

74-40

## 肆、文獻探討

文獻探討方面，茲就現行臺灣各級學生之學力狀況描述起，而後再參照國內外之學力品管方案，以研擬、建立出本研究品管機制之架構與理論基礎，陳述如下：

### 一、現行臺灣各級學生學力狀況

針對臺灣學生的學生學習成果，我國目前參加了三種國際評比，分別是國際數學與科學教育成就趨勢調查 (Trends in International Mathematics and Science Study, 簡稱為 TIMSS)、學生能力國際評量計畫 (The Programme for International Student Assessment, 簡稱為 PISA) 和促進國際閱讀素養研究 (Progress in International Reading Literacy Study, 簡稱為 PIRLS)。TIMSS 最近一次調查是在 2007 年，PISA 和 PIRLS 則是在 2006 年。一般說來，TIMSS 較偏重學科知識學習成就的評量，PISA 偏重認知能力的評量 (PISA 稱之為科學能力)，PIRLS 是幫助了解當地學童的閱讀技巧。以下茲就臺灣學生於各項評比之學力狀況進行陳述：

#### (一) TIMSS 數學成就評量

國際教育學習成就評量委員會 (The International Association for the Education Achievement, IEA) 成立於 1959 年。其目的主要在了解各國學生數學及科學學習成就及其各國文化背景、學習環境、教師因素等影響因子之相關性；同時，對於連續參加的國家可以進行縱向的趨勢比較，以協助參加的國家瞭解其在教育改革或課程改革的成效 (林碧珍、蔡文煥，2003)。我國自 1999 年開始參加 TIMSS 國際評比，當時主要的評量對象為八年級學生。之後 2003 年和 2007 年也再次參加了四年級與八年級的評量。我國歷年來在 TIMSS 的數學評量成績都表現不錯。TIMSS 1999 八年級學生的成就是全世界第三名 (Mullis, et al., 2000)。在代數的排名為第一名，在分數與數感及資料呈現分析與機率的排名，均為第三名，在測量與幾何的排名，均為第四名。TIMSS 2003 八年級學生的成就仍居世界第四名，在幾何及代數的排名，均為第三名，在數、測量及統計的排名，均為第四名 (張秋男主編，2005)。TIMSS 2007 八年級學生的成就則位居首位，在數、統計、幾何和代數的排名則分別位居 3、4、1、1 名 (Mullis, et al., 2008)。

四年級學生在 TIMSS 2003 世界的排名為第四名，在數、數型和關係及資料的排名均為第三名，在測量及幾何的排名，均為第四名 (張秋男主編，2005)。四年級學生在 TIMSS 2007 的排名為第 3 名。至於數、幾何圖形與測量和資料的排名則分別為第 3, 4, 4 名 (Mullis, et al., 2008)。

在如下 TIMSS 2003 及 TIMSS 1999 八年級數學成就達到各等級國際基準

點人數百分比的趨勢表 4-1 中，2003 年與 1999 年領先群各國達到各等級的人數百分比變化都不大，日本的優級人數百分比，由 1999 年的 29%降低到 2003 年的 24%。由表中之國家資料顯示未達到初級者（未達 400）之學生，新加坡則 1999 年 1%至 2003 年維持不變、臺灣從 1999 年 5%到 2003 年降至 4%、韓國從 1999 年 1%到 2003 年提高至 2%、香港及日本則 1999 年 2%至 2003 年維持不變、匈牙利從 1999 年 7%到 2003 年降至 5%、荷蘭從 1999 年 4%到 2003 年降至 3%、比利時從 1999 年 3%到 2003 年提升至 5%、斯洛伐克從 1999 年 4%到 2003 年提升至 10%；依 1999 年之未達到初級者（未達 400）之學生人數百分比高低其各國之由高至低排序為：匈牙利、臺灣、斯洛伐克、荷蘭、香港、日本、比利時、新加坡、韓國。依 2003 年之未達到初級者（未達 400）之學生人數百分比依其各國之由高至低排序為：斯洛伐克、比利時、匈牙利、臺灣、荷蘭、香港、日本、韓國、新加坡。

表4-1 TIMSS 2003 vs. TIMSS 1999 八年級數學成就達到各等級國際基準點人數百分比的趨勢

國家或地區	優級 (625)		高級 (550)		中級 (475)		初級 (400)	
	2003 年	1999 年	2003 年	1999 年	2003 年	1999 年	2003 年	1999 年
新加坡	44	42	77	77	93	94	99	99
臺灣	38	37	66	67	85	85	96	95
韓國	35	32	70	70	90	91	98	99
香港	31	28	73	70	93	92	98	98
日本	24	29	62	66	88	90	98	98
匈牙利	11	13	41	43	75	75	95	93
荷蘭	10	11	44	47	80	82	97	96
比利時	9	19	47	57	82	86	95	97
斯洛伐克	8	11	31	42	66	79	90	96
國際平均	8	10	28	31	56	57	80	80

依下表4-2的數據顯示，若以達到優級的人數百分比，臺灣居第四位，臺灣只有16%的國小四年級學生列為優級，比例低於新加坡、香港及日本，其它國家依次分別為英格蘭、俄羅斯、比利時、拉脫維亞、立陶宛；依高級基準點排序，其國家之排序分別為 新加坡、香港、臺灣、日本、英格蘭、俄羅斯、比利時、拉脫維亞、立陶宛；從中級基準點排序，臺灣位居第2位，日本則退居第五，其它國家由高至低排列分別為香港、新加坡、比利時、拉脫維亞、立陶宛、俄羅斯、

立陶宛；依初級基準點排序，臺灣與香港和比利時並列，其餘國家分別為日本、新加坡、立陶宛、拉脫維亞、俄羅斯、英格蘭。由表中之國家資料顯示未達到初級者（未達400分）之學生，依人數百分比高低其分別為英格蘭（7%）、俄羅斯（5%）、拉脫維亞（4%）、立陶宛（4%）、新加坡（3%）、日本（2%）、香港（1%）、臺灣（1%）、比利時（1%）。

表4-2 小學四年級學生數學成就達到各等級人數百分比（TIMSS 2003）

	優級（625）		高級（550）		中級（475）		初級（400）		未達 400 （%）
	%	Rank	%	Rank	%	Rank	%	Rank	
新加坡	38	1	73	1	91	3	97	6	3
香港	22	2	67	2	94	1	99	1	1
日本	21	3	60	4	89	5	98	5	2
臺灣	16	4	61	3	92	2	99	1	1
英格蘭	14	5	43	8	75	11	93	11	7
俄羅斯	11	6	41	9	76	9	95	9	5
比利時	10	7	51	5	90	4	99	1	1
拉脫維亞	10	7	44	6	81	7	96	7	4
立陶宛	10	7	44	6	79	8	96	7	4
國際平均	9		33		63		82		

## （二）PIRLS成就評量

促進國際閱讀素養研究（Progress in International Reading Literacy Study，簡稱 PIRLS）是一項由國際教育成就調查委員會（International Association for the Evaluation of Educational Achievement，簡稱 IEA）所實施的閱讀理解評定測驗，這項計畫的目的為研究不同國家教育政策、教師教學之下的四年級學生的閱讀能力表現，而臺灣也於2001年開始參與此項學童閱讀成就的檢測。

PIRLS的測驗結果幫助參與國家與地區了解當地學童的閱讀技巧，並且將這些四年級學生的閱讀素養放置於國際脈絡之下，提供測量和解釋不同國家教育體系之間的差異，藉以檢測學生閱讀學習成就及其是如何受環境相關因素的影響，PIRLS不僅提供參與國家瞭解該如何提升學生的閱讀素養，也可以作為各個國家調整閱讀教學和政策的參考，並進一步協助改善全球的閱讀教學和學習。

如下表4-3，2006年，臺灣學生閱讀成就未達400之人數百分比為3%。若以低於初級（未達400分）之百分比由低至高進行排名，臺灣依序落後於俄羅斯、加拿大、盧森堡、香港、義大利、拉脫維亞、奧地利、比利時等國。

表 4-3 學生閱讀成就達到 PIRLS 2006 各等級國際基準點人數百分比

	優級 (625)		高級 (550)		中級 (475)		初級 (400)		未達400 (%)
	%	Rank	%	Rank	%	Rank	%	Rank	
新加坡	19	1	58	3	86	11	97	12	3
俄羅斯	19	1	61	2	90	2	98	5	2
加拿大亞伯特省	17	3	57	4	89	4	99	1	1
保加利亞	16	4	52	10	82	17	95	19	5
加拿大卑詩省	16	4	56	5	88	6	98	5	2
加拿大安大略省	16	4	54	7	87	8	98	5	2
英格蘭	15	7	48	15	78	21	93	21	7
盧森堡	15	7	56	5	89	4	99	1	1
香港	15	7	62	1	92	1	99	1	1
匈牙利	14	10	53	8	86	11	97	12	3
義大利	14	10	52	10	87	8	98	5	2
紐西蘭	13	12	45	19	76	23	92	23	8
加拿大新斯科夏省	13	12	48	15	82	17	96	17	4
美國	12	15	47	17	82	17	96	17	4
丹麥	11	15	52	10	85	14	97	12	3
德國	11	15	52	10	87	8	97	12	3
瑞典	11	15	53	8	88	6	98	5	2
以色列	10	18	40	23	70	24	85	24	15
蘇格蘭	10	18	40	23	77	22	93	21	7
拉脫維亞	8	20	46	18	86	11	98	5	2
斯洛伐克共和國	8	20	43	21	80	20	94	20	6
奧地利	8	20	45	19	84	15	98	5	2
比利時法蘭德斯區	7	23	49	14	90	2	99	1	1
臺灣	7	23	43	21	84	15	97	12	3
國際平均	7		41		76		94		

### (三) PISA成就評量

有關PISA的成就評量，是包含以下兩者，茲分述如下：

#### 1、國際閱讀素養的評比PISA

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劃，稱為「國際性學生評量計劃」( Program for International Student Assessment, 簡稱 PISA )。近年來已成為國際間為提升各國學生素質所重視的重要評比。PISA 測量 15 歲學生在閱讀、數學、科學等三個領域的素養 ( 林煥祥, 2008 ), 目的不在於知識獲取的多寡, 而在於是否能擷取有用知識、反省思考、解決問題, 進而能面臨二十一世紀進入知識爆炸的社會挑戰。

我國在 2006 年參加 PISA 三個領域的評量, 在 57 個調查國家中, 我國 15 歲學生的數學成績名列第 1, 科學成績排名第 4, 唯獨閱讀成績名列第 16, 比同在亞洲的韓國的第 1 名、香港的第 3 名, 都頗有差距, 這樣的結果使得國內的閱

讀素養教育又再次成爲各方討論的焦點 (林巧敏, 2008; 張淑媚、朱啓華, 2008; 游常山, 2008)。

PISA 將受測學生在閱讀的成績表現，每個面向再分爲一到五的五級素養水準 (proficiency levels)(林煥祥, 2008:191)。每一個層級的素養都有定義內涵 (OECD, 2006)。臺灣與亞洲鄰近受測各國之間閱讀素養水準的差異表現，由下表 4-4 可以得知，臺灣受測學生成績達到「層級五」的比重僅佔全體的 4.7%，遠落後於韓國 (第一名) 的 21.7%，香港 (第三名) 的 12.8%，日本 (十五名) 的 9.4%。同時，臺灣學生達成「層級四以上」(「層級四」與「層級五」) 的受測學生僅有 26.3%，也與韓國 54.4%、香港 44.8%、日本 30.9% 有一段差距。第四級、第五級閱讀能力是攸關未來國家競爭力的重要指標，如何提升臺灣閱讀素養水準，相信是未來教育方向要加強的重點領域。臺灣學生僅有 3.8% 低於層級一 (低於 334.75 分)，韓國爲 1.4%，香港 1.3%，日本 6.7%。顯示臺灣學生的閱讀素養仍有很大的提升潛力。

表4-4 臺灣與亞洲鄰近受測各國之間閱讀素養水準的差異表現

	臺灣	日本	韓國	香港
低於層級一 (低於334.75)	3.8%	6.7%	1.4%	1.3%
層級一 (334.75~407.47)	11.5%	11.7%	4.3%	5.9%
層級二 (407.47~480.18)	24.4%	22.0%	12.5%	16.5%
層級三 (480.18~552.89)	34.0%	28.7%	27.2%	31.5%
層級四 (552.89~625.6分)	21.6%	21.5%	32.7%	32.0%
層級五 (625.61分以上)	4.7%	9.4%	21.7%	12.8%

## 2、PISA 數學素養評量

PISA(林煥祥主編 2008)，是由經濟合作暨發展組織(Organization for Economic Co-operation and Development, 簡稱 OECD)所委託的計畫，於 1990 年代末期開始對 15 歲學生的數學、科學、及閱讀進行持續、定期的國際性比較研究。每次評量會從數學、科學及閱讀三個領域中選擇一個主要領域，進行深入評量，另外兩個領域仍會進行施測，但涵蓋的面向較少。PISA 2000 年主要領域爲閱讀，2003 年爲數學，2006 年爲科學。我國則於 2006 年第一次參與此項調查計畫，這一年參與的國家共有 57 國。

PISA 與一般調查針對特定學校學科的能力有所不同，在數學方面，PISA 強

調數學素養的評量，希望有助於反映未來課程改變為知識應用取向，而非只是知識的獲取。數學素養之意為「個體能夠辨認和瞭解數學在世界上所扮演的角色，能夠進行有根據的評斷，並且針對個體在生活中的需求來運用或者投入數學活動，以成爲一個積極的、關懷的、以及反思的國民」。

在下表 4-5 中有關 PISA 2006 各國在（數學）各素養層級水準之百分比及排名表中顯示出，臺灣有 3.6% 學生未達成層級 1（未達 357.8）。OECD 國家中芬蘭和韓國等國家該國學生未達成層級 1 之人數百分比均低於 10%。若以低於層級 1 之百分比由高至低進行排名，臺灣名列第 45，依序落後於日本、香港、芬蘭、韓國、亞塞拜然、愛沙尼亞等國，顯示後續的教育決策上對於低分群部分的數學學習應該值得關注。

表 4-5 PISA 2006 各國在（數學）各素養層級水準之百分比及排名

	層級 6 (>669.3)		層級 5 (607.0~669.3)		層級 4 (544.7~607.0)		層級 3 (482.4~544.7)		層級 2 (420.1~482.4)		層級 1 (357.8~420.1)		低於層級 1 (未達 357.8)	
	%	排名	%	排名	%	排名	%	排名	%	排名	%	排名	%	排名
瑞典	2.9	18	9.7	22	20.1	19	26.0	16	23.0	27	12.9	36	5.4	39
葡萄牙	0.8	38	5.2	34	14.2	35	23.3	31	25.3	12	18.7	20	12.4	21
加利亞	0.6	40	2.5	42	6.7	42	14.9	42	22.0	30	23.9	13	29.4	10
捷克	6.0	7	12.3	14	19.1	23	23.0	35	20.5	36	11.9	40	7.2	32
日本	4.7	12	13.4	11	24.3	6	26.3	11	18.3	46	9.6	47	3.4	48
香港	9.0	3	18.7	2	25.6	2	22.7	36	14.4	53	6.6	55	2.9	51
特內格羅	0.1	45	0.8	48	4.0	49	11.8	48	23.3	25	28.4	4	31.6	9
希臘	0.9	37	4.2	38	12.6	38	23.2	33	26.8	4	19.0	19	13.3	19
臺灣	11.8	1	20.1	1	22.4	14	19.4	38	14.3	55	8.3	52	3.6	45
達	0.1	45	0.5	52	1.4	56	3.3	55	7.5	56	15.5	25	71.7	2
國	0.1	45	1.2	45	5.4	46	14.1	44	26.0	8	29.7	1	23.5	16
亞塞拜然	0.1	45	0.8	48	6.5	44	34.5	1	47.6	1	10.4	42	0.2	58
沙尼亞	2.6	21	10.0	20	23.3	8	30.2	2	21.9	31	9.4	48	2.7	53
韓國	9.1	2	18.0	4	25.5	3	23.5	30	15.2	54	6.5	56	2.3	56
OECD verage	3.4		10.1		9.1		24.3		21.9		13.5		7.7	

## 二、國內現行學力品管機制或其它措施

綜觀國內有關學力品管或學力檢測之方案，十分多元，但細究時，多可分爲兩大類，其一是學力成效的評估，即是以評量、診斷學生學力表現爲核心；其次，

則是以執行補救教學或扶助弱勢為重點之方案，茲分別簡介如下：

### (一)國內學生學力成效評估機制簡介

國內針對學生學力進行成效評估研究，除國家教育研究院籌備處所執行之臺灣學生學習成就評量資料庫（Taiwan Assessment of Student Achievement, TASA）外，以及各縣市政府所自行研發之學生學習能力測驗亦屬之。主要目的除了建立學生學習基本能力資料外，更可檢視並了解目前臺灣地區學生學習成效，進而作為市政府教育局及學校推動補救教學之重要參據。以下分別就 TASA 以及各縣市所研發之學習能力測驗進行簡要描述：

#### 1、臺灣學生學習成就評量資料庫

教育基本法第九條明定中央政府之教育權限，該條第六項為：「教育統計、評鑑與政策研究」。因此，透過評量可以用來瞭解學生在各學習領域（科目）的表現，評鑑學生的學習優劣。所以教育部於九十三年四月二十日國民中小學九年一貫課程推動工作小組第四十三次會議決議：「有關學生學習成就調查是研訂課程與教學政策之重要參據，有必要進行常態性之資料建立」。據此，教育部乃於九十三年函請國家教育研究院籌備處針對國民中小學學生學習成就建立常態性之資料庫，做為研訂課程與教學政策之重要參據。本處旋即規劃自小四、小六、八年級至高中職二等就學階段學生，於國語文、英語文、數學、社會與自然等學科之臺灣學生學習成就評量資料庫之建置計畫，而其根本目的在於除追蹤、分析學生在學習上變遷之趨勢外，更為瞭解國內學校教育之現況，作為教育部研訂課程與教學政策，縣市政府教育局及學校推動補救教學之重要參據(臺灣學生學習成就評量資料庫網站，2004)。

自 2005 年 TASA 首次針對上述各年段學生進行學力檢測以來，至今亦累積不少成就分析結果，顯示臺灣學生學習成就資料庫確可以有效監控學童學習成效，對於學習狀態亦可進行深入的分析，了解學生學習的弱勢。

#### 2、苗栗縣 96 學年度國民中小學學生學習成就診斷補救計畫

苗栗縣提出的「國民中小學學生學習成就診斷補救計畫」，目的在了解苗栗縣各區域（各校）學生之學習表現，以提供縣內研究各區域影響中小學學生學習表現因素之實徵資料，此外可診斷各校學生學習情形，給予積極性的補救教學措施，協助學生有效學習。而其提升學習弱勢學童學習成就策略之實施流程如下圖 4-1 所示(彭富源，2007)，強調著整合教育部專案計畫資源(如現有之教育優先區計畫、新近實施的攜手計畫－課後扶助與精進教師課堂教學能力計畫)，透過教育局輔導團的規劃，多頭並行，並從各年級檢測試題的命題、審題、修改與複審，成績登錄資料庫的建置與觀念宣導，到施測結束後的學習成就檢測資料統計分析，訂定補救教學標準、篩選受輔學生，以及各領域補救教學策略彙整。

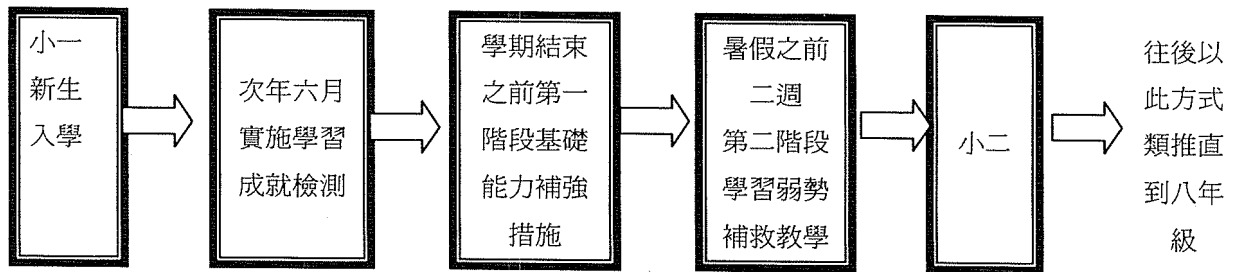


圖 4-1 苗栗縣實施補強與補救教學流程圖

該學習成就診斷補救計畫乃針對全縣學生之國語、英語、數學、社會、自然此五大領域進行施測，而該測驗是屬普測，苗栗縣內除啟智班、在家教育巡迴班以及在家教育學生，其餘國中小 1 至 8 年級學生皆須參加本測驗(苗栗縣國民中小學學生學習成就診斷補救計畫網站，2009)。施測題目來源為國中以基本學力測驗之題目外加國教輔導團(國中)研擬之新題。國小題目主要以國教輔導團(國小)研擬之題目為主。實施的學習領域及年級詳見下表 4-6：

表 4-6 苗栗縣國民中小學學生學習成就診斷補救計畫施測年級與科目表

領域	階段別							
	國民小學						國民中學	
	一年級	二年級	三年級	四年級	五年級	六年級	七年級	八年級
國語文	檢測	檢測	檢測	檢測	檢測	檢測	檢測	檢測
數學	檢測	檢測	檢測	檢測	檢測	檢測	檢測	檢測
英語	---	---	檢測	檢測	檢測	檢測	檢測	檢測
社會	---	---	---	---	檢測	檢測	檢測	檢測
自然與生活	---	---	檢測	檢測	檢測	檢測	檢測	檢測

### 3、臺北市舉辦國民小學基本學力檢測

從 94 年起開始實施國小六年級學生基本學力檢測，檢測結果作為教育部政策制定及強化教師教學參考。並藉以了解九年一貫的實施成效。對象為臺北市國小六年級學生（不含在家學習之學生）。國語文及數學領域採選擇題答方式，為普測。至於國語文及數學領域中，開放性試題部份則是採抽測。施測結果將公佈平均學生之通過率，以及全市指標通過率之對照圖(臺北市政府教育局，2008)。

### 4、臺南縣國民小學低年級學生「快樂學習，多元評量」實施計畫

臺南縣自 97 學年度（97 年 8 月 1 日），針對國民小學低年級學生提出「快樂學習，多元評量」實施計畫，其目的主要是調整低年級學生學習評量的方式，透過多元評量達成學生快樂學習及潛能開發的目標，並藉以落實教師教學的專業發展，提升教學品質。

該計畫實施對象為全縣國民小學一、二年級學生，實施的領域有語文學習領



域（本國語文）、數學學習領域、健康與體育學習領域、生活課程、綜合活動學習領域。該計畫特別之處在於取消各領域定期評量紙筆測驗，由各領域教師規劃平時評量，以多元評量為原則。各學習領域評量方法、內涵與等級區分標準，應由各校課程發展委員會及任課教師審酌教學需求及學生日常表現自訂，並於學習初以口頭或書面通知等方式向學生及家長說明。各領域任課教師，應運用多元評量方法，詳加紀錄學生學習情形，且於學期進行中，如有必要應隨時以書面通知向家長傳達學生之學習歷程。

該計畫之預期效益在於提供學生在不同領域中，多元的、適宜的評量方式。落實多元評量精神，積極促進教師教學專業發展，確實提升教學品質，以及透過多元評量的實施，營造學校最佳學習環境，達致學生快樂學習的目標。

綜合上述，茲將相關訊息濃縮如下表 4-7 所示，此外，除上述縣市外，其它各縣市有關於學生學科能力檢測相關計畫，亦歸納如下表 4-8：

表 4-7 臺北市、苗栗縣、臺南縣提升學習能力計畫比較表

內容	縣市別		
	苗栗縣	臺北市	臺南縣
計畫名稱	國民中小學學生學習成就診斷補救計畫	國民小學基本能力測驗	國民小學低年級學生「快樂學習，多元評量」實施計畫
實施領域	國語、英語、數學、社會、自然等五個學習領域	語文（含本國語文及英語）、數學領域之總結性紙筆測驗。	語文學習領域（本國語文）、數學學習領域、健康與體育學習領域、生活課程、綜合活動學習領域
對象	全縣國中、小 1 至 8 年級學生	臺北市國民小學 6 年級學生	全縣國民小學一、二年級學生
預期效益	作為學生學習成就診斷補救依據	作為學生學習成就診斷補救依據	1.提供學生於不同學習領域中多元的學習評量方式。 2.落實多元評量精神，促進教師專業發展，提升教學品質。 3.營造學校最佳學習環境，達致學生快樂學習目標。

表 4-8 各縣市學生學科能力檢測相關計畫之比較表

縣市	執行計畫名稱	實施領域	對象	預期效益
臺北市	國民小學精進課堂教學能力計畫	小一至小五之語文(含國語及英語)、數學領域之總結性紙筆測驗	小六	分析測驗結果做為教學之參考
臺北縣	學生能力檢測	國語、英語、數學	小一至九年級(每年段擇一科目抽測)	
宜蘭縣	進行國小學生基本學力測驗命題與施測		小六升國一	了解學生起點行為並提供各校進行補救教學
苗栗縣	國民中小學學生學習成就診斷補救計畫	國語、英語、數學、社會、自然	小一至九年級	作為學生學習成就診斷補救依據
臺中市	國民小學數學能力測驗	數學	小四	
彰化縣	國民小學學生學習成就評量	國語、英語、數學	小六	1.提供教育當局研擬課程及教育政策之參考。 2.瞭解本縣國民小學學生學習之基本能力及九年一貫課程實施的成效。
雲林縣	國民中小學學生線上學習成就評量實施計畫	數學	小五及八年級普通班學生(抽測)	1.瞭解本縣學生學習狀況一般情形及城鄉學校學生能力差異。 2.依據診斷評量結果，以作為研擬相關數學教學措施的依據及參考。
臺南市	國民小學改進學習評量實驗計畫			1.改進學習評量方式，提升學生學習成效。 2.促進教師教學改善，提升教學品質。 3.營造學校最佳學習環境，使學生樂於學習。
臺南縣	國民小學低年級學生「快樂學習，多元評量」實施計畫	語文(本國語文)、數學、健康與體育、生活課程、綜合活動	全縣國民小學一、二年級學生	1.提供學生於不同學習領域中多元的學習評量方式。 2.落實多元評量精神，促進教師專業發展，提升教學品質。

### 5、教育部攜手計畫

教育部為協助國民中小學(以下簡稱各校)落實弱勢低成就學生學習輔導，特擬定「攜手計畫」針對學習需要幫助之學童，提供課後扶助措施。篩選資格除國中基測提升學校得斟酌校內情形自訂外，一般學校須兼具下列二種情形：

(1)具有下列身分之一：

- a.原住民學生。
- b.身心障礙人士子女。
- c.外籍、大陸及港澳配偶子女。
- d.低收入、中低收入家庭學生及免納所得稅之農工漁民子弟。
- e.隔代教養及家庭失功能子女(含單親)。
- f.身心障礙學生(含經鑑輔會鑑定為疑似身心障礙學生)。但另有其他輔導方案者(如已接受資源服務者)，以不重複服務為原則。
- g.其他因文化不利、重大傷病、懷孕等因素，確有學習輔導需求，經學習輔導小組會議審查通過者。

(2)學習成就低落需補救者：

- a.都會地區以單一學科班級成績後段百分之二十。
- b.非都會地區則以單一學科班級成績後段百分之三十五為指標。(都會地區指直轄市、省轄市及縣轄市。)

九十八年五月由直轄市政府教育局、縣(市)政府(以下簡稱各縣市)，定期規劃辦理全縣(市)學習評量說明會，邀請各校承辦組長參加，加強評量教育訓練及資格篩選事宜說明。並由各校成立學習輔導小組，規劃學習輔導相關事宜，並辦理全校說明會(或配合全校集會加強宣導)，讓全校教職員充分瞭解相關內容，務使需要扶助學生均能獲得課後扶助資源。

參與評量的方式為每年六月由各校各班級任老師或國語文、數學、英語、自然與生活科技、社會等學習領域任課教師或由家長提出申請。申請時應填寫申請表並取得家長同意書，送校內承辦單位進行初步篩檢，篩檢通過者得接受測驗。測驗內容主要以臺南大學所研發之電腦化測驗之國語文科、數學科評量或其他科目實施測驗。依據測驗結果交付各校學習輔導小組會議，進行綜合研判定學童學習安置型態。

補救教學有兩種方式可以進行：

- (1)於課餘、課後時間及寒暑假進行外加式補救教學：以課餘時間(含早自習)進行為原則，午休不實施，惟學校有特殊情形，專案報經縣市政府同意者，不在此限。
- (2)於正式課程時間進行抽離式補救教學：須經評估正式課程時間抽離實施效果

較佳，確有必要抽離實施者，須經學習輔導小組會議通過、家長同意後為之。

補救教學的內容在國語文與數學科一至九年級均可實施；英語科三年級以上始得實施；自然、社會：七年級以上視需要實施。補救教學教師應就學生起點行為進行學習診斷，設定適合學生需求之個別化學習目標，並依學生能力進行分組教學，提供個別化之學習輔導，並建立學生個人學習輔導歷程檔案紀錄，做為長期輔導追蹤之參考依據。

## (二)國內學生學力成效評估機制執行成果分析與建議

雖然國內現存如此多元的學力成效評估方案，但其間仍存在許多問題，需進一步分析與探討。其中，對於這類學科能力測驗，臺灣已有許多縣市進行，實施的內容、程序、對象以及效益皆有不同，可說是一人一把號，各吹各的調。未有基本的原則與認識，所以本研究將對學科能力測驗訂出基本的原則作為外來實施的依據。其次本研究檢視臺北市、苗栗市與臺南縣相關計畫後進行分析，主要目的是提供中央教育主管機關對於學生學科能力檢測相關措施進行審思，針對實施內容在行政工作、測驗專業與未來發展提出若干問題與建議。

### 1、學科能力測驗應具備之原則

- (1)回饋性：測驗結果報告，不單只是學者研究的工具，更不是束之高閣的文本。它必須能回饋於接受施測的學童，落實測驗的教育責任。
- (2)立即性：如果施測結果不能立即回饋，也是徒然。就目前我國現行教育體制觀之，即使前一個單元未能學會，必定會強迫學童進入下一個學習單元。如此惡性循環，對於學習效果必定打上問號。如果施測結果可以展現立即性，除了可以使學校和教師清楚的知道哪些學生需要補救，更能立刻將報告結果回饋於學生。
- (3)專業性：報告的分析需要專業，透過專業除了知道哪些學生需要補救，甚至是科目、單元，都能透過測驗清楚掌握學生學習狀況。

### 2、學科能力測驗現況分析與建議

有關現況分析與建議方面，茲就幾方面進行說明，分別是行政事務工作、測驗專業發展與未來發展探析，說明如下：

#### (1)行政作業未有常設性組織

各縣市計畫多屬於短期（多半為一年期非長期性）計畫，所以在組織編制上未見常設性單位進行計畫執行。以現行的計畫觀之，執行單位多屬所在地區學校組成的臨編組織，未有固定單位進行。如未來，學科測驗能持續發展，行政的銜接與效率尚有待考驗。

## (2)學校行政人員與教師功能有限

平日學校的相關事務已讓行政人員十分忙碌，教師也爲了教學與學生照顧上疲於奔命，如果再參與相關的行政事務，顧此失彼。所以對於施測規劃等相關事務性工作所能提供的幫助有限。短期內或許還有若干熱誠人士參與，長期看來會將相關的行政事務將會被學校行政人員與教師視爲燙手山芋。

## (3)試題品質與難易度未能掌控

現行測驗方式是以各縣市所提供之題庫進行施測，但試題品質未受檢驗，組卷由學校教師自行選題，如此一來試題層次是否會依選題者的喜好有所不同(如選題者較偏好記憶題，忽略理解題)，以及測卷是否符合評量架構之基本原則等等，上述狀況將會使測驗的專業性未能彰顯。

## (4)測驗結果分析專業性有待評估

延續上述，在試題品質上未能有效控管，在後續分析上也頗令人憂心之慮。施測分析結果主要目的是精進課堂教學，如果只單以分數結果了解學生學習狀況，對於精進課堂教學效果有限。以數學爲例進行說明：兩個學生同樣都得 80 分，但是其中一個是錯在幾何類型題目較多，另一位是錯在代數類型較多，但缺乏專業分析，是無法得知結果，亦無法針對個別的學生進行合適的補救教學，使得施測的功能將有所限制。

## (5)試題數量是否能滿足測驗需求

目前各縣市依據該縣市所提供題庫光碟進行組卷施測，但試題未來如不能持續發展，必定面臨試題重複使用，不但使測驗的品質較有疑慮，也讓分析結果變得不可觀，進而影響補救教學的正確性。

## (6)試題未針對各年級實施

針對臺北市、苗栗縣與臺南縣的現況進行分析，施測對象臺北市爲國小六年級學生、臺南縣爲國小一二年級學生，只有苗栗縣是一到八年級學生。由上述得知，除苗栗縣外，另外兩縣市只針對部分的學生進時施測，未能普遍性的進行，也就是地方教育行政機關無法有效掌握所在地區學生的學習狀況，缺乏施政的參考。

綜合上述，就現行學力成效評估機制而言，本研究建議由中央教育主管機關統籌規畫有關學生學科測驗之常設單位。進行行政規劃、試題研發與試題分析，以求試題品質的穩定與具備專業的分析結果，並提供受測學校補救教學內容與策略，使精進課堂教學計畫與學生補救教學更能有效實施。

## (三)國內執行補救教學或扶助弱勢方案簡介與成果分析、建議

於學力成效評估之外，對於學習落後之學童，最重的要則在於能有效的補救

與扶助，而在國內相關的方案中，雖然有如教育優先區或夜光天使專案等方案之推動，但當是以現行教育部積極著手進行的國民中小學之「攜手計畫」(<http://asap.moe.gov.tw>)，最具整合性、廣泛性與持續性。該計畫主要規劃對於學習落後之學生實施補救教學，使弱勢及學習成就不佳之學生，有提升學習成就之機會。而該計畫的實施，也成為國內國民中小學補救教學之實施依據，所以對於提升學生學習成就扮演著重要角色。

行政院於 2005 年核定教育部所提出之「攜手計畫」，該計畫主要目的為提供弱勢族群學生課業輔導，以縮小國民中小學學習成就低落學生之學習落差，並提供學習成就低落國中學生適性分組學習，以及多元學習方案。在該計畫方案中，已詳細規定各類實施要點，包含受輔對象的篩選方式、教學人員的資格、補救教學課程的內容與實施方式、考核機制等(教育部攜手計畫課後扶助網站，2009)。而為調查該方案是否有需精進之處，本研究曾另以問卷調查方式，針對國中小教師就攜手計畫之執行內容進行探討，探討項目包含「受輔對象」、「新增受輔對象篩選的方式」、「執行補救教學人選」、「配套措施」、「補救教學課程的內容適當與否」、「是否已達到預期提昇成就低落學生學力的成效」、「攜手計畫是否需要改進」等。最後，得總有效國中小教師樣本人數為 1933 名(詳細抽樣過程請見研究方法中之問卷調查)，而其分析結果，可提供中央主管機關作為改進攜手計畫之參考，在此僅簡要說明，以作為背景訊息，詳細可參考陳伯璋、林世華、謝進昌、陳清溪、曾建銘、林宜臻、蔡明學、謝佩蓉、周慧玲、金冠宇(2009a)之研究；研究中可以發現受訪教師認為受輔對象的篩選，應以客觀的測驗成績為依據，並加入教師對於學生的課堂表現進行判斷，而執行補救教學的教師應以受過教育專業訓練與實際經驗者為主，如果是現職教師，在配套措施方面受訪者認為，希望透過「降低原授課班級人數」、「降低授課時數」，以及「提供合理津貼」等相關方式進行。而補救教學課程內容可以由各校自訂，這樣較具教學彈性。不過受訪者同時也希望中央可以提供教材，或是訂立課程標準。最後在成效評估上，也希望中央能訂立一套評估標準，使學校與教師對於補救教學有明確的方向與指引。

### 三、國際現行學力品管機制

在評閱國際現行學力品管機制方面，本研究茲以美國、英國、日本與中國之品管架構為標的，探討如下：

#### (一)美國沒有落後的學生法案

在美國，教育原本是地方權責，但在中小學的部分，則有(美國)聯邦政府在 2001 年通過的沒有落後的學生法案(NCLB)法。NCLB 法案的立法，顯示(美國)聯邦政府對中小學的教育介入較深。其中，楊巧玲(2007)整理了美國教育的核心價值，由「公平」到「卓越」到現今 NCLB 法案的「標準本位」。各州政府各自擬定有一套因應 NCLB 政策中績效任務系統之內涵與執行的政策(例如張奕華

(2004)整理之 NCLB 下政府的密蘇里州的反應機制)或執行高中會考。大學教育幾乎是各校或各州(公立大學院校)自主，所以評鑑很少是全國性，許多評鑑措施，主要由學校自行決定，只有公立大學才會受州政府之一些規範與約束(彭森明，2008)。

NCLB 的緣由，係因(美國)國家教育統計中心(NEAP)的資料指出：非裔與拉丁裔十七歲的學生，其閱讀與數學的水平，僅與十三歲白人學生相等。因而，NCLB 法的目標，是預計在 2014 年前將此「學業成績的差距」加以接合。其中，NCLB 訂明：

- 1、每家學校必須到 2014 年，提升每位學生之數學和閱讀水平至其就讀年級水平(熟練)。
- 2、每家學校必須展示其在為每個子群組之目標，每年有足夠的進展(AYP)—這包括五個最大的族裔組，英語學習者，傷殘人士，和低收入學生。
- 3、接受聯邦政府 Title One 資助之高貧學校，如果在兩年內沒有達到 AYP，則必須支付補習費用或提供交通以便如學生想轉往別的學校。
- 4、同樣的學校如果學校在六年內沒有達到 AYP 要求，則必須重新組織，教職員換新，或作出一些大型之改變。

如今有關學生學力表現的相關統計報表，也公佈在 National Assessment of Educational Progress(NAEP)的網頁(<http://nces.ed.gov/nationsreportcard/nde/>)中，簡潔而有力的讓執政者與教育者很快的瞭解其中的訊息，例如在 NAEP 2007 年 Digest of Education Statistic 中，發表出兩個有關閱讀與數學公開的教育統計報表，隱約可從 1999 年的數據中，看出為何 2001 年會通過 NCLB 法案訴求的重要統計現象。以取自 NAEP2004 的資訊中，在 1999 年的數據中，可發現 17 歲的黑人在閱讀方面的平均表現是為 264(標準差 1.7)，而 17 歲的西班牙裔兒童的平均表現則為 271(標準差 3.9)，雙雙都只達 13 歲白人兒童的水準平均值為 267 分(標準差 1.2)；而在數學表現中，亦有類似的現象，顯示出非裔與拉丁裔十七歲的學生，其閱讀與數學的水平，只達到十三歲白人學生的學習成就，落後於同年齡的白人學生(NAEP, 2004)。在此背景之下，故到了 2001 年，NCLB 法案極欲想解決這種現象。然而，反觀 2004 年的數據，此種現象還沒有顯著性的改善，也許 NCLB 法案到 2004 時的實行的效應還沒有產生，如此，預期到 2014 年的報表中，能藉由這樣的統計資料，便可以明確的指出，NCLB 的政策下產生何種成效及是否達成其預定的目標。

美國在教育普及化之後，所有學童都可上學，由於地區與家庭環境差異等因素，造成學生個別差異，而各地學校教學標準與要求不一，因此學生學習成就參差不齊，若沒有統一標準，程度很低的學生也可以畢業。因此為了確保所有學生達到某種程度與標準才能畢業，各州除了要求學校教師加強教學與輔導，要求學生通過各科考試外，還可能要求學生通過全州統一會考。聯邦政府推動的 NCLB

法案也鼓勵各州推行高中畢業資格考試(High School Exit Exam, Graduation Exit Examination,..)，達到標準的學生才能畢業。目前美國約有半數的州有此規定(Department of Education and Early Development, Arkansas 2008)。畢業資格考試項目各科不大一致。以路易斯安那州(Louisiana)為例，其畢業資格考試(Louisiana Graduation Exit Examination)包括五科目：(1)閱讀與寫作知能(language arts)(2)作文(written composition)(3)數學(mathematics)(4)科學(science)(5)社會科學(social studies)( State of Louisiana,2008)。但加州只要求考閱讀與寫作知能(English language arts)數學(mathematics)和初等代數(algebra I)。

MCAS，1993年麻州通過「教育改革法(Education Reform Act)」，規定高中10年級學生必須參加「麻州綜合測驗(MCAS, Massachusetts Comprehensive Assessment System)」的英文與數學兩科，高中生除必須符合各學校區的畢業規定之外，尚必須通過這兩科會考之標準才可以獲得高中文憑(High School Diploma)，學生若未通過這兩科會考之標準，仍可以畢業，但不能獲得高中文憑。

麻州教育綜合測驗(MCAS)以實行十餘年，期間即使有不同意見，例如：某些教師反應增加教學負擔，有些教育學者批評考試干擾教學及有些能力不是考試可以測出，但麻州教育廳基於大多數家長的支持及為融合聯邦政府 No Child Left Behind 的政策，因此維持這項制度，並準備在2010年以後，高中生要獲得高中文憑必須加考自然科中的生物化學、物理，或工程科技中之任何一科。

「麻州綜合測驗(MCAS)」主要是根據麻州課程大綱(Curriculum Framework)的學習標準(Learning Standard)來評量學生的表現。評量結果分學生個人、學校、學校區三個層級作報告分析。麻州教育當局認為，MCAS可以有以下用途和功能：

- 1、改善教學如追蹤學生學習進展，如蒐集診斷性資訊、找出需要額外協助或補救教學的學生、發掘出目前課程和教學的優缺點。
- 2、學校和學校區績效責任：教育委員會(Board of Education)可以訂出改善的標準來看各校及各校區是否持續改善其教學成效。
- 3、學生的績效責任：學生提早被告知要通過英文和數學會考才可拿到高中文憑，又有重考之機會，因此此一會考也有激勵學生自我負責的功用 (駐波士頓文化組，2008)。

因應美國聯邦政府 NCLB 法案，各州政府皆擬定有一套 NCLB 政策中績效任務系統之內涵與執行的政策，也紛紛建立屬於各州的標準化測驗，以衡量學校的績效。由於 NEAP 數學與閱讀約隔四年會測一次，可監測全國及各州的成績與三個年級的表現，然而並不敷各州的要求，因而，各州紛紛開始建立標準化測驗，自增加測驗頻率，以至每年測驗並增加測驗的年級。此外，亦有某些學校採重視閱讀與數學的策略而排擠到社會領域的時數。

NCLB 法案要求每家學校必須展示其在為每個子群組之目標，每年有足夠的進展(AYP)，而這包括五個最大的族裔組，英語學習者，傷殘人士，和低收入學生。此從 NAEP2004 的統計數據可以看出，非裔與拉丁裔的平均分數，在低年級



是有進步的。然而，這個要求易使位於這類族群豐富的學區學校，很容易被不公平的評為不好的學校。

有關接受聯邦政府 Title One 資助之高貧學校，如在兩年內沒有達到 AYP，必須支付補習費用或提供交通如學生想轉往別的學校。這個部份 NCLB 實行以來，已有部份學校接受到這項處分的實施，然而移校就讀的效果如何，尚還有待追蹤研究的評估。至於同樣的學校如果他們在六年內沒有達到 AYP 要求，必須重新組織，教職員換新，或作出一些大型之改變。這項因為實行年限的關係還在觀察中。但有意見顯示應該把學校選擇與補充服務，應針對不能達到 AYP 的特定團體學生，而非把整個學校標籤為「失敗」。

美國的法案有充分自我監督機制，可以對下一次法案的重新授權作準備，但此部份在臺灣是相當薄弱甚至缺乏的(楊巧玲,2007)。本研究期望藉由美國 NCLB 以說明，一般需要追蹤或監控的統計數據，通常也是政策所關切的焦點。美國可以借鏡的地方是其標準化測驗的精神與長期追蹤的教育統計資料。而臺灣也可以確認我們想要的教育核心精神與重視哪些議題(縣市或特殊族群等..)並加以追蹤。臺灣自 2005 年開始建置臺灣學生學習成就資料庫 TASA，係屬於標準化測驗的精神，目前已在建置的階段，期許將來亦可建立具有貫穿各年級間的一把量尺教育報表，以供執政者與教育者參考。

## (二)英國評量架構

### 1、英國教育現況、目的與起源

英國在 1944 年的〈教育法〉中確立了教育的三個階段：初等教育(5~11 歲)、中等教育(11~15/16 歲)、繼續教育(16 歲之後，包括高等教育)，其中，中學教育階段屬於免費義務教育，學生可根據其年齡、程度及性向進入文法學校(grammar schools)、現代中學(secondary modern schools)、或技術學校(technical schools)等就讀(周愚文, 2008, p.51)。而根據英國政府的官方統計，2007 年 1 月英格蘭小學班級的平均生師比是為 26.2:1；在教學上，小學通常為不分組教學，沒有能力分班，僅少數科目或部份教學方法採班級內分組進行的方式。

在課程方面，是於 1988 的〈教育改革法案〉開始採用「國定課程」(National Curriculum)、並首次定義了義務教育中的「關鍵階段」(key stages)，其下主要分為 4 個階段，而後國定課程評量(National curriculum assessment)即依各個階段，設立能力標準，建立起各年級評量的基礎(DCSF, 2008)。

在權則分配方面，由於英國是屬地方分權，評量的方向與原則是由中央訂立，但實際作法與細節(如教師評量)則由各學校自行決定，搭配政府與家長的監督以建立起國家的評量網絡。

### 2、評量架構

英國的評量體系，主要是在義務教育的時程中，分成 4 個主要的關鍵階段，如表 4-9 所示，而每個階段下都有其重點評量科目與評量方式，其中評量科目是

由國定課程中所組成，而評量方式主要區分教師評量與國家測驗兩種。以下茲分初等教育與中等教育分述之：

表 4-9 英國各關鍵階段之評量方式一覽表

年齡	階段	年	評量方式
3-4	學前階段		
4-5			
5-6	Key Stage 1	Year 1	
6-7		Year 2	英語、數學、科學的教師評量
7-8		Year 3	
8-9	Key Stage 2	Year 4	
9-10		Year 5	
10-11		Year 6	英語、數學、科學的國家測驗與教師評量
11-12		Year 7	
12-13	Key Stage 3	Year 8	
13-14		Year 9	英語、數學、科學的國家測驗；其它基礎科目的教師評量
14-15		Year 10	某些學生接受 GCSEs
15-16	Key Stage 4	Year 11	大部份學生會接受 GCSEs 或其它國家資格檢定考試

註：GCSEs：中等教育認證科目(General Certificate of Secondary Education)

#### (1)初等教育

一般來說，英格蘭的小學招收 5~11 歲的學生，如表 4-9 所示，是屬於英國國定課程中的第一關鍵階段 (key stage 1, 5~7 歲) 和第二關鍵階段 (key stage 2, 7~11 歲)。英格蘭的初等教育皆採用英國國定課程。根據國定課程的第一關鍵階段及第二關鍵階段，英格蘭初等教育的核心科目包括英語、數學、和科學，此外還有基礎科目包括：資通科技 (information and communication technology, ICT)、設計與科技 (design and technology)、歷史、地理、藝術與設計 (art and design)、音樂、體育 (PE)、以及宗教教育。

除國定課程之外，英格蘭的小學還規劃許多非國定科目或課程，例如經濟、環境教育、政治與國際 (包含歐洲認同、歐洲歷史文化與社會經濟)、公民與品德教育、性教育、藥物教育等。此外，英格蘭已規劃在 2010 年之前，全面提供第二關鍵階段的學生現代外語教育。而威爾斯的非國定科目或課程則是將重心擺在對其文化、經濟、環境、歷史、及語言等特性的認識和了解。

在學習評量方面，架構如表 4-9 所示，在每一關鍵階段結束前，都會針對核心科目進行評量，而這些科目包括：語文（英語）、數學、和科學。其中，評量方式，英格蘭小學在第一關鍵階段結束前（7 歲），是由教師對學生的學習情形對照國定課程能力指標（attainment targets）的設定進行等級評量，而對於等級評量，如表 4-10 所示，是將整個義務教育分成 8 個層次（詳細如附錄一所示），其中，大部分初等教育學生的等級是介在第一級到第三級之間，而初等教育的目標乃希望學生在第一關鍵階段結束前達到第二等級。而第二關鍵階段結束前（11 歲）則是同時利用成就評量（即國家測驗）及教師評量來評定等級，其中，大部分學生的等級多落在第二級到第五級之間，而教學的目標則是希望學生能達到第四級，此外，若是學生無法達到該預期等級，則必須進一步接受補教教學。

表 4-10 學生於各關鍵階段預期達到的水準

關鍵階段	水準範圍	年齡	預期達到的水準
Key stage 1	1-3	7	2
Key stage 2	2-5	11	4
Key stage 3	3-7	14	5/6

## (2) 中等教育

英格蘭的中等教育範圍係指提供給 11~19 歲學生的各類型學校教育，其中 11~16 歲為義務教育。中等教育的前五年（11~16 歲）屬於義務教育，其中 11~14 歲是屬國定課程的第三關鍵階段，而 14~16 歲是為第四關鍵階段。

英格蘭的中等義務教育包括國定課程、宗教教育、以及各校自訂之課程，且所有接受公款補助之中等學校（maintained schools）皆應提供性教育課程，內容應至少包括對愛滋病、愛滋病毒（HIV）、和其他性傳染病的認識。此外，中學也必須提供毒品菸酒教育、工作相關的學習、和生涯教育。

英格蘭第三關鍵階段的國定課程包括核心科目（英語、數學、和科學）和基礎科目（設計與科技、資通科技、歷史、地理、現代外國語（modern foreign language）、藝術與設計、音樂、體育、公民）。國定課程之外，自 2000 年 8 月起，中學生都應修習品德與健康教育（personal, social and health education, PSHE）；而公立學校都應提供宗教教育課程，但家長有權力決定自己的孩子是否參加此課程。而在評量上，如表 4-9、4-10 所示，是由國家測驗與教師評量來評定學生在整個能力層級中，是否達到預期的第 5/6 的水平。

到了第四關鍵階段，學生可以選擇自己想要修習的中等教育認證科目（General Certificate of Secondary Education, GCSEs）（Directgov, 2008a），可選擇的範圍包括超過 40 種一般科目，以及 8 種應用類科（與工作相關的學習科目），這些科目有部份與國定課程重疊。一般而言經過五個學期（將近兩年，也就是 14~16 歲）的學習後便可參加中等教育認證考試，各校也會視學生的程度而允許資賦優異的學生提早考某些科目。中等教育認證考試不限定每次應考幾科，許多

科目的成績會包括平時學習成果及考試結果，成績評定以八個等級表示：A\*、A、B、C、D、E、F、G，若成績不達最低標準則標示為 U，表示無法拿到認證。

根據 2002 年教育法案，英格蘭的第四關鍵階段國定課程開始與前三階段有明顯區別，使該階段的義務教育更具有彈性，並自 2004 年九月起頒佈實施，其中的改變包括：(1) 減少義務教育的核心科目，目前的核心科目有：英文、資通科技、數學、科學、公民、體育、和宗教教育。(2) 將生涯教育與性教育囊括進義務教育的學習領域。(3) 所有接受公款補助之中等學校皆應提供與工作相關的學習。(4) 除了核心科目之外，各校應就以下四領域：藝術、設計與科技、人文學科、及現代外國語，各領域至少應提供一門課程。此一「14~16 歲學生增加彈性方案」(Increased Flexibility for 14- to 16-Year-Olds Programme, IFP) 爲了使學生有更多職業或工作相關學習的機會，還請當地的擴充教育學院與這些中學及代辦單位合作，使中學生有機會到學院上課，或者每週安排一至兩天接受職業訓練教育，學生也因此能爲職業或工作相關的證照考試作準備和學習，並發展其對工作／職業狀況的認識和了解。

16 歲~18 歲以上是屬義務後教育，學生所接受的義務後教育不在國定課程的範圍，因此學生學習的課程內容將依照其所選擇的認證考試及認證單位所公佈的課程大綱來安排，可能的選擇包括：進階級子認證 (Advanced Subsidiary qualifications, GCE AS)、進階級教育認證 (GCE A-level)、進階級職業教育認證 (A-levels in applied subjects)，總共包括了超過 80 種學術科目及職業科目 (Directgov, 2008b)。

### (三)日本評量架構

#### 1、日本全國學力調查沿革

日本於 1960 年代就已展開全國性學力測驗，由於學校及地區競爭過熱，1964 年停止全體調查，1965 年起改以抽樣方式，由於國家層級舉辦之學力調查被判定違法，1966 年起全面停止。1982 年起以局部之中小學生爲對象，2002 年起以局部之高中生爲對象，再度展開學力調查。2004 年當時的文部科學省(相當於我國教育部)中山成彬部長爲因應學力下降以及確保次世代人材基盤的義務教育品質的提升，客觀掌握學童學習狀況的必要性考量下，2005 年 6 月「2005 經濟財政營運和改革構造的基本方針」的內閣會議(首相官邸，2005)以及中央教育審議會答詢「創造新時代的義務教育」決定實施全國性學力調查(文部科學省，2005a)，同年 11 月文部科學省召開全國性學力調查實施方法等之專家檢討會議(文部科學省，2005b)，討論具體實施的方法(文部科學省，2006)。睽違 43 年，2007 年日本再度以中三及小六全體爲對象(※僅少數學校以引發競爭爲由拒絕參加)，展開國語與數學的學力調查 (文部省，2007)，以掌握分析各地區學童的學力與學習狀況，檢討教育及教育施策之成果與課題以求改善，提升學童之學習效果及提升學習意願，進而保障義務教育機會均等及維持與提升全國水準。

## 2、以普查方式實施之緣由

學力低落問題受到重視下，基於學校教育的現狀和課題充分地把握的必要性、國際學力調查結果發現學力和學習意願有下降傾向、建立保證義務教育質量架構的請求，而實施學力調查。學力調查採取全國普查方式實施的原因在於：全國學力狀況趨勢分析的抽樣性調查，對各學校教學助益不大，而全國性普查除提供教育政策的確立與「學習指導要領(指日本的課程標準)」修訂等之依據外，可以提供地方教育委員會、學校以及教師改進教學；其次，國際評比排名與學習意願下滑，例如：日本過去在 TIMSS 與 PISA 等國際評比的數學、自然科學等成就測驗上，有著良好的成績表現，日本學生的 TIMSS 小四數學排名：1995 年與 2003 年第 3 名、2007 年第 4 名；八年級數學 TIMSS 1995 第 3 名、TIMSS-R、2003 與 2007 都是第 5 名。日本 PISA 數學排名則從 2000 年的第 1→2003 年的第 6 下降至 2006 年的第 10 名(林宜臻，2009)；最後，建立保證義務教育質量架構的請求：1980 年代以來，日本為改變填鴨教育實施「寬鬆(裕)教育」政策，1998 年版的學習指導要領將原有教學內容削減了三分之一，該政策被認為是導致日本學生學力下降的重要原因，因此，為提高學校和學生的競爭能力及促進學習，文部科學省實施全國性學力普查。

## 3、調查目的

- (1)全國義務教育機會的均等與水準的提升，由各地區學生的學力與學習的掌握與分析，檢驗教育與教育施策的成果與課題，以謀求改善。
- (2)各教育委員會、學校等從與全國狀況的關係中，掌握各自教育及教育施策的成果與課題，謀求改善的同時，確立後續的檢證與改善的機制。
- (3)各校得以掌握學生的學力與學習狀況，幫助對學生的教學指導和學習狀況的改善等。

## 4、調查內容

其調查內容，主要以學科中的國語、算術/數學為主。

- (1)出題範圍：原則上以該調查年級先前的學習內容為主。
- (2)出題內容：以「知識」和「活用」為出題種類。

「知識」算術/數學 A 以會影響後續學年學習的內容及生活中活用不可或缺的知識、技能為主。

「活用」算術/數學 B 將知識、技能等活用於生活中各種場面的能力，以及為解決各種課題的構思及評鑑、改善能力等之內容為主。

問卷調查(生活習慣和學習環境等)其下分為學生與學校問卷，其中，學生問卷：以接受調查該學年的學生為對象，調查其學習意願、學習方法、學習環境及生活等；學校問卷：調查學校的指導方法以及學校的人力、物力上等教育條件之整備狀況。

## 5、實施方式

本調查以紙筆測驗及問卷調查方式進行，以瞭解學生對語文及數學的基礎知識與基本技能的掌握和活用情況，以及相關的背景和條件，本文將就下表 4-11 之 2009 年 4 月 21 日進行的學力調查為例探討之。

國立學校無論小六或九年級階段全數參加，公立學校也幾乎全數參加，惟 2009 年九年級階段下降至 96.9%，私立學校無論小六或九年級階段參加比例最低，由六成下降至 43.1%(小六)及 46.6%(九年級)。

表 4-11 日本學力調查範圍

校別	年度		
	2007 年	2008 年	2009 年
國立學校	75 校 (100.0%)	76 校(100.0%)	76 校(100%)
小六 公立學校	21,889 校 (99.8%)	21,685 校(99.8%)	21,482 校(99.8%)
私立學校	108 校 (60.0%)	88 校(47.1%)	85 校(43.1%)
合計	22,072 校 (99.5%)	21,849 校(99.4%)	21,643 校(99.3%)
國立學校	81 校(96.3%)	79 校(97.5%)	77 校(95.1%)
九年級 公立學校	10,050 校(98.0%)	10,121 校(99.3%)	9,851 校(96.9%)
私立學校	416 校(60.5%)	374 校(53.4%)	330 校(46.6%)
合計	10,544 校(95.7%)	10,574 校(96.4%)	10,258 校(93.6%)

## 6、日本學力調查的特色

### (1)精進調查工具

#### a. 試題內容生活化

日本借鏡 PISA 國際學力調查，將試題除「知識」外，同時將「活用」列入試題是一大特色，A 卷的「知識」主要考查學生不學會將影響後續學年的學習內容，以及生活中活用時不可缺少的知識與技能為主。B 卷的「活用」主要考查學生能否將知識與技能等活用於實際生活中的能力以及解決課題的構思、評鑑、改進等的能力。

#### b. 題型多樣

A 卷由選擇題、簡答題組成，B 卷由選擇題、簡答題和問答題組成。

#### c. 問卷內容廣泛

問卷調查包括對學生和學校相關背景的調查，除了涉及學生的學習興趣、態度、學習方法等，還有學習環境、生活等方面的內容。因此，調查結果不僅可用

於學校教育，也有利於學生家長協助學校教育，改善學生的生活習慣、家庭環境等影響學力的因素。

#### (2)變更調查年級

爲了實施全國學力調查，文部科學省於 2005 年 11 月 8 日設置「全國性學力調查實施方法等之專家檢討會議」(文部科學省，2005b)歷經七次會議討論具體實施的方法，翌年(2006年)4 月公佈了關於全國學力調查具體實施方法的報告(文部科學省，2006)。較以往之相關調查係針對小學五年級學生進行，該報告指出，爲了掌握義務教育水準，選擇小學六年級及九年級較爲合適，由於調查在新學期 4 月進行，因此測驗考查範圍原則上以調查年級之前的指導事項。

#### (3)調查結果同年公佈

學力調查於 2009 年 4 月 21 日進行，2009 年 8 月 27 日就公佈調查結果，12 月提出調查結果報告書。

#### (4)公開調查問題

日本教育部於調查實施後，除提供統計分析之相關數據外，亦公開調查問題、正確回答事例、題目主旨以及解答類型。

#### (5)規範公開方式

日本會公佈全國性調查結果，包含公佈全國、都道府縣、地區規模(大都市 核心城市、其他城市、鎮村、邊遠地區) 的調查結果，其次，分析與公佈學生的生活習慣及學習環境、學校教育條件的整備狀況等與學力之相關情形。

#### (6)提供給教育委員會、學校等調查結果

考畢，日本會提供都道府縣、市鎮村、學校等之調查結果，同時，不歸還學生答案卡，但透過學校提供個人正確回答與誤答狀況之通知，此外，不明確指出各個市鎮村名、學校名等，避免學校排名及過度競爭。但市鎮村、學校得以向家長等說明自己的結果，最後，調查結果不作為入學考試資料用。

#### (7)提供教學改進建議

評量結果除指出學童答對率外，並分別針對內容指出學習上之問題點，並提出教學改進之建議。

### 7、日本學力調查的問題點

#### (1)「活用」型的問題仍待努力

2009 年公佈的學力調查結果發現，秋田、福井、富山等縣市因年仍名列前茅，表現好的縣市，除了學校指導有力外，家庭中的學習習慣及正確的生活習慣，

也是影響學力的提升。原先排名最差的沖繩及大阪等縣市，向排名第 1 的秋田縣學習，沖繩甚至與秋田交換教師，大阪市加強計算問題的反覆練習，小學階段屬於「知識」類型的問題，成績接近平均值，排名也由 41 名提升至 33 名，但「活用」的問題仍待努力，中學則連續三年全國最後(產經新聞，2009 年 8 月 28 日)，

#### (2)鄉鎮村別及學校別的資訊是否公開

規制改革會議 2008 年 12 月的答詢中指出「使用多數的公費未提供國民相對的資訊」，內閣府委託民間調查公司於 2009 年 1 月進行調查，結果發現：約 7 成的家長認為應以學校別公開成績，約 9 成的縣市教育局則主張不宜公開(朝日新聞，2009 年 8 月 3 日)。鳥取縣教育委員會無視於文部科學省要求不能公開學校別及市町村別(指鄉鎮村)的調查結果，首開公佈學校別及市町村別全國學力調查結果之例(読売新聞，2009 年 9 月 08 日)。

#### (3)全國性普查是否合宜

全國性普查的方式 2007 年耗資約 77 億日幣，2008 年及 2009 年耗資約 58 億日幣，2010 年的預算也達 57 億日幣，被指出「分析所得結果與經費和勞力間不成正比」(朝日新聞，2009 年 8 月 24 日)，民主黨建議以抽樣調查方式取代全員參加的學力調查(產經新聞，2009 年 9 月 2 日)，隨著 2009 年 9 月 16 日民主黨鳩山由紀夫新政權的誕生，教育部長川端達夫指出重新檢討以抽樣方式進行調查以及增加科目等(每日新聞，2009 年 9 月 17 日)。

#### (4)體制的矛盾

梅原利夫(2007) 指出(1)測驗分數的高低成爲唯一的評量基準—教育課程由各學科、道德及特別活動構成，片面性的評量基準違反教育宗旨；(2)與「多樣學力與生存力」的學力觀矛盾—爲提高 PISA 調查成績。(3)委由民間公司參與獨攬造成不安—評量題編制雖由文部科學省負責，但試題及答案用紙的印刷、梱包、發送、回收、計分、和數據處理、公告資料製作等作業則委託精通考試的民間教育產業擔負，其中結果的數據及問卷的個人情資由民間公司獨佔引發不安。(4) 外部評鑑扭曲教育評鑑—評鑑應由自己評鑑、相互評鑑、外部評鑑、綜合評鑑等機能性組合之，方利於教育活動的有效進行，由外部評鑑取代評鑑，教師與學校目標由外部設定，將不能因應事實進行適切的診斷。

### (四)中國評量架構

「建立中小學生學業品質分析、回饋與指導系統」專案與「國家基礎教育品質監測中心」共同構成了中國國家基礎教育品質保障系統。

#### 1、中小學生學業品質分析、回饋與指導系統

##### (1)建立的背景



中國隨著基礎教育已從普及邁向教學品質全面提升的新階段，而教學品質大範圍的提高，有賴於科學的宏觀調控和指導，在基礎教育課程改革深入推進的階段，適應新時期基礎教育發展的需求，2003年，在教育部基礎教育司的指導下，中國教育部基礎教育課程教材發展中心組織成立了「建立中小學生學業品質分析、回饋與指導系統」專案組。其成員包括課程教學、心理學、教育測量、教育統計、電腦軟體程式設計等各方面的專業人員。中國教育部課程中心主任朱慕菊針對中國當前沒有一個基礎教育的資料平臺，不能對各地區的基礎教育情況作橫向的比較的現狀，力圖通過建立國家級、省級常模，進行歸因分析，指導教學，從而促進義務教育更均衡地發展，指出「建立國家中小學學習品質分析與指導系統」的重要性(中小學學習品質評價的新動向，2007年3月8日)。

## (2)目的

「建立中小學生學業品質分析、回饋與指導系統」專案設置目的在於：(1) 真實反映學生學業品質狀況；(2) 提高教育決策科學化水準；(3) 教育教學診斷和品質提高提供依據；(4) 提升教師和相關人員的專業水準；(5) 引導社會樹立正確的品質觀。

對學生學習品質進行評量監測，主要是為決策和指導服務，其意義在於：(1) 反應學科教學多方面的指標資訊；(2) 對學校和地區教學過程進行診斷分析；(3) 測試之後對學校和地區教學工作進行針對性指導(董洪亮，2008年2月18日)。針對中小學生學業品質進行檢測、分析、指導。經由資料的收集、分析和回饋，幫助各級教育行政、教研部門，有針對性地改進教學中的問題，提高學業品質。來自一線的優秀教師的教研人員，面對全面提高基礎教育品質，在經驗積累的基礎上，學會運用科學的手段。

## (3)實施概要

中國的評量架構，其調查對象是以小學三年級和中學八年級，而其調查內容包含學科調查與問卷調查，其中，學科調查是以三年級測試語文、數學兩個學科、八年級測試語文、數學、英語、科學四個學科，而問卷調查是包含：學生問卷，以調查的內容包括家庭環境、學校環境、課業負擔、師生關係、與同學的關係、學習方法、學習態度等；教師問卷，調查的內容包括教師基本情況如學歷、任職經歷、職稱，教師對新課程的看法，教學策略和教學方法，教師對學校管理的看法，教師面臨的困難，教師的教育觀和教師專業發展等；最後，校長問卷，調查的內容包括學校硬體條件、校長辦學理念和教育觀、校長的學校管理方法等。

評量架構方面，各學科學習品質評量架構主要依據課程標準，考核學生對學科核心知識與技能的理解和掌握，尤其重視科學方法的掌握、解決問題的能力、收集與分析資訊的能力和反思批判能力。評量架構是將課程標準中對學生的要求轉化為操作化、行為化的描述，將之分為內容與能力兩個向度。

#### (4)特色

##### a.命題內容的特色

以課程標準為依據，在基礎知識與基本技能的雙基基礎上，重視核心素養與核心能力，使用真實的情境和任務，以客觀性試題考察學生高層次認知能力，突顯學生未來發展的核心素養和核心能力。其命題內容包括：(1) 學科核心知識、技能的理解和掌握；(2) 綜合運用所學知識，解決實際問題的能力；(3) 收集與分析資訊能力、反思批判的能力，以及對重要學科方法的掌握。

##### b.試題編製的特色

每一道試題編製的都經過專家審議和多種形式的試測，經過層層篩選，明確標定每一道題目的難度、鑑別度、反應曲線等資訊，並運用布魯姆分類法對能力進行界定，同時結合索羅分類法 (SOLO Taxonomy)設計題目選項，以測驗的方式瞭解學生解決問題時表現出的認知發展水準，更精細區分學生能力。

##### c.評量結果解釋的特色

評量結果的解釋同時結合標準參照與常模參照，標準參照確定學生處於怎樣的水準，常模參照比較不同類別學生的表現，使教育行政和教研部門等超越個人經驗的侷限性，從不同的層面，瞭解學生在學業上的水準和特點，能更準確地判斷自己地區或學校各學科的整體水準及優弱勢，得以有針對性地進行改進和指導，還可以利用資料資訊追蹤研究，連續若干年瞭解學生學業品質的發展變化狀況。

##### d.測驗結果處理的特色

###### (a)測試專家分析而不解說因果關係

專案組負責命題、抽樣、測試過程管理、閱卷、資料錄入與整理、資料分析、報告撰寫，以及教學改進等項目，呈現學生在各學科不同內容領域和能力維度上的表現及其在典型題目上的作答情況，分析學生學業成績和學習環境因素的關係，培訓一線教研員幫助解讀報告，過程中，測試專家只對數量關係進行分析，因果關係的解說由當地的管理者、研究者和教師承擔。

###### (b)即時回饋系統

12月底完成測試，隔年3月前後，完成資料處理和報告撰寫。4月，國家專案組組織專家對全省的測試情況進行省一級的回饋。5月，國家專案組支援所有測試科目教研人員的專門研修活動，進一步學習科學命題、科學評量的相關技術內容，更深入地解讀省、市、縣區三級的分析報告。6月底，省教研室組成專家組，進行巡迴式的回饋活動，並要求各地級市提交利用測試結果改進教學工作的規劃，規劃當中包括測試結果、原因分析、改革重點以及改革的計畫和保障等等內容。

### (c)依據測試結果展開課堂教學改革活動

各地教育行政部門以對測試結果的分析為依據，展開一系列針對課堂教學的改革活動。以江蘇省為例，在測試結果分析的基礎上，省教研室組成專家進行巡迴式回饋活動，同時要求各地級市提交利用測試結果改進教學工作的專門計畫和報告。在市級回饋的基礎上，組織了縣區級的回饋、研討與交流活動。實證資料面提供當地教育的優勢和不足，市教育局組織政府領導、教育行政部門領導、學校校長等為學科教學品質進行專門研究，利用 3 個月時間對所有小學的教學狀況進行拉網式調查研究，瞭解師資情況、設備條件情況和課堂教學情況，推動教學管理與課堂教學改革，針對性執行改進措施。

### (d)資料分析報告內容特色

提供的資料分析報告內容包括：學生在各學科上的總體表現；學生在各學科不同內容領域和能力層次上的表現；不同群體學生的比較；學生在每題的作答情況；典型題目分析；問卷調查結果分析。根據測試結果的統計分析，以學校或區域、學科為單位，針對學習狀況、學習能力、教學過程、管理過程進行實際診斷，並提出相關的改進建議。

### (e)由需求者自籌經費提出申請參加測試

基於「中小學學習品質分析與指導系統」專案裨益於：提高教育行政部門、教研部門的決策和指導水準；促進教學品質的大範圍提高和基礎教育的均衡發展；引導評量方式的轉變。以江蘇省為例，2006 年年底，江蘇省教育廳投入 600 萬人民幣左右的專項經費實施覆蓋全省 13 個地市、106 個區縣 11 萬多次的抽樣測試(董洪亮，2008 年 2 月 18 日)。

### (f)利用定錨題探討趨勢

以課程標準的要求為參照，為保證歷次採集到的資料本身具有連續性和可比性，題型、題量以及每一類題型的檢測重點等方面具有高度的穩定性，而且每一次測試的試題都與「定錨題」的類型相同，瞭解學生學業品質的發展變化狀況。

## 2、中國教育部基礎教育品質監測中心

### (1)緣起與角色

相對於中國教育部基礎教育課程教材發展中心於 2003 年組織成立的「建立中小學生學業品質分析、回饋與指導系統」專案組，2007 年 9 月中國中央機構編制委員會辦公室批准成立「教育部基礎教育品質監測中心」，在教育部的直接領導下委託北京師範大學建立專業機構。中國胡錦濤總書記在 2007 年 10 月 15 日在共產黨第十七次全國代表大會上的報告重申「優先發展教育，建設人力資源強國」，強調實施素質教育，更新教育觀念，深化品質評量制度改革，建立與新的品質觀相適應的評量制度，以保證學校基礎教育品質的全面提升，指明教育督

導評估工作的發展方向 (鄭富芝, 2007 年 11 月 09 日)。同年(2007 年)11 月成立「教育部基礎教育品質監測指導委員會」, 由課程、學科和教育評量、學業測量等方面的專家組成, 為教育品質監測工作提供學術諮詢和支援, 用以加強對基礎教育品質監測工作的宏觀指導與管理, 確保教育部基礎教育品質監測中心權威性和科學性。中心負責擬定基礎教育品質監測標準, 組織實施全國基礎教育品質監測工作, 為各地展開基礎教育監測工作提供技術支援和業務指導。

## (2)目的

陳小嫻教育部副部長在由聯合國教科文組織國際教育規劃研究所與上海市教育科學研究院聯合主辦的「變遷世界中的成功教育體系：教育系統效能監測與評估」政策的研討會指出：基礎教育監測是科學決策的必然要求, 是進一步轉變政府職能、改善宏觀管理的重要措施(王軼軒, 2009)。中國教育部部長周濟在教育部基礎教育品質監測中心揭牌儀式上, 指出要建立基礎教育品質監測體系, 全面準確把握基礎教育的品質狀況, 科學診斷我國基礎教育品質存在的問題和原因, 教育決策者可以根據監測所提供的資訊對未來的教育進行調整。(佚名, 2008 年 2 月 14 日)。對基礎教育品質進行全面的監測和科學的評量, 全面準確掌握學生的真實素質狀況, 其目的在於：(1) 探索人才培養的規律, 促進各地區、各學校全面實施素質教育；(2) 提供實踐和理論依據, 全面提高教育品質；(3) 提高教育督導的針對性和時效性, 使教育督導工作監督有據、指導有方, 提高督導的服務水準。

## (3)實施方式

### a.監測內容

周濟教育部部長指出基礎教育品質監測要：(1) 按照基礎教育品質標準進行；(2) 監測範圍要覆蓋學生德、智、體、美全面發展的各個方面；(3) 要特別重視監測學生的創新意識和心理健康水準。監測的重點領域包括學生的身心健康水準、學生的綜合素質、學生的學業成績, 考查中小學生在品德、智力、體質等方面的全面發展狀況, 以及獨立思考能力、創新意識、實踐能力、影響學生發展的教育環境等(佚名, 2008 年 2 月 14 日)。

### b.監測對象

以抽樣方式對基礎教育階段所有在校學生進行整體監測。

## (4)特色

### a.探索教育發展的規律

「依法監測, 服務決策」原則下, 以法律/法規以及基礎教育課程改革標準為依據, 採用科學的監測方法, (1) 診斷基礎教育品質存在的問題和原因, 幫助尋求解決方案；(2) 探索基礎教育發展的規律、人才成長的規律和教育管理的規

律。

#### b.綜合監測

國家基礎教育品質監測中心主要從：(1) 學生的思想品德和公民素養；(2) 身體和心理健康水準；(3) 學業水準和學習素養；(4) 藝術素養；(5) 實踐能力和創新意識；(6) 影響學生發展的教育環境與社會環境等六個方面進行綜合監測。

#### c.提供支援和指導

中心負責擬定基礎教育品質監測標準，組織實施全國基礎教育品質監測工作，為各地展開基礎教育監測工作提供技術支援和業務指導。

### 3、中國評量架構所提供之經驗、建議

#### (1) 設置改善與監控體系

中國為保障國家基礎教育品質系統，設置「國家基礎教育品質監測中心」與「建立中小學生學業品質分析、回饋與指導系統」。「國家基礎教育品質監測中心」屬於外部的「品質評估與監控」系統，探索教育發展的規律，面向政府報告基礎教育品質的現狀，為教育決策提供資訊、依據和建議。「建立中小學生學業品質分析、回饋與指導系統」面向基層，為基層教育行政、教研部門直接提供所有資料資訊，整個過程對一線教師開放，是一種內部的「品質保障與改善」系統，著眼於基礎教育品質的改善和保障。為教育系統內部和外部改善與監控各執其司下，得以「問責」與「回饋」，保證國家基礎教育品質並促進其不斷提升。

#### (2) 執行綜合監測

監測機制一般著眼於學業水準和學習素養，中國的「國家基礎教育品質監測中心」則針對基礎教育階段的學生德、智、體、美以及影響學生全面發展的環境因素等進行綜合監測。

#### (3) 提供科學化測試系統

提供科學化測試系統，轉變政府職能、改善宏觀管理，使教育行政和教研部門等超越個人經驗的侷限性，瞭解當地教育的優勢和不足，以及學生在學業上的水準和特點。此外，由需求者自籌經費提出申請參加測試，以免各地自行架設測試系統，疊床架屋。

#### (4) 建立監回饋與指導系統

中國的「建立中小學生學業品質分析、回饋與指導系統」，進行教學診斷並指導教學，幫助地方與學校逐步提升品質，重視分析與指導，以保障課程計畫的落實和達到課程標準要求。

#### (5) 分析而不解說因果關係

只對數量關係進行分析，因果關係的解說由當地的管理者、研究者和教師進行解說，將更能因地制宜，進行適切的診斷，能更準確地判斷自己地區或學校各學科的整體水準及優弱勢，得以有針對性地進行改進和指導。

#### (6) 即時回饋

各地教育行政部門方能依據測試結果，即時展開針對性的對策，進行課堂教學改革活動。

#### (7) 提供學生學業品質分析報告

提供呈現學生在各學科上的總體表現及其在不同學科內容領域和能力維度上的優勢和不足，並分析了學生學業成績和學習環境因素的關係的學生學業品質分析報告，方利於課堂針對性教學。

#### (8) 試題內容生活化

日本學力調查，將試題除「知識」外，同時將「活用」列入試題，「活用」的試題主要考查學生能否將知識與技能等活用於實際生活中的能力以及解決課題的構思、評鑑、改進等的的能力。

#### (9) 檢測認知發展水準

結合索羅分類法 (SOLO Taxonomy) 設計題目選項，以測驗的方式瞭解學生解決問題時表現出的認知發展水準，更精細區分學生能力。

### 四、本建議方案之思路與國際趨勢

在前文，本研究已詳盡描述國內相關的品管方案(例如：攜手計畫)與美國、英國、日本等國家之學力品管架構，而面對如此多元與豐富的資訊，要如何有效的規劃或者從哪裡開始執行?則變成一個重要的課題。研究團隊在依循 Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)的想法下，以他們所認為監控系統的資訊來源是包含下列五種，作為參考建議，更指引本研究建置學力品管之思路，作為未來執行的方向，陳述如下：

#### (一) 採用現行可得資料(Status quo)

對於監控系統的資料來源，Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)初步的想法很簡單，他們認為如果現行的資料已經能夠滿足或回答所需的問題時，研究者可直接採用，不必另行開發或建置，但這方面的資料，來源可能非常複雜(或時常變動)，必須視當下時間，所能夠獲得哪些訊息為主。

#### (二) 機制的整合(Patchwork)

Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)所提出的另一項選擇，是

認為研究者能考慮有效的整合目前現存的各式品管機制，而這個概念是有點像將上述可能雜亂無章的資訊來源，以系統化、制度化的方式，進行整合的動作，亦即只納入定期性出版或調查的方案，例如：研究者能考慮整合 NAEP(National Assessment of Educational Progress)、TIMSS 與 PISA 等方案，以形成另一種整合性品管監控機制。

### (三)週期性的研究(Cyclical studies)

在上述方案中，Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)認為可能受限於時間、經費等因素，導致每年能搜集到的訊息，會很表面或簡略，因此，他們認為可以透過週期性的研究，分年度針對不同的主題進行更深入的訊息搜集，例如：在過往時，研究者可能受限於某些因素，以致無法同時(或深入)針對學生成就、教師或學校背景變項等逐一進行評量與調查，但利用週期性的研究模式，研究者可考慮將時程切割成幾年完成，在第一年，研究者可只針對學生成就進行評量(例如：採用實作評量取代紙筆評量，以獲得更深入訊息)，而第二年，再執行教師、學校背景訊息之調查等，如此，研究者每年就能以單獨、較充裕的資源深入調查所需的訊息。

### (四)援用現行方案(Piggyback)

這項方案主要是 Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)對於上述機制整合的另一種想法，他們認為當各方案無法整合(或整合難度過高)時，研究者是可以單獨採用(或必要時，加以擴增)一項完整的方案，例如：研究者能考慮只採用 NAEP 評量方案，以作為品管機制，但其中，為滿足自身所需，可視情況加以作些微的調整。

### (五)獨立方案(Independent)

Shavelson, McDonnell, Oakes, Carey 與 Picus(1987)的最後一項建議，就是在上述訊息來源，都無法滿足研究者所需時，可以考慮自行建置一套獨立的品管機制，但此時研究者不僅需要衡量其必要性外，還必須審慎考量要如何重新建置之問題。

本研究團隊綜觀國內現有相關方案與機制後，認為臺灣目前不論是自行建置或國際的參與，都存在著學生接受過多的測驗、評量之疑慮，加之各縣市、中央單位仍不斷的推出新式品管方案，致使現行可用的訊息或資訊，可謂十分豐富，因此，團隊成員認為實無再獨立建置之必要性，而在進一步思考其它可行的層面時，研究者認為若以機制的整合為考量時，實務上，是有其執行的難度，主因在於現存的評量機制，不論在目的、決策單位、未來規劃上，多存在歧異性。綜合上述，研究團隊會是以援用現行方案，再加以擴充、修訂為建議目標。

在確立執行的大方向後，其細節的規劃，則是本節下一步描述的重點，而在上述的文獻探討中，研究者大致能獲得品管機制的實體架構，但對於其意涵，則

有必要預先加以陳述。研究團隊是根據 Koretz(2008)、Koretz 與 Hamilton(2006) 等人對於國際趨勢與潮流的剖析，作為借鏡，而在本建議方案中融入這些想法，其中，大致能歸納出以下幾個方向，陳述如下：

### (一)增加教育人員與學生的利害關係

1980 年左右，對於教育成效的評定，並不是那麼倚賴在高風險測驗 (high-stakes exams)，但 Koretz 與 Hamilton(2006)認為近年則逐漸朝向以測驗結果檢視成效為導向，導致利害關係(stake)之重心移轉至教育人員與學生身上，間接使得教育人員跟學生所必須承受的壓力增加，其中，最重要的例子，就是美國 NCLB 要求學校每年都要有足夠的進展(AYP)。

### (二)發展多元題型

由於成本與便利性等因素的考量，使得選擇題型在 20 世紀大型測驗中所運用的比率，始終高居不下，但 Koretz 與 Hamilton(2006)認為在 1980 年末與 1990 年初，有另外一項呼聲不斷的傳入，那就是許多教育學者與改革者開始強調使用實作評量(performance assessment)作為檢定學生能力的工具，因為，他們希望能評量較複雜、進階的知識或技巧，同時，期望著能推升評量引導教學之功能，在如此訴求下，傳統選擇題型所扮演的影響力，也逐漸受到影響，例如 NAEP 在 1996、2000 與 2003 年數學評量中，也納入近一半的建構反應試題(constructed response items)(NCES, 2004)，至今，雖然受限於經費因素，許多測驗始終是以選擇題型為首要，但實作評量或其它多元題型之訴求，仍不斷受到重視。

### (三)依賴以標準為基礎的測驗結果呈現方式

在 1980 年代後期，決策者與教育學者開始訴求建立一套應用在所有學生的標準，以期能向學生、教師溝通所期望的成就水平，而這項想法不久即獲得美國聯邦立法同意，法案 The Goals 2000: Educate America Act of 1994 的成立，就促成內容標準(content standard)與表現標準(performance standard)的使用，接續，在由 National Council on Education Standards and Testing(NCEST)規劃進行(NCEST, 1992)。在這項改革下，研究者對於測驗結果呈現亦逐漸由傳統常模參照模式(norm referenced)，轉變成以標準參照(standard referenced)的解讀方式，強調在某個水平上，學生學會了什麼知識、技巧或能力，至今，例如 NAEP 就已完全採用標準參照的方式，進行測驗結果之詮釋(請見網站 The Nation's Report Card：<http://nationsreportcard.gov/>)。

### (四)客製化的測驗與標準模式

對於內容標準與表現標準的建立，所強調的是由各州自行發展，Koretz 與 Hamilton(2006)認為此舉直接促使著各州對於客製化測驗的需求，他們期望能適切的聯結課程、測驗與標準之關係，以進行學生測驗結果之詮釋。而此項進程，



在伴隨著 NCLB 的鼓吹下，更是快速的融入於各州的活動中。

### (五)納入學習不利學生的需求

近年來，隨著大型評量之推動，對於標準或施測流程的一致性要求，亦逐漸的推升，但其中，對於學習障礙學生或英語受限的學生，是否也具備同樣要求，則變成另一個關注重點。在隨著美國 Title I of the Improving America's School Act of 1994 立法的成立，強調接受經費補助的學校，都必須將學習不利的學生納入評量架構後，間接促使學校開始針對不同需求之人士，發展適合或需經修訂、調適的測驗工具，接續，加上 NCLB 的推展，又重新要求在測驗上，要呈現出不同種族、社經、學習不利與語言受限等四個群體之結果，再再促使納入這群學習不利學生的需求。

### (六)強調群體的測驗結果報告

伴隨著以測驗為基礎的責任績效系統建立，其內容所強調的都是以學校層級的獎賞或懲罰為核心，因而，Koretz 與 Hamilton(2006)認為間接促使各州亟欲建立一套能提供學校或年級層級等群體的測驗結果解讀系統，以彰顯其執行成果，而此舉亦逐漸推升對於群體測驗結果解讀之重視。

### (七)強調報告學生學習成就的進步量

在早期，對於學生、學校表現之評估，多是以跨區域(cross-sectional)的資料比較為基礎，但時至今日，Koretz 與 Hamilton(2006)認為注意力已開始轉移，逐漸強調著學生隨時間表現或分數的進步，例如：在大型的 K-12 評量中，則多是以監控學生改變為目標。而這種改變的呈現方式，除了 NCLB 所採用的世代比較模式(cohort to cohort model)，也可以是以縱貫性研究(longitudinal approach)或稱增值模式(value added model)，進行個別學生長期性的追蹤，並評量其成長，亦或是採用準縱貫性研究(quasi-longitudinal approach)(Linn, 2000)，例如：以今年某學校四年級學生的平均分數，對照該校前一年三年級學生的平均表現，以進行改變量的監控。但不論採用何種研究方法，可以肯定的是決策者對這項趨勢的重視，是不斷的在增長中。

綜合上述，大致可以發現目前國際的潮流與趨勢，不外乎是上述幾點要件，而研究團隊就是在這些想法的導引下，規劃本研究建議之學力品管架構與細節，茲將詳細流程陳述如後。

## 五、本研究初步建議方案

從上述國內外現行的品管機制或措施中，大致可從其經驗中，粹取出學力品管機制中幾項重要的元素，包含有受輔對象的篩選、評量工具、基本學力標準、教學人員、配套措施、補救教學、成效考核工具等等，而此要素即為本研究建議方案之組成元素，在上述理論、實務基礎下，架構出整個面向，進一步詳細說明

如下。

在針對本研究學力品管機制，提出建議方案前，必須先澄清本方案所持之先備假設，雖然本方案只著重在提昇或監控學生的學習成就，但不表示不重視其它能力培養，例如：體適能、品格陶養等；此外，對於品質的管控雖然僅限於學生的學力品質，但老師教學能力、學校品質的提昇，亦期望政府單位應該給予相對的呼應；最後，若將學力品管的層面分為群體為基礎(group base)與個人為基礎(individual base)時，本研究所提之方案，是比較傾向以個人為基礎，意即對於學生學力的評量、監控，以至於最後的補救等，都強調必須深及至個人；而比較不同於單純監控整個學校或整個縣市群體之學力表現。

本研究對於建立國中小、高中職學科能力品管機制，所提之規劃方案，欲擬從現有之國家層級測驗來建立品管機制。綜觀之，研究團隊認為以臺灣學生學習成就評量資料庫(TASA)、國中基本學力測驗(<http://www.bctest.ntnu.edu.tw/>)、大學學科能力測驗(<http://www.ceec.edu.tw/>)等最適合用以作為我國學力品管機制中對於學生學力評量之工具，其餘諸如：各縣市測驗、診斷工具，其限制可參考文獻探討之分析或臺灣參與之 TIMSS、PISA 等國際評比方案，則有所必須遵循國際間統一之規範，是無法作完全彈性的運用，同時，國際評比方案多屬抽樣設計，並不適合作為學生全面進行補救教學依據。

國家層級測驗中，例如：臺灣學生學習成就評量資料庫，其建立的目的是在於追蹤、分析學生在學習上變遷之趨勢，進而瞭解國內學校教育之現況，作為教育部研訂課程與教學政策，縣市政府教育局及學校推動補救教學之重要參據(臺灣學生學習成就評量資料庫，2008)，就目的上而言，是符合學力品管之初步要求，應能有效提供品管機制中，作為篩選受輔對象的參據；此外，鑑於 TASA 施測的科目是廣及國文、英文、數學、社會與自然等學科，且對象是包含國小四、六年級、八年級與高中職二年級等，幾乎含蓋義務教育中的各主要階段，同時，考量 TASA 評量架構與九年一貫與高中職課程綱要的密切關聯，更促使延伸 TASA 作為學力品管機制時，對於後續補救教學內容的設計會貼近現行的課程。此外，其它國家層級測驗，如：國中基本學力測驗與大學學科能力測驗，在樣本推論性上，是較臺灣學生學習成就評量更佳，因其幾乎含蓋整個母群(九年級、高中三年級)的學科能力表現，因此，在本建議案中，對於國家層級測驗的選擇，這兩項測驗方案是值得納入延伸作為學力品管機制的考量。

本建議案的是可依循診斷、補救的模式進行，其概念是如下圖 4-2 所示，各級學生(如：國小四、六、八年級與高中職二年級)，在接受國家測驗(如臺灣學生學習成就評量或國中基本學力測驗)的評測後，以篩選出需接受輔導的學生(其中，亦可彈性將特殊或弱勢學生免測自動納入受輔)，接續，再透過積極的補救

教學以提昇受輔學生的學力。此外，本建議案中最重要的工作項目，是針對各年級基本學力的界定(同樣是以如國小四、六、八年級與高中職二年級為首要)，將直接影響學力品管機制中評量的組成要素，亦是診斷學生學力不足部分與作為補救教學課程設計的依據。

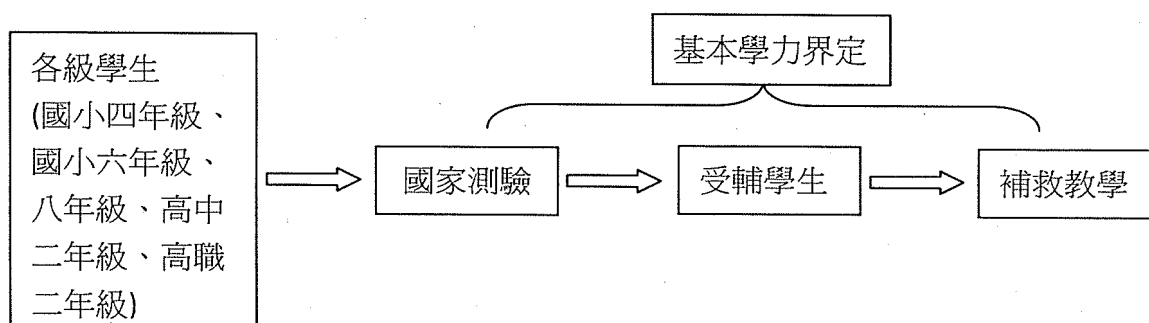


圖 4-2 國家測驗延伸之品管機制示意圖

國家層級測驗雖具有上述作為學力品管機制評量之特性，但目前仍有些許部份需要修補，才足以運用於學力品管，茲分別陳述如下。

就國中基本學力測驗與大學學科能力測驗而言，其施測的對象，是以九年級與高中三年級為主，因此，就樣本的廣泛性上，仍需要國小樣本與評量，才足以使學力品管機制覆蓋各年齡層。此外，這兩項國家層級測驗又因高度涉及學生個人權益，在做為學力品管機制的考量上，則需要十分謹慎。

就臺灣學生學習成就評量而言，鑑於 TASA 對於國中小各年級是屬於抽測，並非是全面性普測，因此，若期望以 TASA 為核心，並由中央機關主導，進行學力品管機制之建置，則可能須考慮採取普測且逐漸將試題內容擴及各個年段，才足以作為品管機制中針對施測對象進行學力診斷的工具。接續，對於學科能力的基本水準的設定，以確立經篩選之受輔學生所缺乏且必備之基本能力。同時期，後續的補救教學課程亦需隨之推動，以期能達到消極偵測學力、積極補救的功能。

另一方面，上述針對 TASA 實施完全普測的建議，是依賴中央具有龐大資源的建議作法，另外，本研究亦提出一項彈性的建議做法，如下圖 4-3 所示，原則上，是將學力品管機制中的評量，分中央、縣市與學校三個層級來實施。各層級有各自之權責，由中央統一監控、績效評核與配套措施的建立，由國家教育研究院主導 TASA 專責建立統一的評量題庫，但實際施測時，則是由各縣市提出評量需求的申請，再經由各級學校對學生進行普測，最後，透過國家教育研究院對測驗結果的解讀、回饋，經與教師溝通，以篩選出需受輔學生，進行補救教學。如此一來，同樣能瞭解各縣市學生的學力狀況，優點不僅能有效降低中央資源的負

擔，且依舊能利用經標準化流程所建立之題庫，此作法若推行得當，更能有效整合目前各縣市自辦的學力檢測的資源，但缺點則是若無嚴謹的監督機制(如對題目使用的規範)，易造成各縣市對於施測結果解釋、試題運用的偏差。除此之外，若欲採行此方案，現行 TASA 亦需進行些微調整，首先，是題型與評量層次的擴增，當前 TASA 題目多是以選擇題型與少許建構反應試題(如國文作文、數學應用題)為主，但這些題型對於反映出學生基本學力，有其限制，因此，後續題型與評量層次的擴增(如高層次能力試題)是有其必要性；其次，有別於現行 TASA 以群體進行測驗結果解讀的方式，教師的補救教學多是以個人為基礎，因此，對於受試個人施測結果解讀模式的建立，亦是方案順利執行的重要一環。

最後，關係到本建議方案能否有效推動的重要前提，則是配套措施的建立。首先，教育部必須預先建立整合、溝通的平臺，以利各單位業務之推動，同時，確立與協助各縣市教育局具有承擔全縣進行施測的能力，其間例如：評量制度的建立、測驗使用中可能發生的問題、各縣市施測期程等，都必須預先擬訂；其次，是根據 TASA 的評量架構與內容，以精確界定基本學力，此舉目的在於施測過後，能偵測出學生所真正缺乏能力或知識，並作為後續補救教學課程設計的重要指標；此外，教師的培訓則是關係到是否能有效提昇受輔學生能力的核心，若執行機關(如教育部)與教師間無法有效傳達建立品管機制的目的，則易產生弊端，如形成以評量領導教學，妨礙了品管機制的預期效用。

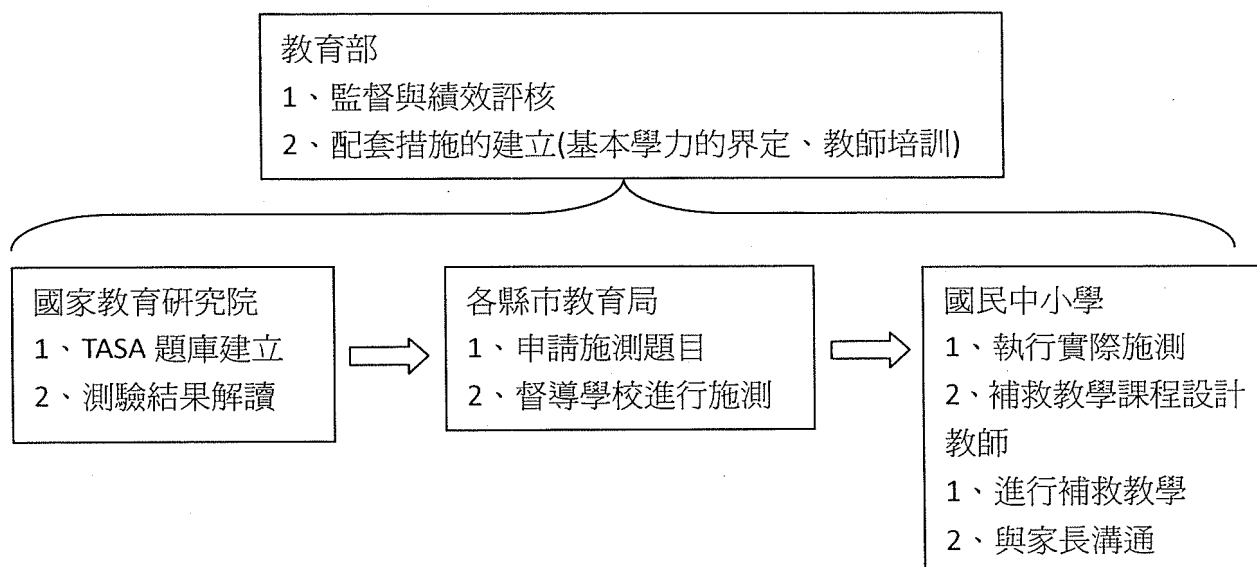


圖 4-3 TASA 延伸品管機制之各級單位層責

此外，若鑑於資源之有限性下，另有一項彈性作法，可供參考，即是針對國中小學生，仍是採上述建議方案，以個人為基礎的模式進行，但對於高中職學生

則可考慮以群體模式進行，即是預先設立監控單位，再利用 TASA 施測之結果，監控各學校、縣市之整體表現情況，當發現若某縣市或學校有過度低落或異常表現時，可由監控單位轉知該校或該縣市，予以加強補救；此舉，不僅是考慮到高職學生多重視技術層面，對於學科知識的需求相對較輕微，此外，加諸現行高職的學科科目過多，很難逐一予以補救，最後，目前針對高中職學生，亦有多項如會考、留級方式被相繼提出討論，若此時再加入學力品管措施，不免是增加實行的複雜度。

綜整之，本建議方案之原則是採中央監督、地方分權的方式進行，實際執行時，各權責單位則必須針對各自工作內容，進行設計與規劃，例如中央所分配到的工作內容之一，是必須界定基本學力，因此，何謂基本學力？各年段在各學科(如國文、數學)的基本學力為何？都必須在中央監督下架構起來；此外，如國家教育研究院所分配的工作內容，是測驗題庫的建立與測驗結果的解釋，則該單位必須依循著中央所訂立之基本學力，建立題庫，並針對測驗結果建立系統性的解讀模式等，同理，國民中小學之補救教學課程設計，亦是在此原則下進行。如此，各地方之權責則會在中央監督下，架構起完整、統一的學力品管機制。