述的定義將數學能力分為主要的與次要的兩個方面。數學能力主要的方面包括數學能力 (mathematic competences) 與數學的大概念 (mathematic big ideals)。數學能力的次要方面一為數學課程要素 (mathematic curricular strands)，另一為情境與脈絡 (situation and context) (OECD, 1999)。

數學能力是指一般的數學技巧與能力，例如問題解決、使用數學語言與數

學模式。OECD/PISA 這方面的數學能力是一種一般數學技巧的非階層性呈現，但與各階段的教育都有關聯。這些數學技能包括：

1. 數學思考模式；
2. 數學辯證技巧；
3. 模式建立技巧；
4. 問題呈現與解決技巧；
5. 表徵化技巧；
6. 符號、形式與技術技巧；
7. 溝通技巧；
8. 輔助與工具技巧。

在實際的評量實施時，OECD/PISA 採取的方式是將這些能力區分成三個較大的等級，這些能力等級如下：

等級 1：再製、定義與運算；

等級 2：問題解決的連結與統整；

等級 3：數學思考、規約化與洞察力。上述的八種能力在所有的能力等級中都可能扮演某種角色。因此，這些數學技能很難歸入某一個能力等級內。

等級 1 能力：再製、定義與運算。這一層級包含是實的知識、表徵、等式的認識、數學物標的與特性的記憶、例程序的執行、標準代數的應用與應用技巧的發展。符號與公式的操作與表達以及計算亦包含在這一等級的能力。在測驗上可用測驗題或半開放形式的題目來測量。這一等級尤其是與符號、形式與技術技巧有關。等級 1 數學能力的實際例子如下：

Ex1：方程式的解決

$$7x - 3 = 13x + 15 \quad (x = ?)$$

Ex2：平均數的計算

7, 12, 8, 14, 15, 9 等數的平均數為何？

Ex3：繪出 69% 所占的部分

等級 2 能力：問題解決的連結與統整。此一等級連結不同數學的要素與範圍，必須統整解決簡單的問題。學生需要學會選擇何種策略及數學工具來使用。儘管這一等級的問題區分為非例行性，只需要相當低層基礎的數學能力。這一等級與前述的好幾種數學技巧有關，很明顯的解決所給的例題需要一些推理

或論證，因此需要數學論證技巧；學生需要建立問題的模式以進一步解決，因此需要數學模式建立的技巧。此外也需要問題呈現與解決技巧以及表徵化技巧。等級 2 能力的例子如下：

Ex4：你已開了三分之二的車程，車子起動時你的油箱是加滿油，而目前妳只剩 1/4 箱的油，你會面臨問題嗎？

Ex5：瑪麗住在離學校兩公里的地方，馬丁是離學校五公里。瑪麗與馬丁的家離多遠。

Ex6：全班 28 人，女生與男生的比率為 4 比 3，請問這一班女生有多少人？

等級 3 能力：數學思考、規約化與洞察力。此一等級學生要求數學的情境。指認與摘述在情境中的數學特徵以及利用數學解決問題。進行分析、進行解釋、發展自己的模式與策略與呈現數學的論證。等級 3 能力的數學能力，如圖 4.2。

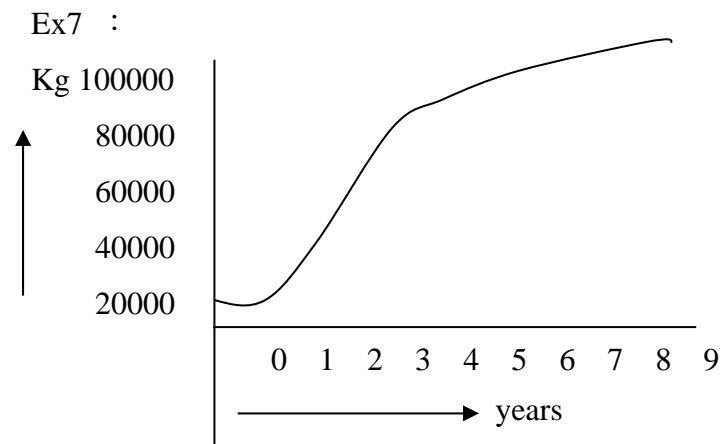


圖 4.2 數學能力應用：魚獲量與時間的關係

左邊的座標是魚塭魚群增加的重量，右邊的則是年數，請問一位漁夫應該等多少年，才能每年抓到最大數量的魚？

OECD/PISA 的數學生活化 (mathematisation) 是透過數學觀念與概念的使用以組織所知覺到的真實。根據已知的知識與技能去組織活動去發掘未知的規

則、關係與結構。這個過程有時候稱為「水平的數學生活化」(horizontal mathematisation)。這個過程所需要的活動包括：

1. 在一般的脈絡裡確認特定的數學；
2. 基模結構化 (schematising)；
3. 問題的公式化與視覺化；
4. 發現關係與規則；
5. 在不同問題中重組相似性。

一旦問題被轉化成數學問題即可用數學工具予以解決。數學工具可以用來操作與定義真實世界問題的數學模式。此一過程即所謂的「垂直的數學生活化」(vertical mathematisation)。垂直的數學生活化有下列活動：

1. 藉由公式呈現一種關係；
2. 證明規則；
3. 定義與調整模式；
4. 組成與統整模式；
5. 規約化。

數學的生活化包括水平的數學生活化與垂直的數學生活化。它是發生在這兩個不同的階段。第一階段是水平生活化，將實際的生活世界轉化成數學世界；第二階段是垂直的生活化，在數學的世界裡，運用數學的工具解決數學問題。將數學問題應用於實際生活情境並不一定就是等級 3 的能力，也可能是等級 2 能力。

Ex8：1980 年國防預算 3000 萬，國家全部預算則為 5 億；1981 年國防預算 3500 萬，而總預算則為 6 億 500 萬。兩年間的通貨膨脹率是 10%。

假設你要參加一個和平團體的相關演講，你想解釋國防預算在這期間是減少的，請問你該如何解釋？

假設你受邀參加軍事學術研討演講，你想解釋國防預算在這一段時間是增加的，你將如何解釋？

(一) 數學的大概念

數學大概念是指在日常生活情境中數學的一套相關概念，因為在真實世界中，我們不會用單一種數學來解決問題，通常都是多種相關的數學概念共同來

解決。OECD/PISA 所使用的數學大概念包括：

- 1.機會 (chance)；
- 2.改變與成長 (change and growth)；
- 3.空間與形態 (space and shape)；
- 4.量化推理 (quantitative reasoning)；
- 5.不確定性 (uncertainty)；
- 6.依賴與關聯 (dependency and relationships)。

OECD/PISA 將這些概念合成最重要的兩個概念，分別是改變與成長以及空間與模型。

1.改變與成長

任何一種現象都是會改變的，而這些改變都可以用數學的方式呈現出來，例如直線、幾何、曲線圖、週期等。而爲了要對控制具有敏感度，Stuart 認爲我們需要下列的能力：

- (1) 用可理解的形式來代表改變；
- (2) 了解改變的基本型態；
- (3) 當改變發生時，知道特殊的改變型態；
- (4) 將這些技巧應用到外在的世界；
- (5) 控制改變的通則，以獲取我們最大的利益。

至於成長則可用代數與長條圖顯示出來，也可以透過實證的方式獲得，例如資料分析與統計。

2.空間與形態

空間與形態是數學上的另一個重要概念，例如房子、橋、都市計畫等。而這些空間與形態和日常生活息息相關，它可以用幾何立體圖的方式表現出來。因此我們必須用各種方法領悟空間的概念，使人類在日常的生活中能更有效的利用空間。這些能力的要求包括：

- (1) 了解物體的相對位置；
- (2) 能有邊看邊做邊想的能力；
- (3) 能有穿越空間、建構形態的能力；
- (4) 能融會貫通實際的生活空間與數學上的空間（如地圖）；
- (5) 能分辨平面上的立體概念。

（二）數學課程基本要素

OECD/PISA 認為傳統數學課程的要素不能也不應忽略，因此將它列入數學能力的次要指標之中，這些內容要素可以確保測驗項目的平衡以及合理的將測驗內容與學校課程相連結。OECD/PISA 的數學內容要素包括：1.數（number）；2.測量（measurement）；3.估計（estimation）；4.代數（algebra）；5.函數（functions）；6.幾何（geometry）；7.機率（probability）；8.統計（statistics）；9.離散數學（discrete mathematics）。

（三）情境

情境是數學能力評量的重要部分，主要是應用數學於不同的情境。選擇數學的方法以及呈現結果都需要依據問題所處的環境。各種情境應提供學生數學方面的參與，並練習情境與情境的轉換。

在 OECD 的評量中，情境也被看成是一種「距離」，最接近的距離是個人的生活，其次是學校生活、工作與休閒運動，然後是社區與社會的日常接觸之生活，更遠的則是科學的脈絡。科學脈絡包括抽象假設的證據、數目的規約或空間的型態等等。在這樣的界定下，OECD 的五種情境評量包括：個人、教育的、職業的、公共與科學的。

三、測驗結構

由於測驗強調實際情境發生的數學問題，因此數學能力相當重要指標之一就是要將數學應用在日常生活情境中。在 60 分鐘的評量時間裡，OECD 的測驗結構如下：

1.題目的內容

OECD/PISA 的數學評量任務的焦點放在「真實的脈絡」(authentic context)，所謂「真實的脈絡」是指能真正的被適用在生活經驗上的，而不是真實的客觀情境，不只是顧慮到受測者的週遭事物，必須將各國的文化情境列入考慮。例如：兩件同樣的衣服與兩杯同樣的飲料是 1200 元，但如果買一件衣服與三

杯同樣的飲料是 900 元，請問衣服與飲料各是多少錢？

OECD/PISA 認為這種題目雖然可以得到客觀真實的答案，但在真實的生活經驗中沒有人會被要求做這種事情。

2.測驗題目的形式

OECD/PISA 的測驗題目形式包括選擇式 (multiple-choice) 題目、封閉式問答題 (closed-constructed response) 與開放式問答題 (open-constructed response)。

(1) 選擇題

主要在測驗較低層次的能力，包括計算與理解的能力。至於較高層次能力的評量則要用開放性問題。

(2) 封閉式問答題

類似選擇題，只有一個正確答案。和選擇題不一樣的是學生無法用猜測的方式猜出正確答案，因為問答題並沒有標示出選擇的項目。

(3) 開放式問答題

不只是要找出答案，還要學生說出思考的過程，解釋答案的意義。通常這是較高層次能力的表現。

3.題目型式與分配比率

在第一循環的測驗中，平均分配兩大概念的題目 (改變與成長、空間與形態)，三階層的能力題目比率則接近 1：2：1，此一測驗的結構如表 4.8 所示。

4.報告的呈現

報告的方式有下列的選擇：單一數學能力尺度的評量報告：每一個數學大概概念分別的評量報告：三種能力等級分別的報告。至於以何種報告方式呈現，依預試資料的分析而定。

表 4.8 OECD 數學能力測驗的題目型式與分配結構

題目的型式	改變與成長			空間與形態		
	能力等級			能力等級		
	等級 1	等級 2	等級 3	等級 1	等級 2	等級 3
一個標記	6 (6)	5 (5)		6 (6)	5 (5)	
多個標記	2 (5)	2 (5)		2 (5)	2 (5)	

每一個大概念包含 15 題

每一個大概念包含 21 分（括號的數字為期望的分數）

8 個多標記題目；22 單一標記題目

第一等級 12 分；第二等級 20 分；第三等級 10 分（趨近 1：2：1）

第四節 PISA 科學能力評量

在 OECD/PISA 的科學能力評量方面，以「證明為基礎」的陳述是科學能力之基礎，藉以了解測驗觀和理論之合理性。若不能測驗、不是科學理論，充其量只是「偽科學」。

一、定義與組織

Millar 和 Oaborne（1998）定義現代科學課程視為「閱讀和融合科學和技術資訊之能力」。總結科學能力的不同觀點，科學能力可視為一個階層。Bybee（1997）有四個層級：最低層次兩個為「名義上科學能力」、「功能能力」、第三為「概念和過程科學能力」、最高層次為「多重領域科學能力」。OECD/PISA 所提出的科學能力與其第三層級較類似。PISA 舉出重要的界定語句，例如：科學能力是使用科學知識界定問題和以「證明為基礎」的結論，了解和幫助決定，探究人類活動自然世界和變遷。

OECD/PISA 提出三個觀點：科學過程、科學概念、情境。前面兩個觀點是

任務結構和學生表現特徵。第三個觀點將確認評估任務的發展。OECD/PISA 架構確認評估的焦點，是基於科學教育的成果。

(一) 科學過程

過程是心智上的行動，使用想像、獲得、詮釋和使用獲得知識與理解的證明或資料。過程使用於關聯目標的範圍，意指世界科學觀念的描繪和更進一步使用科學理解之成果。PISA 提出五個步驟，用以理解科學過程：

1. 確認調查問題的科學性；
2. 在科學調查中界定證據需要；
3. 描述或評鑑結論；
4. 溝通有效結論；
5. 展現理解之科學概念。

(二) 概念和內容

OECD/PISA 使用四個向度來做科學概念選擇：

1. 每日情境的關聯；
2. 下個十年或更久之關聯性；
3. 情境定義視為應該被展現的科學能力；
4. 結合選擇科學過程。

(三) 情境

特別的情境影響表現，所以評估任務時，決定和控制則顯得相當重要。在選擇情境時，科學評估目的，要能接近學生知識技能的應用，和學校義務教育的要求。「真實世界」情境通常有一些問題，影響個人和區域社群的成員。在國際研究中，評估的主題應該與學生興趣和所有國家生活情境有所關聯。OECD/PISA 的內涵趨近這些知識的應用，並且在這些主題中反映「真實世界」的情境。OECD/PISA 評定科學能力的項目與類型如表 4.9 所示。

表 4.9 評定科學能力的項目類型

給 予	要求的回應
確認可調查的科學性問題	
某一研究或步驟的數字資料被搜集或比較	選擇或提出可被測試的問題或意見
某一情況中描述的問題是可被科學性調查的	想出一個可被科學調查的問題
數個問題意見的形成或相關情境的呈現	從中選取可被回答、測試的科學調查
驗明科學調查中需要的證據	
需被測試的某一意見或放置在問題上的假說或刺激物	選擇或提出需測驗意見的資訊或以此意見為據提出預言： a.什麼事項須被比較 b.什麼變項須被改變或控制 c.什麼額外資訊是必須的 d.什麼行動須執行才有利適當資料的收集
汲取或評價結論	
那些結論資料可被汲取	提出適合資料的結論
資料（測驗結果或觀察的資料）	選擇適合資料的結論並給予解釋
資料（測驗結果或觀察的資料）並從中得到結論	從給與的資料（支持或不支持的結論或建議）產生信賴的推理
傳達令人信服的結論	
支持一個結論或建議以及特定聽眾所需的整合資訊	藉由建立在刺激物的適當證據、資料提出一個清晰的辯論給予觀眾
科學觀念的論證了解	
情況可被預言及解釋所需的資訊	產生或選擇一個預測或解釋或根據科學概念了解的額外資訊或根據在問題中或刺激教材中未提供的資訊

二、任務特徵與評估架構

任務通常可視為呈現於情境中一連串的問題。從問題中，去了解學生的科學能力。每一個問題，都有潛在問題之後的理論基礎。例如，PISA 舉了六個主題，每一主題分別有其背後的目的。如要說明「確認調查問題之科學化，使用人類生物的知識，應用在科學生活和健康的領域。」

評估架構有兩個理由：第一、架構是創造任務反應「真實生活」情境；另一理由是藉測驗以了解問題，問題的提出，從不同情境問題的區分更易獲取欲探究之主旨。在所有評估中，有關的科學過程需尋求平衡，以便在 2006 年的科學報告主題中呈現。表 4.10 為「科學過程分數分配之建議」。這個建議表包含的要項分別是：確認調查問題的科學性、在科學調查中界定證據、描述或評鑑結論、有效溝通結論及展現理解之科學概念。

表 4.10 科學過程分數分配之建議

科學過程	得分點%
確認調查問題的科學性	10-15
在科學調查中界定證據	15-20
描述或評鑑結論	15-20
有效溝通結論	10-15
展現理解之科學概念	40-50

OECD/PISA 的目標主要建議概念的理解和科學知識結合「驗證為基礎的結論」。生活與健康科學、地球與環境科學以及科技的科學各占 30-40%的彈性比率。

三、報告量表

為了符合 OECD/PISA 的目標，發展學生成就量表是最基本的，量表可分為主要與次要。報告量表，基於過去成就經驗和從在科學的學習認知發展逐漸

修正。從科學概念和應用認知概念之複雜情境，科學能力從嘗試性結果分析。在 2006 年，將會有詳細的資訊呈現。這些資訊將用來配合現在關心的議題，以提供各參與國家的教育進一步的發展。

第五章 實施策略分析與比較

本章以美國、英國與 OECD/PISA 的評量實施策略為主，並比較其間的差異，以供參考。

第一節 美國學生評量的實施策略

NAEP 利用由上而下的政策執行方式，進行美國學生基本能力評量的教育政策。展現美國教育領域長期屬於地方分權自治的一個新面貌。美國 NCES 成立一個獨立專責機構 NAGB 來專責執行 NAEP 這一項計畫，並授權此機構來制訂相關教育評量方針與政策。

NAEP 有一套標準的評量編制過程，各科目訂有評量設計架構參考依據。

NAEP 的評量編製，根據各科評量架構，邀請教育官員、教育專家學者、學校教師、教材專家、與一般公民，決定評量科目的具體目標。NAEP 透過這樣協商溝通的管道，審查與修訂各科評量的測驗內容，再將評量的具體目標交由各科的測驗專家進行測驗題庫的建立。最後，各科的評量測驗題庫再一次經過審查、預試、修正、選取樣本預先施測與結果分析，才正式通過決定採用。評量是整體學生的表現，而非個別學生的成就測驗 (assessing groups of students, not individuals)

編製評量以專業人員為主導 (deciding what is assessed)；

功能擴增下，實施 NAEP 的州政府與地方教育單位逐漸增加 (expanding NAEP's Mission)；

各科評量架構的建立是測驗編製的基礎 (beginning with a framework)；

發展與編製評量題目過程嚴謹 (developing the assessment questions)；

評量小冊子的完成 (assembling the assessment booklet)；

NAEP 的實施方式非全面性評量所有學生，而是選取足夠且具代表性的樣本學

生來參與 (selecting the sample) ;
鼓勵少數族群學生的參與 (accommodating students with disabilities and/or limited-English-proficient students) ;
被選取要接受 NAEP 評量的學生，於接受評量前，其家長有被事先告知的權利 (informing parents about NAEP) ;
NAEP 評量實施的時間約在一月底到三月中，接受 NAEP 評量應是完全自願的過程 (administering and taking the assessment) ;
NAEP 的評分方式具有信、效度 (scoring NAEP) ;
評量的報告方式以整體學習成果表現趨勢為目標 (reporting student performance) ;
評量結果呈現方式多元，可供資料間相互交叉比較 (putting results in context) ;
評量結果完全公開，有興趣的民眾皆可取得相關資料 (presenting results to the public) ;
建立 NAEP 專屬網站，提供即時最新消息 (NAEP on the Web) ; NAEP 出版品提供豐富、多元且便利的資訊來源。

第二節 英國學生評量的實施策略

一、教師班級評量

英國的教師班級評量是由老師對兒童各方面表現所進行的持續性評量，逐日或逐週的紀錄下來，成爲一種個人的紀錄卡。老師使用的評鑑依據包括兒童的作業、解決問題的能力、參與團體工作的表現等。依據每一個學生的個別紀錄，老師可以安排適合學生的學習活動，並對兒童所達到的水準予以個別評量。

教師每年需把評量結果匯集成一份學童成就報告，此即爲家長所收到的成績單。教師會根據國定課程的標準，核定學生在學科上所達到的等級。一般而

言，七歲的兒童應該達到第二級；11 歲兒童應該達到第四等級；14 歲兒童應該達到第五或第六等級的標準。且此報告需考慮到兒童進步和整個階段的表現，並對照全國等級描述的說明來判斷。

二、全國統一評量

由於老師的評量會受到教師不同經驗及測驗未經標準化的影響，產生無法比較的歧異，因此尚須透過「標準評量工作經驗」(Standard Assessment Tasks, SATs)，在每一個關鍵階段完成後進行全國性的標準測驗。

學校將依學生在三個科目中的學習，與國定課程的這些測驗標準相比，做出評量。測驗後，學校會向家長提出一份成績單，包含教師評鑑結果的報告、學童在全國測驗中的成績，以及與一份同齡學童成績比較表。16 歲學生所參加任何公定的考試 (public examination) 的成績，亦會登載在他們的成績單上。教師有責任在成績發表後，安排與家長見面研討學童的學習狀況。

三、成績評量的登錄與通知格式

評量的登錄依規定，每年必須送交家長知悉，稱之為年度報告 (annual report)，此外學校也必須存檔 (可供學生轉學之用)，因此檔案的登錄是學校的一項重要工作，由於登錄並無一定格式，所以各學校有絕對的自由性，但基本上須注意幾項要點：

1. 紀錄須包括逐年進行的課程及跨科目活動。
2. 紀錄必須顯示學校提供廣泛而平衡的課程。
3. 紀錄必須呈現持續及連續性的說明。
4. 紀錄必須適合不同的學校轉用。
5. 紀錄必須顯示整個班級、小組及個人的活動及評量。
6. 紀錄必須容易處理，盡量不增加老師的負擔。
7. 紀錄必須提供家長、學生及政府有意義的資料。

至於一年一度給家長的年度報告，必須顯示：

1. 學校所提供課程的內容介紹。
2. 各種跨科目活動。

- 3.兒童在各項成就目標或成就水準上的表現。
- 4.指出兒童的優缺點。
- 5.指出學習困難的可能原因。
- 6.說明合理的下一步嘗試計畫。
- 7.紀錄個人特別的成就表現。
- 8.肯定兒童的成就及進步。

全國課程學生成就評量通知方式大可分為二種。一為，統一通知格式；二為，非統一通知格式。

統一通知之格式，主要是用在每一關鍵學習階段最後一年，分別依照學生個人、學校和全國三類表現結果，加以統計並通知學生及其家長。每一類的通知書中，必須包含教師評量和全國性評量，每一學科的整體成就等級。

非統一通知格式，則由地方教育當局或學校自行規範；基本上，格式內容兼具量化記錄和文字描述；量化記錄包括精熟程度、成就等級等，文字描述包括學生學期的學習目標、課程重點、學生表現描述以及優缺點或建議等（陳明印，2004）。

第三節 OECD/PISA 的實施策略

OECD/PISA 代表一個新的承諾，靠著 OECD 國家的政府機構，監控教育系統中學校及學生成就方面的產出。PISA 評量的架構及內涵是國際間形成的共識。評量結果每三年出版，預期這些結果將幫助並激發教育改革與學校進步。特別是針對學校或教育系統有相同之輸入，但結果卻不同的現象，提供檢視的機能。此外 OECD/PISA 也將提供一個國際層級教育系統效能的評量與監控，目前參與的國家已達 30 國，是一項頗為值得重視的超大型研究計畫，尤其是許多寶貴資訊的正確解讀，隱含各種時、空的意義。

OECD/PISA 包含多元評量的範疇，三年一循環補充最新資料，以達到定期的監督學生進步，進而符合關鍵學習的目標。評量對象的年齡群為 15 歲，就義務教育接近結束時進行評估，提供教育系統表現有用的指標。大部分 OECD 國家的教育超過 15 歲，在這個年齡前都有一個比較共同的課程，就此基礎進

行評量相關的知識與技能。知識與技能的定義並不完全以國訂課程為主。

PISA 主要考量的是未來的生活。學校所學的技能與生活所需的技能是不能明確予以區分，但是了解造成這種區分的存在，是非常重要的。OECD/PISA 不排除學校基礎課程的知識與了解，但是評量想了解的是更廣的概念與技能的獲得，以及這些知識的應用。

第六章 結論與建議

本研究探討下列主題：1.各主要國家學生基本能力評量的內涵與優缺點；2.各主要國家推動學生基本能力評量的策略；3.這些學生基本能力評量的呈現方式有何借鑑之處。我國已進行國中階段的「基本學力測驗」，未來如果要全面推動「學生基本能力評量」，可能的配套措施為何？本章就研究的主要結論與建議做一說明。

第一節 結論

一、美國學生基本能力評量之優缺點

(一) 優點與缺點

經過上述 NAEP 內涵之探討，進一步呈現美國學生基本能力評量的優缺點及特色。美國的學生基本能力評量具有下列優點：

- 1.獨立行政組織執行政策；
- 2.標準參照測驗；
- 3.由上而下強制性評量；
- 4.教育經費補助的依據；
- 5.各科評量時間分配適宜，不讓學校過分負擔。

美國學生基本能力評量的缺點包含：

- 1.無國定課程標準可遵循；
- 2.地方教師反彈；

3.評量題目設計的缺失。

(二) 美國的發展經驗可供參考之處

- 1.長期性的執行與規劃；
- 2.閱讀與數學為重點評量科目，國際的比較性評量為主導；
- 3.建立學生能力評量的資料題庫；
- 4.評量題型的架構使長期教育發展趨勢的評量結果可以互相比較；
- 5.專責單位的設立使學生基本能力的掌握持續不間斷；
- 6.重點科目加強評量的密集性；
- 7.學生評量結果的呈現方式有所區分，主要評量（Main NAEP）呈現學生成就的結果，長期發展趨勢評量（Long-term Trend NAEP）提供全國性資料而非學生的個別成就；
- 8.學生基本能力評量的各科目架構，每十年重新檢視更新；
- 9.學生樣本採取抽樣方式，其中包含少數族群的學生代表；
- 10.積極參與掌控學生基本能力的學區可獲較高教育經費的支持。

二、英國學生基本能力評量之優缺點

(一) 優點與缺點

英國發展的評量系統具有下列的優點：

- 1.賦予人民應有的權利
國定課程為所有的兒童，不論是出生、文化、種族、性別、階級、能力以及障礙者，界定了各種完善的學習領域，促進了教育機會的均等，並賦予一個人民所應有的學習權利，期以發展自我，進而自我超越。
- 2.建立一個全國性的標準
國定課程對學生、家長、教師、政府和社會大眾，清楚設定出學生和成就目標的期望。而且對所有學生，在學科中的應有表現，制訂出全國性的標準。這些標準可作為制訂、改進、測量、達成這些成就的指標，甚至作為監測的

評核，以比較個人、團體和學校間表現的參考依據，以了解學生在學習過程中的程度與差異性。

3. 促進連續性和一致性

國定課程引導了一個全國義務教育一致性架構的產生。該架構會促進課程的連續性，而且以充分彈性來確保學生學習的進步。它也會強化學生在各教育階段和各學校的轉換，並提供一個終身學習的基礎。

4. 促進公共大眾的理解

國定課程一致性的課程標準會增進社會大眾對學校教學成效，特別是對義務教育成果的理解。尤其也可藉此提供非教育界人士以及教育專業團體間，包括學生、教師、雇主間，發展出對教育議題的共同討論基礎。

英國在評量實施過程所面臨的問題或缺點方面包括：

1. 教師不願配合的困擾

教師除擔任日常教學外，須依據國定課程要求，必須隨時掌握學生在每一科目的每項成就目標及水準的表現，並加以紀錄。此外還必須配合國定課程的評量，進行標準化測驗，更使老師疲於奔命。加上測驗結果常與老師評量相反，無法幫助老師作出有意義的決定。

2. 信度、效度的問題

- (1) 試題方面：當使用相同的標準或量尺，則無法掌握標準以外的品質；
- (2) 學生方面：學生或學校間，因其主客觀條件不一，使用同一標準加以比較，甚至將其結果公布，實在有失公平；
- (3) 教師方面：無法保證老師評量完全客觀。

3. 評量領導教學的問題

教師專注於學生是否能在考試中得到高分，則會影響老師的教學方式與內容；學習如果只是為了應付考試，則知識常被切割為學習零碎的知識；全國性的評量如果成為主流，則老師平日的行程性評量功效盡失。

(二) 英國的發展經驗可供參考之處

英國的國定課程評量是希望改變以往偏頗的觀念，藉由老師們持續性的紀錄兒童發展以及標準化的國家測驗，提供兒童表現的可靠資訊。此種評量的策略，結合形成性、診斷性、綜合性與評鑑性的功能，有以下的特點：

1. 確認學生已經達到的目標，幫助他們計畫下一個學習目標，也稱為「形成性」的評量；
2. 顯示兒童學習的缺點及困難，幫助其有效克服，也稱為「診斷性」的評量；
3. 給與老師、家長、兒童一個全面性的綜合報告，在固定時間報導學生到達的程度，也稱為「綜合性」的評量；
4. 提供家長、老師、教育當局有關學校表現績效，用以比較學校之間的辦學差異，也稱為「評鑑性」的評量。

三、OECD 學生基本能力評量之優缺點

(一) OECD/PISA 評量的優缺點

OECD/PISA 評量的優點如下：

1. 強調教育系統的評量，可以做為教育系統運作成效的診斷；
2. 評量內涵以生活化能力、終生學習技能的養成為主；
3. 評量具有國際比較的功能，可讓參與國家互相觀摩比較；
4. 評量提供時間序列的比較功能，從 2000 年至 2006 年，其後也已規劃後續的執行計畫；
5. 評量結構嚴謹，評量結果的資訊以專書形式出版，相當透明化；
6. 評量結果資訊的呈現除了顯示平均成績之外，也呈現學生表現的離散程度，具有多元呈現的功能。
7. 評量的結果對各國的教育政策有相當的影響力。

OECD/PISA 評量的問題或缺點包含：

1. 跨國使用的試題，有翻譯上及文化差異上的困擾；
2. 非全面性的評量，閱讀能力、數學能力與科學能力有個別的評量重點、並每隔三年一次，週期有點鬆散。
3. 以抽樣方式進行且侷限義務教育終了階段的評量，比較難達成及時的診斷功能。

(二) OECD/PISA 的發展經驗可供參考之處

- 1.OECD/PISA 是一個跨國合作、比較的典範，所強調的是當前最先進的教育理念，有相當的參考價值。我國於 2004 年加入此一評量計畫，OECD/PISA 評量的運作模式、國際的研討互動，對國內提升學生素質的相關措施將有積極的意義。
- 2.OECD/PISA 採取各主要國家分工合作發展的模式，參與 OECD/PISA 可以與國際接軌，除了有助於國內學者的研究視野之外，也提升台灣的國際能見度。

第二節 建議

本研究以美國 NAPE、英國國家課程評量及 OECD/PISA 為對象，分別探討其中的學生評量內涵、評量策略以及可供參考優劣點。綜合這些分析結果，本研究對於未來國內發展全面性的基本學力評量，主要建議如下：

一、評量機構方面

- 1.慎選發展評量的機構，做長期性、系統性的研究發展；
- 2.採取以校際或是機構分工的設計模式，結合學者專家的專長，可以快速達到分工整合的功效；

二、評量內涵方面

- 1.審慎參考國際的評量架構與評量內涵，做為發展國內的評量系統之參考；
- 2.打破以教科書本為範疇的評量型態，以學習技能及生活技能為評量學生表現的重點；
- 3.建立長期性評量題庫的建置、發展單位。

三、評量實施方面

- 1.初期先以抽樣的方式進行試驗，以建立基本的評量內涵；
- 2.最終以建立全國性的學生評量資料庫為目標。

四、評量結果的呈現

- 1.可以顯現全國性學生表現的資訊；
- 2.可以呈現區域比較及校際比較的資訊；
- 3.評量結果的呈現具有時間序列的意義。

過去台灣師大發展的「基本學力測驗中心」，以國中基本學力測驗為主，目的在做為學生升學之選擇與分發的依據，與國際間發展的基本能力評量工具不同。近年來，國立教育研究院籌備處亦著手進行國內「基本學力評量」的研發工作，惟目前只是研究階段，嘗試建立基本的測量題庫，距離全面推動仍有一段時間。就現階段的發展來看，重新思考國內全面進行基本學力評量的發展策略有其必要。

為今之計，建議教育部先成立「學生基本能力評量任務小組」，配合教育部的政策，就此項任務推動的時程，重新擬定發展計畫，並參考前述建議，做行動策略的規劃。

參考文獻

- 吳知賢 (1996)。英國國定課程剖析及其對我國中小學課程改革的啓示。 **國立編譯館館刊**, 25 (1), 327~347。
- 周淑卿 (1992)。 **英國國定課程之研究**。國立台灣師範大學教育研究所碩士論文, 台北市。
- 林天佑 (2004)。英國中小學教育品質管理策略及方案－近十年教育改革的分析。 **教育研究月刊**, 123, 49~65。
- 陳明印 (2004)。英國國定課程學生成就評量制度。 **研習資訊**, 21 (6), 90~108。
- Binkley, M., & Linndkyla, P. (1997). Teaching reading in the United States and Finland. In M. Binkley, K. Rust and T. Williams (Eds.), *Reading literacy in an international perspective*. Washington, DC: US Department of Education.
- Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.
- Bybee, R. W. (1997). Towards an understanding of scientific literacy. In W. Grabe and C. Bolte (Eds.), *Scientific literacy: An international symposium*. IPN, Kiel.
- Cutler, B. (1992). How is your child doing at school? *British Journal of Curriculum & Assessment*, 3 (1), 21-23.
- Department for Education and Employment (1996). *Results of the 1996 National Curriculum Assessment of 14 Year Olds in England*. London: The Author.
- National Assessment Agency (2005). *National curriculum tests*. Retrieved October 15,2005, from <http://www.naa.org.uk/tests/index.aspx>.
- Millar, R., & Osborne, J. (1998). *Beyond 2000 : Science education for the future*. London: King's College London School of Education.
- National Assessment Agency (2005a). *Key stage 1 (ages 5-7)*. Retrieved October 15,2005, from http://www.qca.org.uk/12818_13017.html.
- National Assessment Agency (2005b). *Key stage 2 (ages 7-11)*. Retrieved October 15,2005, from http://www.qca.org.uk/12818_13016.html
- National Assessment Agency (2005c). *Key stage 3 (ages 11-14)*. Retrieved

- October 15,2005, from http://www.qca.org.uk/12818_13015.html
- National Assessment Agency (2005d) . *At a glance*. Retrieved October 15,2005, from http://www.qca.org.uk/12818_13011.html
- National Center for Education Statistics, National Assessment of Educational Progress (NAEP) . (2005a) . *2005 mathematics assessment: Comparison of the 2005 and 1990-2003 mathematics frameworks*. Retrieved October 5, 2005,from <http://nces.ed.gov/nationsreportcard/mathematics/frameworkcomparison.asp>
- National Center for Education Statistics, National Assessment of Educational Progress (NAEP) . (2005b) . *2005 Mathematics Assessment: Distribution of mathematics questions 2005*. Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/mathematics/distributequest2005.asp>
- National Center for Education Statistics, National Assessment of Educational Progress (NAEP) . (2005c) . *1996 and 2000 science assessments*. Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/science/timedevoted.asp>
- OECD (1999) . *Measuring student knowledge and skills: A new framework for assessment*. Paris: Author.
- The Nation's Report Card. (2005a) . *Overview of the Nation's Report Card*. Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/about/>
- The Nation's Report Card. (2005b) . *What are the differences between long-term trend NAEP and main NAEP*. Retrieved October 5, 2005, from http://nces.ed.gov/nationsreportcard/about/ltt_main_diff.asp
- The Nation's Report Card. (2005c) . *Chronology of National Assessment of Educational Progress (NAEP) assessments from 1969 until 2004*. Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/about/assesshistory.asp>
- The Nation's Report Card. (2005d) . *Schedule for the State and National Assessment of Educational Progress (NAEP) from 2005–2017*. Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/about/assessmentsched.asp>

- The Nation's Report Card. (2005e) . *The Nation's Report Card: Reading*. Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/reading/>
- The Nation's Report Card. (2005f) . *The Nation's Report Card: Mathematics*. Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/mathematics/>
- The Nation's Report Card. (2005g) . *The Nation's Report Card: Science*. Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/science/>
- The Nation's Report Card. (2005h) . *The Nation's Report Card: Science*. Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/science/whatmeasure.asp>
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) . (2005) . *2005 Reading Assessment: What does the NAEP reading assessment measure?* Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/reading/whatmeasure.asp>
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) . (2005) . *2005 Reading Assessment: Distribution of reading question 2005*. Retrieved October 5, 2005, from <http://nces.ed.gov/nationsreportcard/reading/distributequest.asp>
- U.S. Department of Education. (2005) . *U.S. Department of Education Principal Office Functional Statements*. Retrieved October 5, 2005, from http://www.ed.gov/about/offices/list/om/fs_po/osods/os_orgchart.html

國家圖書館出版品預行編目資料

基本能力評量跨國發展經驗之比較研究 / 張鈿富研究主持. - 初版. -- 臺北市：教育資料館, 民 95

面；公分

參考書目；面

ISBN 986-00-5078-3 (平裝)

1. 教育測驗 - 比較研究

521.3

95007914

◎版權所有·請勿翻印◎

發行人：王世英

研究主持：張鈿富

出版機關：國立教育資料館

臺北市大安區（106）和平東路一段 181 號

電話：(02) 23519090

電子郵件信箱：rs@mail.nioerar.edu.tw

出版年月：民國九十五年五月

版次：初版

電子出版品說明：本書同時登載於國立教育資料館網站

網址為：<http://www.nioerar.edu.tw>

工本費：新台幣 250 元

印刷者：承印實業股份有限公司

電話：02-29555282

展售處：五南文化廣場：台中市中山路 2 號

國家書坊：臺北市八德路三段十號 B1

法律顧問：劉祥墩律師

GPN：1009501106

ISBN：986-00-5078-3

統一編號

1009501106



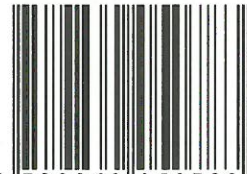
地址：台北市大安區和平東路一段181號

電話：(02)2351-9090

網址：www.nioerar.edu.tw

E-mail：rs@mail.nioerar.edu.tw

ISBN 986-00-5078-3



9 789860 050783