

第二章 文獻探討

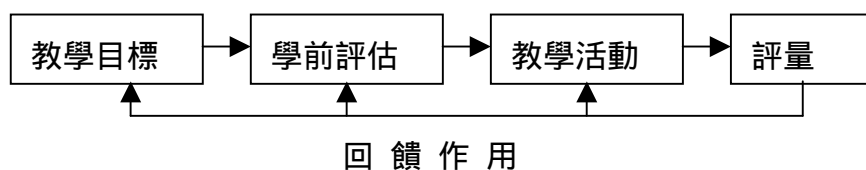
本章旨在說明本研究之理論基礎及研究理念。第一節是先蒐集整理現今教育研究上常用的評量方法，並分析各種方法的限制及探討其用途。第二節是探討現階段國中生的科學概念發展。第三節是比較現階段教育政策（九年一貫）所公佈的自然領域之科學能力指標與國際要求的學生的學科基本知能。第四節主要是介紹 IEA 的歷史背景及此組織的運作方式。第五節對 TIMSS-R 作評量架構的探討，內容包括其研究的範圍、方法及各種不同資料分析的方式。第六節則是研究和深入瞭解 TIMSS-R 的國際報告書內容。

第一節 評量方法研究

本節將就評量的意義、基本原理原則、多元化評量的方式及各種評量方法之以較；來探討評量方法在教學歷程中的重要性，以及在教育研究上的應用。

一、 評量的意義及演進

評量 (evaluation) 是運用科學方法和技術，蒐集有關學生學習行為及其成就的正確可靠資料，再根據教學目標，就學生學習表現的情形，予以分析、研究和評斷的動態過程。在整個教學歷程中，評量是承接轉合的關鍵部分。柯柏勒(R. J, Kibler)在一九七四年出版的「教學與評量之目標」(Objectives for Instruction and Evaluation)一書中，把教學的基本歷程分為教學目標、學前評估、教學活動、評量等四部分，進而闡述四者之間的交互關係，並特別強調評量的回饋作用及積極功能，其模式如下：



從上圖中可知:評量不是教學歷程的終點站,並非表示教學活動的結束。教學評量主要的目的在於分析教學得失及診所學習困難,作為實施補救教學和個別輔導之依據。有些人認為「評量」就是「考試」或「測驗」,把它當作跟學生算總帳或判決學生命運的手段,這是極端錯誤的見解。

一般言之,評量在教學過程中有四項最主要的功能:1.瞭解學生的潛能與學習成就,以判斷學生努力的程度;2.診斷學生學習的困難,作為補救教學及個別輔導的依據;3.估量教師教學的效率,作為教師改進教材、教法的參考;4.獲悉學習進步的情形,可觸發學生學習的動機(簡茂發,68年)。

二、 評量的基本原理原則

教學評量應顧及教學活動的七W層面,即「為何而教」(why)、「誰教」(who)、「教誰」(whom)、「何時教」(when)、「教什麼」(what)、「如何教」(how)、「何處施教」(where)等。教學在本質上是師生共同參與而交互影響的持續性活動,學生的「學」(learning)為主,以教師的「教」(teaching)為輔,故評量的重點在考查學生的學習成就,同時也衡鑑教師的教學效率。在進行教學評鑑時,必須把握五項基本原理:

1.決策原理:在教學過程中,隨時會遭遇各種教學問題,有待分析研判而後從幾個可供解決問題的方案中加以抉擇,採取必要的革新措施,以增進教與學的效果。決策是評量的理論基礎,為一種有次序且週而復始的連貫性歷程,從各方面蒐集正確可靠資料,並參照合理而適當的價值標準,以定取捨。

2.回饋原理:評量旨在教學歷程中提供各種必要的「回饋」(feedback)和「引導」(guide),一方面針對教學上的缺失而檢討改進,另一方面設法突破學習上的障礙,以提高其成就水準。因此,評量乃成為學校教育品質管制(quality control)的有效工具,可促進教育的正常發展。

3.完整原理:評量需要全面性、多元性的綜合資料,並從各個角度和不同觀點加以分析研判,故蒐集的資料愈多、愈齊全,則愈能掌握整體而加以靈活運用,即可避免「盲人摸象」之弊,又能發揮「既見樹木,也見森林」的統合功能。

4.合作原理:無論評量計畫之擬訂、評量工具之設計與編製、評量之實施及

其結果之分析、解釋與應用，均須集結全校師生的力量，大家共同參與，彼此協調合作，始能順利進行，達成教學評量的任務。

5.研究發展原理：評量研究乃一體之兩面，關係至為密切。為求突破目前教學的瓶頸，革新教學措施，以提高教學效率，必須運用評量的方法和技術，進行教學的實驗研究，使理論與實際相互印證，並開拓教學與評量之研究發展的新途徑。

學校進行成績評量時，應注意下列幾項原則：

- 1.成績評量應以學生身心成熟的程度為依據：因為學生畢竟是學生，他是兒童，是青年，而不是成人，所以我們就不應以成人之標準去要求學生。其次，學生與學生之間，其身心發展成熟程度也不一致。在教學時要顧及學生之個別差異，在進行成績評量的時候，也不應忽略學生之個別差異的重要性。
- 2.成績評量是多方面的：過去一般人以為學校的成績評量只限於知識技能的考查，事實上現代國民中學所強調的是全面教育，應該把評量的範圍擴大至教育的全部。除了知識技能之外，應包括學生的品行、人際關係、學習的態度、興趣和方法習慣等方面。當我們從事某一學科的教學時，雖然以該學科本身為重點，但也不能忽略與該學科有關的知識，而且透過各學科之學習，學生在其態度及理想上可能有所改變或進步。因此副學習與副學習之結果，也應納入成績考核的範圍，不要僅以主學習作為成績評量之唯一對象。最近美國心理學者認為學校教學包括(1)認知方面--以學生心智能力之發展為核心；(2)情感方面--以學生之態度、興趣、對人對事之方式以及各種鑑賞能力為重；(3)心理動作方面--以各種技能為主。凡此無一不屬於成績評量之範圍。
- 3.成績評量必須採用多種方式：因為成績評量之範圍非常廣泛，所以應該採取不同的考查方法，以配合教材之內容進行多種方式的評量。評量時所蒐集資料越齊全，其結果越客觀，越能符合成績考查的要求。
- 4.成績評量應注意學生平時的學習活動：除了舉行月考、期考之外，應將學生平時學習的情形予以記錄，這樣才能了解學生在學習過程中的缺點，

指導他及時改正。應該避免到學期之末，才將學習成績算一總帳。如能隨時注意學生進步的情形，使學生了解本身進步的狀況，必能增進其學習的興趣，提高其學習成就水準。

5. 成績評量應由教師、學生與其他有關人員共同參與：以往一般人認為成績評量是老師的責任，只有老師有權力來評定學生學習的成果，但從學生學習成果綜合評定的觀點來看，教師、學生家長、學生本人及其他有關人員均可提供成績評量的各種資料。根據這些多方面的資料，才能作一個比較周全的評斷。（簡茂發，68年）

三、 多元化評量的意義、方式與各種評量方法之比較

（一）、多元評量的意義

Wolf, Bixby, Glen, & Gardner(1991)強調紙筆測驗雖然能測量認知領域的學習結果，但在技能、情意領域則有其先天限制，如文章、圖畫、實驗報告、自然科學展覽設計、演講、握筆、打字和小組合作學習能力均難以運用紙筆測驗評量。教師教學必須兼顧認知、技能、情意之學習結果，且評量應分析「應該怎樣表現」（知識）、「真正表現行為」（實作）之間差異，方能檢討教學成效與實施補救教學，教師宜善用「多元評量」方能適切評量學生真正的各項學習表現。

李坤崇(民 88、民 90a、民 90b、民 91a)認為傳統以紙筆測驗為主的教學評量，具有計分客觀、批閱迅速、易於團體施測、及激勵學生認知學習等優點，且能充分發揮公平、客觀、省時、省錢的功能，然卻衍生下列十二項問題：(1)評量目標較少顧及教學目標。(2)評量內涵忽略技能、情意。(3)評量方式過於偏重紙筆測驗。(4)評量時機較忽略形成性評量。(5)評量未能營造公平、良好施測情境。(6)評量結果解釋較少鼓勵 增強學生。(7)評量結果解釋過於依賴量化測量。(8)評量認知過於強調記憶層次。(9)命題觀念與技術仍有待加強。(10)測驗試題編排未以學生為中心。(11)評分缺乏客觀標準與自省思維。(12)家長分數至上觀念難以消除。可見，傳統教學評量與人性化、多元化的教學評量相去甚遠。

Lazear(1999)認為教師實施多元教學的意願高於多元評量，其因有二，一為教師較缺乏多元評量經驗與知能，實施較不易獲得家長與同儕支持；二為家長不

習慣看自己與他人孩子的分數，更不習慣去分析與鑑定孩子的各項學習與關鍵能力。然而，身為教師應當為所當為，而非取決於個人意願。教師當務之急在於提升評量專業素養，而非迴避多元評量。

張稚美(民 89)提出紙筆測驗、教師觀察(如真實或實作評量)、多元智慧成長檔案之三角檢核法(詳見圖 2-1.1)，來評量學生學習成果，且此三項各佔學年成績的三分之一。紙筆測驗以評量學生認知層面的學習結果。教師觀察係教師先系統建立評量知能與評量指標，再對學生實施有目的的「觀察」與「記錄」，此方式頗能評量學生技能、情意層面的學習結果。成長檔案評量可完整呈現學生整個學習歷程，提供學生自我評量與習慣性省思的機會，若再將家庭納入評量學生學習與成長歷程，更能為學生搭起學校與家庭的橋樑。

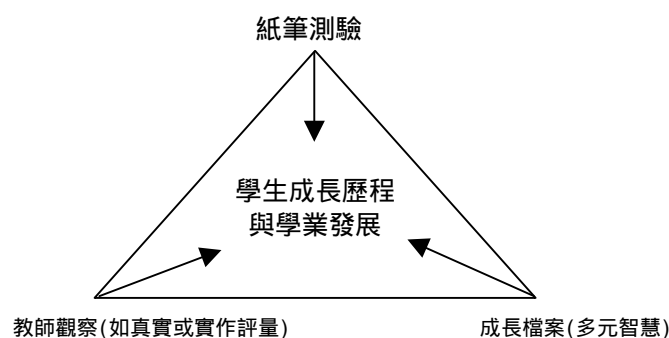


圖 2-1.1 學生的學習成長歷程評量結構圖

張稚美(民 89)雖然針對圖 2-1.1 發展出系列的評量方式與具體策略，亦能兼顧紙筆測驗、教師觀察、成長檔案三種評量方式，且各佔三分之一，然而評量方式可更多元，評量人員可更多元、且評量內涵不一定侷限於多元智慧。

Gardner(1993)從不同的思考向度去看學生的學習成長，強調未來評量將具有下列的特色：(1)評量(assessment)重於測驗(testing)：通常測驗為幾個目的而實施，無法確實反映出外在環境實際的需求，而測驗更可獲得個人技能或潛能，提供個人有效的回饋及外在環境有效的訊息。(2)評量乃教學歷程中，簡單、自然發生的措施：評量乃教學歷程中的一部份，應與簡單、自然的學習情境高度結合，而非外加。(3)評量應具生態效度(ecological validity)：評量宜在「真實工作情境」的相似狀態下進行，方能對個人的最終表現做出最佳的預測。(4)評量工具應具「智慧公平」原則：大部分測驗工具著重語文及邏輯 - 數學兩種智慧，然此兩種

智慧較強者評量結果通常亦較佳，工具應顧及各種智慧強勢者。(5)利用多元測量(measures)工具：評量宜採取多元的測量工具，以測出不同能力的各種面向。(6)顧及個別差異、發展階段及各種不同的專業知識：評量時宜考量個體的個別差異、不同發展階段，及不同專業知識，所產生評量結果的差異。(7)評量素材趣味化：評量應利用一些有趣及能引起動機的素材，讓學生樂於進行。(8)以學生利益為前提：評量旨在幫助學生瞭解其學習優缺點，應將評量訊息回饋給學生。

田耐青(民 88)提出符合多元智慧理念的評量，應具備下列五項原則：(1)評量長期化：長期評量可對學生的作品或表現進行連續性的觀察，並可提供學生對學習做持續性的反省。(2)評量多元化：評量內涵包含內容和技巧的評量，評量人員包括教師、家長、與學生。(3)評量回饋化：評量結果應該能為教學提供資訊，回饋予學生、教師及家長。(4)評量兼顧非正式評量：學生在班級互動的參與度與日常表現，反應學習態度應予兼顧。(5)評量激勵主動的自我評量：學生有能力評估自己學習的優缺點，能清楚說出自己學到的知識或概念，更能確認自己應如何運用適當的思考與學習歷程。

綜合田耐青(民 88)、李坤崇(民 88、民 90a、民 90b、民 91a)、張稚美(民 89)、Airasian(1996)、Gardner(1993)、Lazear(1999)、Wolf, Bixby, Glen, & Gardner(1991)之觀點，闡述多元評量的意義，多元評量係以教師教學與評量專業為基礎，依據教學目標研擬適切的評量方式、評量內涵、評量人員、及評量時機與過程，並呈現多元的學習結果，以提供更適性化的教學來增進學生成長。

(二) 多元評量的方式

評量應依據評量目標選取適切的評量方式，而非僅限於單一的客觀紙筆測驗。國中小教師依其評量目標常選取的評量方式，至少包括紙筆測驗、實作評量、軼事記錄、口語評量、檔案評量、遊戲評量、動態評量等七項，茲扼要闡述理念，並附實例說明之。

1. 紙筆測驗

多元化評量並非廢除紙筆測驗，而是降低紙筆測驗在現今教師評量的比率，加重其他評量方式的比率，依據評量目標彈性運用各種評量方式來適切評量學生學習成果。然而，現今國中、國小教師因在師資養成階段，修習評

量學分數遠低於修習課程、教學學分數，使得教師在紙筆測驗的專業素養仍有待加強。

2. 實作評量

余民寧(民 86)、李坤崇(民 88)、吳鐵雄、洪碧霞(民 87)、陳英豪、吳裕益(民 80)、Airasian (1996)、Aschbacher(1991)、Fitzpatrick & Morrison(1971)、Linn & Gronlund(1995)、Stiggins(1987)、Wiggins(1992)等觀點，實作評量定義為：具相當評量專業素養的教師，編擬與學習結果之應用情境頗為類似之模擬測驗情境，讓學生表現所知、所能的學習結果。國中國小常用的實作評量包括實作評定量表、實作檢核表、軼事記錄、口語評量、檔案評量、遊戲評量等六項。

李坤崇(民 88)指出實作評量具有下列特質：(1)強調實際生活的表現；(2)著重較高層次的思考與解決問題技巧；(3)重視學生學習個別差異；(4)適於年齡幼小、發展較遲緩學生；(5)促進學生自我決定與負責；(6)講求評分、標準與人員的多元化；(7)強化溝通與合作學習能力；(8)兼重評量的結果與歷程；(9)著重統整化、全方位、多樣化的評量；(10)強調專業化、目標化的評量；(11)強調教學與評量的統合。

實作評量能彌補傳統紙筆測驗過於僵化、內容與現實脫節、重視結果忽略歷程等缺失之不足。然與傳統紙筆測驗比較，實作評量具有下列缺失：(1)較要求應用於實際生活使得設計較不易；(2)評分較客觀的紙筆測驗主觀，較易受評分者個人特質影響；(3)評分較客觀的紙筆測驗費時費力，較難用機械計分；(4)較難進行團體間學習結果之比較(李坤崇，民 88；余民寧，民 86；陳英豪、吳裕益，民 80；Airasian，1996；Linn & Gronlund，1995)。

3. 軼事記錄

軼事記錄乃教師觀察學生日常生活表現，詳細寫下重要而有意義的偶發個人事件和行為的紀錄。軼事記錄通常做為評量佐證資料，而非評量的唯一依據，因學生日常生活的點點滴滴難以在教師設計的評定量表、檢核表、檔案資料、或紙筆測驗中完全呈現，若能以教師在班級情境的直接觀察作為佐證資料，當能提高評量結果的效度(李坤崇，民 88)。

4. 口語評量

常用之口語評量有二：一為「口試」(Oral Examination)、一為「問問題」(questions)。「口試」較常用於總結性評量，如國語可用演講、辯論、口頭報告、經驗分享、故事接龍來評量，數學採放聲思考，解題經驗分享、日常應用心得分享、口頭報告、表演等方式來評量。「問問題」較常用於形成性評量，教師於教學過程以問題問學生乃常見的師生互動模式，只是較少教師將問問題納入教學評量，將其視為教學評量的一部分(李坤崇，民 88)。

5. 檔案評量

「檔案評量」在傳統紙筆測驗難以評量學生適應未來能力、其他評量方式難以讓學生充分自我表現、及強調學生自主學習的思潮下，漸受重視。用之「學生」的檔案評量，可用於九年一貫課程七大學習領域，亦可用之特殊教育。用之「教師」的檔案評量，我國未來即將實施的教學評鑑、教師分級制度，美國已廣泛用於職前教師課業、畢業資格、求職證照核發、新進與實習教師的專業成長與教師評鑑、經驗教師的評鑑、教師證照的再核發、教師績效獎金的核發等。可見，檔案評量將是廿一世紀寵兒。

「檔案評量」旨在突破以班級為單位，改以學生個人為單位，請每個學生均設計與製作個人數學學習檔案，就特定主題連續收集資料，經綜合統整呈現，以系統的展現學生個人學習數學的歷程與成果。為達成教學與評量的結合，教師宜與學生充分討論以決定單元目標與檔案內容，且學生必須參與整個評量過程，方能建構出創造、有意義的數學學習，評量時除教師評量外，尚可請學生自評、家長評量、同儕互評(李坤崇，民 88)。

學習檔案的內容可依其評量和使用目的而有不同的蒐集重點和組織結構，Henderson(1995)將學習檔案分為下列五種類型：(1)展示檔案(showcase portfolio)：展示檔案呈現學生自己挑選的作品樣本，旨在慶祝或分享成就，展示檔案製作係請學生從其收集的作品中挑選出自己最滿意或最喜愛的作品，並附上作品的說明和省思組織而成的，展示檔案可顯示學生的個人特性、自主權和成就。(2)紀錄檔案(documentation portfolio)：紀錄檔案係定期挑選學生的在特定技能或學習領域的代表作品組成的檔案，通常由老師

和學生一起選擇，旨在觀察學生在一段時間持續的進步情形。(3)評鑑檔案(evaluation portfolio)：評鑑檔案係根據教學或評量目標設計檔案內容，要求學生根據規定內容選擇或製作項目而成的，通常用於評定學習成果或大規模評鑑，如學校校務績效評鑑、課程與教學評鑑。(4)歷程檔案(process portfolio)：歷程檔案乃呈現一個活動單元或專題研究所有的材料，目的在描述學習歷程的努力、進步和成就，或進行學習診斷。(5)綜合檔案(composite portfolio)：綜合檔案兼具上述檔案類型的特點，實際應用上較有彈性，然因其兼顧數項使用目的，實施上需花費較多的時間和精神；另外，教師須先充分掌握綜合檔案的使用目的，方能有目的、有系統地蒐集學生有意義且有代表性的作品和表現，來達成既定的評量目標。

檔案評量可適切評量學生應用、推理、分析、綜合、評鑑等高層次認知行為，讀、說、寫與其他實作技巧等技能，作文、各種報告、美術、音樂或其他藝術或科學作品等成果，以及學習態度、性序、學習動機、努力情況、求知精神等情意，漸受教師的青睞與使用，然為求更適切運用，提出下列原則供參酌：(1)檔案評量必須與教學相結合。(2)檔案評量應與其他評量並行，檔案評量不應作為評量學習結果的唯一評量工具，尚必須輔以其他評量方式或工具，如傳統紙筆測驗、口試或公開展示方式。(3)檔案評量應實施多次、階段的協助或省思，請學生持續一段時間主動收集、組織與省思學習成果的檔案，以評定其努力、進步、成長情形。(4)檔案評量應顧及可使用資源與學生家庭背景差異。(5)實施檔案評量應採漸進式、引導式模式，由觀摩檔案範例、再製作小規模檔案、後製作較大規模檔案。(6)檔案評量若用之評鑑應力求慎重。

6. 遊戲化評量

遊戲化評量不僅能激發學生參與興趣，更能讓學生在遊戲中評量、在遊戲中成長。較常見的遊戲化評量方式乃過關評量、分站評量、踩地雷、猜猜看、填空高手。對語文程度較低、語言表達能力較弱的學生，難以從紙筆測驗、專題報告或檔案評量來評量時，遊戲化評量相當適用。遊戲化評量讓學生參與遊戲，身歷其境通常對學生均極具吸引力，小從小學、大到研究所的學生均喜歡以遊戲化取代呆板的紙筆測驗或其他評量方式(李

坤崇，民 88)。

為發揮遊戲化評量的功能，減少可能缺失，李坤崇(民 88)建議教師實施遊戲化評量時，注意下列原則：(1)遊戲不可與教學目標脫節，不宜流於純粹遊戲。(2)以安全為最重要原則，設計遊戲應多徵詢其他教師、主任、校長對活動安全性的意見與評估。(3)擬定嚴謹實施計畫，執行確實與預留人力，用心檢討做成記錄，彙整成一份遊戲化評量資料檔。(4)因年齡不同選用適合的活動，設身處地的站在遊戲化評量對象的角度，來設計屬於該年齡層的活動，而非設計教師喜歡的活動。(6)說明以室內或書面資料為主，輔以戶外或口頭說明，避免學生因吵雜漏聽遊戲規則，更可留下資料作為往後參考。(7)活動單設計力求美觀大方，提高學生參與興趣。(8)協助活動或評量者的行前講習，以提高活動安全性與結果正確性。(9)教師或協助者宜事前模擬遊戲，務求減少意外事件。(10)循序漸進、累積經驗、自我增強，避免過高自我期待衍生挫折。

7. 動態評量

動態的兩層意義：(1)了解受試者動態認知歷程與確認認知能力的變化情形，著重評量學習歷程與認知改變。(2)著重評量者與受試者的互動關係，強調評量與教學結合(Haywood, Brown, & Wingefeld, 1990)。因此，動態評量乃教師運用「前測--教學介入--後測」的主動介入模式，經由充分溝通互動歷程，持續評析學生教學反應與學習歷程，剖析教學前後認知能力的發展與改變，進而提供發展或改變所需的教學介入的評量方法，可見動態評量乃一種結合教學與診斷的評量模式。

綜合江文慈(民 82)、邱上真(民 85)、莊明貞(民 88)、莊麗娟(民 85)、Haywood, Brown, & Wingefeld(1990)、Haywood, Tzuriel, & Vaught(1992)、Lidz(1991)等人的觀點，動態評量之特性如下：(1)兼重學習結果與學習歷程：Campione, & Brown(1987)認為動態評量不僅評量過去既有知能或經驗的靜態結果，更在於評量成長、改變的動態歷程與學習預備度。(2)兼重回溯性評量與前瞻性評量：動態評量不僅著重目前的表現與成就(回溯性評量)，亦重視未來發展的潛在水準(前瞻性評量)。Campione, & Brown(1989)認為動態評量除評估學生目前的表現水準外，亦重視學生如何達到此水

準，及最大的可能表現水準。(3)兼重鑑定、診斷與處方：動態評量除鑑定與診斷學生學習潛能與認知歷程外，更提出處方性的教學訊息，來協助學生學習。(4)著重認知能力的可塑性：傳統評量認為智力本質是天生、不可改變的，但 Lidz(1991)強調認知能力的可塑性(modifiability)與活動性(activity)。(5)著重師生雙向溝通的互動關係：動態評量遵循「前測--教學介入--後測」的程序，教師於評量學生時，不是中立的觀察者，而是協助學生學習的教學者，扮演雙向溝通互動的協助者。(6)融合教學與評量：傳統評量區隔教學、評量為兩種教育歷程的不同概念，遵循「先教學而後評量」的教育模式。動態評量將教學與評量融合，採取「在教學中評量、在評量中教學」的教育模式。

(三) 各種評量方法之比較

教學評量可分為「形成性評量」(formative evaluation)和「總結性評量」(Summative evaluation) 兩種。前者係在教學過程中，就教師的教學情形與學生的學習表現加以觀察和記錄，通常採用評定量表為工具，進行非正式的評量；後者係在教學活動之末或結束之後，以定期考試或測驗的方式考查教師的教學成果與學生的學習成就，通常採用標準化學科測驗及教師自編課堂測驗為工具，進行正式的評量。茲將上述兩種評量方式與一般所謂的「診斷性評量」(diagnostic evaluation)加以比較如表2-1.2：

表2-1.2 診斷性、形成性、總結性之比較

比較項目	診斷性評量	形成性評量	總結性評量
功能	決定學生的成熟度、預備狀態、起點行為、與學習有關的特質，予以分組並診斷學習困難的原因。	提供學生進步的回饋資料，指出教學單元結構上的缺陷，以便實施補救教學。	在某一教學單元、課程或學期之末，就學生們的學習成就進行評量，決定其成績的等第、及格與否。
時間	教學之初或學習困難時	教學進行中	教學之末
評量重點	認知、情意、技能方面的行為、身心及環境因素。	認知方面的行為表現。	一般以認知行為為主，但有些科目也涉及技能、情意方面的學習結果。
工具類型	學前診斷、標準化成就測驗、診斷測驗、教師自編測驗、觀察和檢核表。	為教學需要而特別設計的評量工具：評定量表、作業及其共同訂正、口頭考問、實際演示、問題研討。	期末或教學單元結束時的考試。
行為目標 行為樣本之選擇	必備的起點行為；單元目標的要項和教學有關的學生身心特質及環境因素所涉及的行為。	教學單元層次結構中所有相關的行為項目；教材細目、學習的動機、態度、方法、習慣等。	依教學目標和教材內容的相對重要性而擇定評量項目，使其有適當的比例分配。
項目難度	為診斷必備的知能和基本技能，大部分試題是簡易的，通常 P 在 .65 以上	隨實際情況的需要而異，未能事先決定。	大部分的試題的 P 在 .35 至 .70 之間，但很容易、很困難的試題也有
計分	常模參照和標準參照	標準參照	通常是常模參照，但有時也可能是標準參照
通報分數的方式	把各方面的知能程度以側面圖呈現	把個人在教學單元層次結構中的各項細目及格與否的組型呈現出來，以供了解個別情形。	根據行為目標列出各項分數和總分

教學評量又可分為「常模參照的評量」(norm-referenced evaluation)與「標準參照的評量」(criterion-reference devatuation)兩種。前者以同年級學生學業成就的平均水準為參照點，比較分析學生之間的差異情形；後者則以事前決定的絕對性標準為衡斷的依據，判定學生成效的及格與否，而賦予優劣的評斷。茲將此兩種評量方式比較如表2-1.3：

表2-1.3 常模參照的評量與標準參照的評量之比較

比較項目	常模參照的評量	標準參照的評量
量尺定準點	中間，事後決定	兩端，事前決定
參照點性質	實際的、相對的	理想的、絕對的
評量功能	鑑別；比較團體成員之間的差異，找出最具有潛能的	鑑定；找出超乎某一特定的能力水準以上
數據性質	分數的變異性愈大愈好	注重各題反應與效標之間的關聯性
結果表示法	百分等級，標準分數	及格或不及格
計分制	常態等第制	傳統百分制

第二節 科學概念發展

一、科學概念的科學史發展

科學可分經驗科學和非經驗科學兩類。韓培爾（C. G. Hempel，何秀煌譯，民 58），指出：「非經驗科學是指像純數學和邏輯，他們的命題（propositions）基本上可以不經過參照經驗發掘而加以證明。經驗科學志在發掘、描述、說明和預測那些在我們生活中，所發生的種種事項。因此命題必須以我們經驗中的事實檢驗。同時，命題只有在他們有著適當的經驗證據的支持之下，才能夠被接受。」人類會作經驗科學的探討，以致會做結構性的科學探討是歷史發展的結果。（鍾聖校，民 79）

化學家 Nash（1963）認為科學是察看世界的一種方式。考察歷史上人類察看世界的方式，是經過感性直觀、直觀外推、純粹理性思考、機制的類比，而進入注意結構關係的思考。

（一）感性直觀

所謂感性直觀是指用感覺能力、未經概念化、未經推論，而直接掌握和認識事物的形象及特殊性。直觀的積極意思是對事物做立即和直接的認識（apprehension），消極意思是沒有任何推論過程（reasoning process）的介入。例如古希臘的重要哲學學派—畢達哥拉斯學派、愛利亞學派等等，以不同的解釋方式說明某特定的「感性存在物」是世界的統一基礎。例如水、土、氣、火（四元）是萬物生滅變化的根源，他們共同的特點是不在借助於超自然的神力觀，而以生活經驗中的自然現象來說明自然（發靈頓，B. Farrington，程石泉譯，民 57）。

（二）直觀外推

直觀外推是指以人的行為和感覺外推，來解釋自然現象。卡普拉（F. Capra，蔡伸章譯，民 75）認為理性與直觀乃是人類心智功能的兩種相輔相成的模式。「理性思考是直線型、集中化及分析性的。它屬於智能的領域，它的功能在分解、測度與分門別類，因此理性知識往往是細分性的。而在另一方面，直觀的知識則是建基於意識的擴大而產生的一種直接的、非智能的實體經驗，因此，它往往是綜合性的，非直線型及整全的。」

卡普拉認為過分強調理性思考容易造成文化的失衡，以西方來說是科學主義宰制了文化的各層面。但反觀中國，則因直觀外推思考方式的氾濫而影響了科學發展。

（三） 理性思考之不足

理性思考固然有助於分析事理，但是自然界現象法則的發現，卻不是單純的靠理性思考和觀察就可以達成。

以亞里斯多德為例，它注重系統的觀察紀錄和理性思考研究科學。在分析宇宙「實在」(reality) 之後，發現它們具有形式因、物質因、動力因及目的因（傳偉勳，民 57）。科學研究的目的只在獲得有關形式方面的知識。

發靈頓 (B. Farrington , 民 57) , 認為「形式」論的科學觀應用於動植物的研究頗見功效，但不適合用在物理學。物理過程不像生物過程那樣，可以在任何事物上目睹一種形式趨於完成。因此，發靈頓說：「物理科學之發展必須借重於”能力觀”，而非”形式論”」

李約瑟 (民 73) 在「大滴定」一書中指出：用「機制的類比」(analogy of mechanism) 來說明自然，是導致現代科學成功的最基本概念之一。

（四） 機制的類比

「機制的類比」實際上是一種研究事理的觀點，是人類注意到宇宙事物存在的規律性、關聯性和計量性。劉清峰 (民 77) 以「結構觀」來表達這個概念，並且較詳細的指出結構性思考的特點。

結構性思考有下述特點：

1. 結構性思考是把不同的經驗類比起來，整個理論體系中要顯示出一致的有序性。
2. 結構性思考需要突破本體觀。所謂本體觀是指不同類事物，在宇宙中處於不同等級的位置，屬於性質完全不同的東西。
3. 結構性思考對不同事物的思考是指採取其某一性質來考察，把它數量化，看做一種關係結構的一部份。

二、科學概念形成在認知心理學基礎上

在心理學派或研究模式中，主張認知作用的性質是注意事物間關係的，大致為格式塔心理學、皮亞傑理論和訊息處理模式中的建構論。

（一） 格式塔心理學

由魏泰瑪（M. Wertheimer, 1880-1943）柯勒（W. Kohler, 1887-1967）及考夫卡（K. Koffka, 1886-1941），分別以不同的方式說明認知的本質是注意到事物間的關係。

（二） 皮亞傑理論

從皮亞傑觀點來看，在任何與環境接觸的時候，認知必然同時含有同化和順應。藉同化和順應作用，人內在的認知結構與外在事物呈現的關係結合起來，形成某種認知結果。

皮亞傑的「同化—順應」模式要求探討者提出關係假設是符合人類認知本質的。因為它說明：（1）吾人內在認知結構系統可以同化外在現象；（2）認知結構系統也受外在事象的關係形貌規制，由順著外在事象的形貌關係而於內在認知結構之上，在形成一種組型（pattern），始能與外在事物關係相對應。

（三） 訊息處理模式

Siegler（1983）指出訊息處理的研究方式，可能在目前研究認知發展的策略中居於領導的地位。它把人的心靈類比像電腦一樣，認知系統運用或操作從環境中送入，或在心理系統中以儲存的訊息。

Neisser（1967）分析訊息處理的認知理論，發現可分為兩種幾近相反的觀點；其一與聯結論一脈相成，另一則傾向結構論，可稱之為基模建構模式（Schemata Construction Model）。能夠充分解釋結構科學探討可能性的是後者。基模建構模式的認知觀點是承襲巴雷特、皮亞傑、夏赫特、托爾曼……等人對結構或建構行為的重視，主張在心理歷程中，回憶和思想都是建構的歷程。

第三節 九年一貫自然領域之科學能力指標

本節宗旨在於瞭解現階段台灣的教育政策，實施國中小九年一貫課程改革的課程目標、國中階段學生的分段能力指標與與探討科學素養的真實內涵。

一、九年一貫課程改革的課程目標

台灣現階段的新課程改革，即九年一貫課程是依據民國 89 年 9 月 30 日的 台(89) 國字第 89122368 號令所頒佈的課程綱要公佈實施。

新課程的基本理念是希望將科學的知識再拉回到自然事物及自然現象本身，從觀察自然萬物的變化去思考其因果法則。由認知這些自然現象和自然的演變規則，使學生能應用自然運作的原理，進一步能有各種創造與發明。回到科學的本質讓我們藉由學習科學，學會如何去進行探究活動：學會觀察、詢問、規劃、實驗、歸納、研判，也培養出批判、創造等各種能力。特別是以實驗或實地觀察的方式去進行學習，使我們獲得處理事務、解決問題的能力。瞭解到探究過程中細心、耐心與切實的重要性。同時我們也應該了解科學與技術的發展對人類生活的影響，學會使用和管理科學與技術以適應現代化的社會生活。透過學習使我們能善用各種科學與技術、便利現在和未來的生活。

自然、科學、技術三者一脈相連，前後貫通，我們對其有以下幾點基本的認識：

1. 自然與生活科技的學習應為國民教育必需的基本課程。
2. 自然與生活科技的學習應以探究及實作的方式來進行，強調手腦並用、活動導向、設計與製作兼顧、知能與態度並重。
3. 自然與生活科技的學習應以學習者的活動為主體，重視開放架構和專題本位的方法。
4. 自然與生活科技的學習應該培養國民的科學與技術的精神及素養。

並依照以上的這些基本認識來規劃課程。所以新課程的課程目標是：

1. 培養探索科學的興趣與熱忱，並養成主動學習的習慣。

2. 學習科學與技術的探究方法及其基本知能，並能應用所學於當前和未來的生活。
3. 培養愛護環境、珍惜資源及尊重生命的態度。
4. 培養與人溝通表達、團隊合作以及和諧相處的能力。
5. 培養獨立思考、解決問題的能力，並激發創造潛能。
6. 察覺和試探人與科技的互動關係。

二、九年一貫的教學目標

國民中小學九年一貫課程最主要的特色就是以培養生活所需的「基本能力」做為課程目標。因此各學習領域課程綱要內所提示的教學目標也都以該學習階段學生所應養成的能力來表述，稱之為「分段能力指標」。本領域的這些能力統稱之為「科學素養」。

「科學素養」乃是經由科學性的研討活動可以獲得的「學習成果」；由於學校裡「教學」的概念一再的演進(例如美國六〇年代的科學教育特別加入了「操作中學習」的強調，而有「科學過程技能」教學目標被標示出來。八〇年代提倡的 STS 教育理念和建構主義的教學模式，而特別提及科學應用、創造力等能力的培養)，教學的「目標」由原初著重「科學概念的知道、理解」推展到同時注重科學過程技能、思考智能(或稱解決問題能力)、科學知識與技能的應用，強調批判、創造力(創思、設計、製造...)等思考的習慣和能力的學習，以及比較具有整合性的素養如對科學本質的體認、科學態度等等的養成。

所以說，由於社會活動演進快速，個人生活在一個資訊發達、環境多變的情境下，所需的「能力」也不同，教育既然在培養國民適應生活與改善生活環境的能力，因此社會環境的變遷、教學模式的改變(也即研討活動的內容、方式的改變)，學生獲得的學習內容也不同(見圖 2-3.1 所示)(陳文典，92 年)。

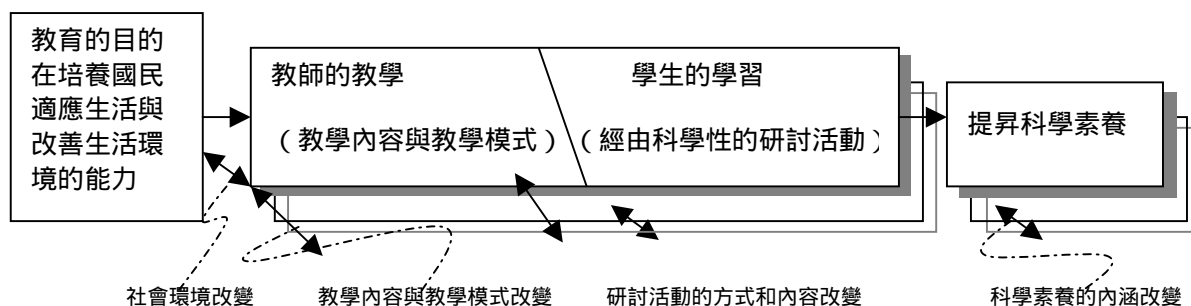


圖 2-3.1 (陳文典，92 年)

本領域先將總課程目標「培養國民生活所需的十項基本能力」，轉換成「科學素養」，而以科學與技術認知、科技發展(的認識)、(科學)過程技能、思考智能、科學本質(的體認)、科學態度、科學應用、設計與製造等八項來表述。並在學童成長的各個學習階段上依此八項科學素養訂定教學目標，為「分段能力指標」。因為在某學習階段的教學應達到這個「指標」，所以，「指標」就是我們在該階段要評量的基準。為了更清楚瞭解該項能力指標的意義，可以參閱該項的能力在全部四個學習階段的指標(見附錄三)，全部閱讀之後將可得知該項能力的「涵義」，比較上下在不同學習階段的指標可以得知李生成長發展的情形。這樣，在評量時才能掌握內容，且知道學生學習的優(超前)、缺(落後)情況(陳文典，92 年)。

三、國中階段學生的分段能力指標

「素養」蘊涵於內即為知識、見解與觀念，表現於外即為能力、技術與態度。實際上內外之分，也僅係應陳述之方便而已。自然科學學習所培養之國民科學素養，依其屬性和層次來分項，分成科學探究過程之心智運作能力(以下簡稱「過程技能」)，科學概念與技術的習得(以下簡稱「科學與技術認知」)，對科學本質之認識(以下簡稱「科學本質」)，瞭解科技如何創生與發展的過程(以下簡稱「科技的發展」)，處事求真求實、感受科學之美與力及喜愛探究等之科學精神與態度(以下簡稱「科學態度」)，資訊統整、對事物作推論與批判、解決問題等整合性的科學思維能力(以下簡稱「思考智能」)，應用科學探究方法、科學知識以處理問題的能力(以下簡稱「科學應用」)，以及如何運用個人與團體合作的創意來製作科技的產品(以下簡稱「設計與製作」)等八項來陳述。其中「科學認知」涉及教材內容，其內容詳列於參考資料中(附錄三：自然與生活科技學習

領域教材內容要項)，作為教學活動設計之參考。茲將自然與生活科技課程，學生在學習各階段所應習得之能力指標列之於後。在設計教學活動時，宜依指標所提示的基準，於教學中達成之。

1. 過程技能
2. 科學與技術認知
3. 科學本質
4. 科技的發展
5. 科學態度
6. 思考智能
7. 科學應用
8. 設計與製作

經由科學性的研討活動，可以增長其人的能力，我們將之稱為科素養的提昇。那麼，「科學素養」這個「能力」，究竟有多少的內涵呢？為了敘述上的方便，在本領域課程綱要上把它分成八項，而這八項又可歸成四類（陳文典，92年）：

1. 概念與技術認知類：科學與技術認知、科技發展(的認識)。
2. 探討與思考智能類：(科學)過程技能、思考智能。
3. 科學精神與態度類：科學本質(的體認)、科學態度。
4. 應用科學解決問題類：科學應用、設計與製作。

茲將科學素養包含的各項能力列在表 2-3.2 中（陳文典，92年）。

表 2-3.2 科學素養包含的各項能力

科學素養	二分法	四類分法	八項	能力表徵
	知識認知	概念與技術認知類	科學與技術認知	S ₁ 知道、理解及運用科學概念 S ₂ 會操作儀器及有製作的技術
			科技發展	S ₃ 科技發展 (的認識)
	科學智能	探討與思考智能類	過程技能	S ₄ 探討某問題時，具有觀察、比較分類、組織關聯、研究推斷等心智運作能力 -----L ₁ 傳達 (溝通與表達) -----
思考技能			S ₅ 綜合能力 S ₆ 推論能力 S ₇ 批判能力 S ₈ 創造能力 S ₉ 解決問題能力 (評估問題、提出策略、分派工作、安排流程、執行實現、評鑑成果)	
科學智能	應用科學解決問題類	科學應用	-----L ₂ 知識、技術應用於處理問題中----- -----L ₃ 科學方法及思考習慣運用於處理問題中-----	
		設計與製造	S ₁₀ 有設計及製作技能 -----L ₄ 為自己的需要去設計與製作出作品-----	
科學智能	科學精神與態度類	科學本質	-----L ₅ 科學本質的體認 -----	
		科學態度	-----L ₆ 投注與熱忱 ----- -----L ₇ 細心與切實 -----	

表中 S₁ 至 S₁₀ 項是比較可以「短時間」觀測到的，例如由紙筆測驗、實驗報告、課堂上問答、專題報告、現場觀察等方式評測可得到。由表中 L₁、L₇ 都是歷經長期的學習之後，表現於處理問題、解決問題的能力上、以及做事的態度與精神。這些學習的成果需要由學生長時間的表現去做評斷(例如某學生最近參與探討的態度改變了，顯得自信、積極而且興緻勃勃，或是作品的精細度突然改進不少等等)。

四、科學素養的內涵

(一)、科學

「科學」，以知識的內涵來說，科學可以說是「運用科學方法所獲致的知識」。那麼，什麼是科學方法呢？

人類智識的成長與累積，形成現在的文明社會。在我們生活中所運用的各種智識，有些則是價值的判斷、目標的取捨，有些則是傳達能力，有些則是執行能力。為了傳述的方便，我們勉強把它分為宗教、藝文、科學來說明。宗教是個人的一種價值觀，一種對生命的態度。它看來隱而不顯，但是卻影響整個人的生活模式。藝文則是語言、音樂、藝術、舞蹈、衣著等傳達感情意念的能力。而科學則是依據事實或是以事實來考驗的知識，依賴這些知識，使我們真實不虛的去選擇及做合理有效的行動。

我們可以參酌認知心理學(其實也不一定必要，就直接運用自我經驗的反省即可)，來探討人們如何運用外在的資訊，來建構或激發他的思想，如圖2-3.3 (陳文典，92)：

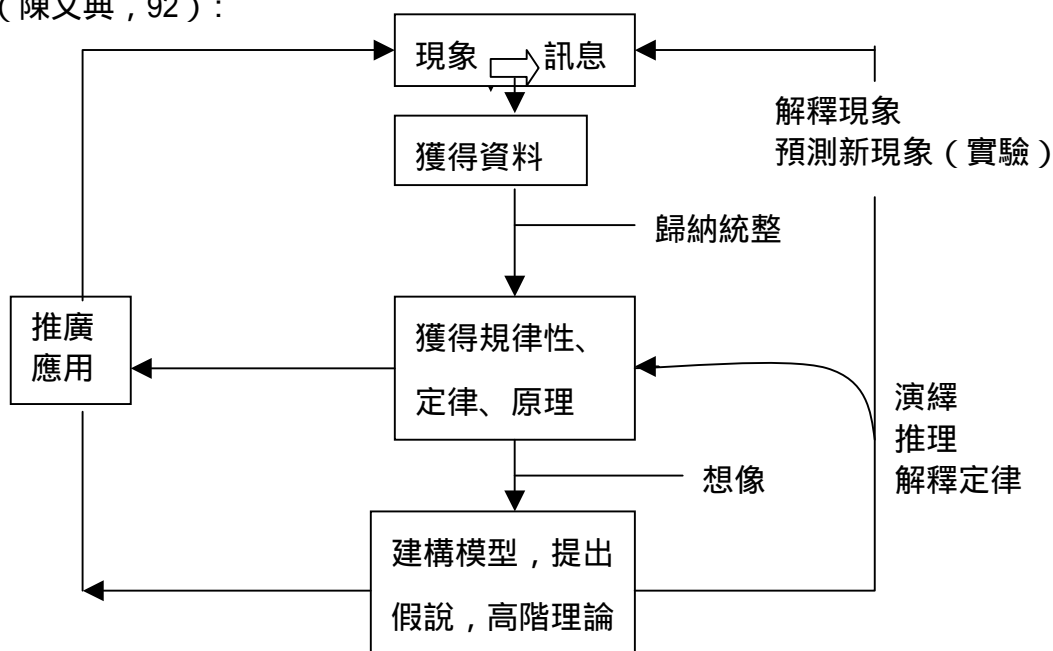


圖2-3.3 認知建構與發展過程

在這個認知的建構與發展的過程，歸納法是將眾多的資料和知識加以批判、統整而獲得「通則」的一種思考過程及方法。因此，通則的成立，是建立在這些事實之上的。不過，有時，遇合這些事實的通則可能不只一個，所以通則的

成立與否，還得受到更多資料、事實的考驗，或是和既有的理論、模型相互辯證。歸納法所獲得的知識，可說是「到目前為止，沒有論據可以證明它錯」的知識。

至於假說或理論的提出、模型的建構，則是先要在心中有一個一個想，認為真相應該是「那個樣子」，再依「設A為真，則B亦為真，或C必為偽」等推理，來解釋各種已知的定律和現象，或預測事情應該如何如何。

不管是定律、原理、假說、模型與理論，由這些構成的知識體系，可稱之為「科學」。這樣建構的知識體系，在本質上具有一些特性：

1. 它的成立依憑外在事實的支持。也就是說，它要能夠接受客觀事實的考驗，或詮釋各種現象。
2. 知識體系內具有邏輯協調一致性，不得相互矛盾，或有困呆不相連局的情形。
3. 精神上它是開放的。因為現象的觀察是無止境的，目前建構的科學是否真的內部已經邏輯協調一致性，則不僅是不斷地在檢核，而它所預測的現象是否發生，或是有了新發現的現象，該理論是否尚能完滿解釋它，一連串這些考驗，使科學更趨堅實，或使科學概念再度更新，這兩種都有可能。所以科學的知識可以說是對於事物迄至目前最佳的詮釋。

基於以上所提的特性，科學可說是經過切實考驗的知識，故可信度極高且最切合實際，而同時卻又具有開放的特性。

我們再回過頭來看看獲得科學知識的認知過程；有對現象的「觀察」，它涉及到觀察什麼？怎麼觀察？而要知道觀察什麼及如何觀察，則觀察者心中一定事前要有一些先備知識或探求的目標。再者，觀察所獲得的訊息，經過整理與分析，依據這些資料當論據，提出一些統整性的概念、規則性等，或是依據心中預設的模型、理論，去做邏輯性的推論，來解釋或證明發生的現象和成立的定律，來預測可能發生的現象等。這些依據外在事實當論據，來建構知識的思考模式，我們稱之為「科學方法」。我們知道，像音樂、藝術或宗教等的知識，並非經此思考模式而得的，故不屬於科學的知識。

在從事科學的探討時，為了所獲得的資訊要正確，觀察及執行時必須切實而細心，為了所獲得的結論要經得起客觀事實的考驗，歸納及推理時要篤實誠

實，不得馬虎作假，以免落於虛妄。因而如此切實細心做事，就事論事的態度，可稱之為科學的精神。

我們可以把探求科學時，所培養的科學精神，以及探求科學時所運用的觀察、整理、分析、歸納、推理等心智活動能力(可稱為科學方法的運用能力)，和獲得的科學知識，統稱為科學素養。我們稟承科學的態度和精神，運用科學方法去探求科學知識，而獲得認知上的成長以及能力的培進。科學教育的目標在此，而我們的科學教學也應依此科學的過程來進行(陳文典，92)。

(二)、科學概念的認知

科學探索滋長發展了我們的智識。它包括科學的名詞、事實、概念、原理、定律、理論、模型及各式應用等方面的知識。

1. 科學的語彙(scientific term)：

科學家用以稱呼存在的實體、物體、事件、週期、各分類部門、組織、結構等。語彙於傳達時被引用，係一種定義，不包括相關、定律、模型及理論。例如魚、人、書本、中國、風、等。

2. 科學的事實(scientific fact)：

透過操作型定義的方法和程序，對自然現象、事件或物質性質的觀察與度量，產生一個可反覆驗證的、確定的結果，成立一個科學的事實。例如真空中的光速、一大氣壓下水的沸點、東西的長度等。

3. 科學的概念(scientific concept)：

經由不同的發生方式及途徑，被觀察或偵測具恆常性的自然現象、性質或事物特性，可用語詞或符號來表徵的稱為科學的「概念」。例如在定容積下，壓力與溫度成正比，例如物體的熱脹冷縮等。許多概念又是由一些基本的概念建構起來，例如描述概念間的關係也是一種概念。基本的概念不多，像長度、質量、電荷、時間等。許多概念由關係式來界定或方程式來陳述，例如密度(質量/體積)、波義耳定律等，它常在某特定條件下才成立，瞭解這些成立的要件也是另一概念。

4. 科學的原理(scientific principle)：

對系列的觀察、度量所呈現的規則，作一般化的、統整式的陳述或數學表

示，科學上的原理雖是一種一般化的陳述，但在應用上則尚未達到定律的水準，例如巴斯卡原理、阿基米德原理。

5. 經驗定律(empirical law)：

經驗定律是經由觀察與度量，建立兩個或兩個以上概念之間的相關關係。這些定律是應用或建構理論、模型的基石，例如斯涅耳折射定律、刻卜勒定律、歐姆定律。

6. 科學的理論(scientific theory)：

經驗定律是經由觀察與度量所得的經驗以及原理和經驗定律，創造性地提出一個統攝性的想法。此一想法可能藉助語文或數學來陳述。此一想法，可普遍地解釋這些現象，並且能預測系列相關現象，而獲得證實。有時基於相同的論據而有不同的理論，科學家往往傾向於接受那最簡單、最容易理解、最具預測能力，是能符合經驗定律的理論，例如原子論、分子動力論、自然選擇理論、板塊理論、量子論。

7. 科學的模型(science model)：

一種代表，可見的、或數學或語文描述的，用以幫助描述或瞭解科學的現象、理論、定律、組織、結構等。例如波動模型、粒子模型、電流模型、地球大氣的溫室模型等。

8. 遍通性的定律(universal law)：

歷經考驗仍無法否定它，而且解釋現象的涵蓋面很大的定律，它們是科學的基石，如萬有引力定律、能量守恆定律、動量守恆定律、電荷守恆定律等。

9. 科學應用(application of science)：

運用經驗定律、理論的預測和推論，設計特殊的環境使現象得以發生，即是一種人為的情境，例如發電機、電動機、電燈、電腦等。經過科學知識的發展，基本的定律之瞭解，運用這些理論及定律，人們研發科技，做各種巧妙的發明（陳文典，92）。

（三）、科學智能

「科學探索」在於致力於對現象做不斷地檢視，以及對時下適用的解釋是否

能充分而適當地詮釋這些現象做嚴謹地審思。對於科學家所提的研判不應武斷地設定它僅是一種試驗性暫時的理論。權威不能決定或設立科學的知識，科學乃是科學家源自自然真實的一種陳述。

這些「科學探索」的倫理可說是立根於它的運思過程中。這些過程：觀察、分類、度量、解釋資料、推斷、傳達、控制變因、模型及理論、形成假說、預測等，它不一定也不必要依序發生，但是每一過程都可有不同的層次和不同的分歧發展，這使得科學的運思過程深具有機性，靈活及切合情境地做有效演進。

1. 觀察(observing)：

經由直接感官或藉助儀器，有目的的、專注的、多向性的、切近的客觀檢視，察覺事象的特徵並加記錄、詮釋，而非隨意泛泛的看過。

2. 分類(classifying)：

依某屬性、某特徵或某功能，將事物作系統地分類。

3. 度量(measuring)：

界定觀測量，運用儀器對物體、系統或現象做量化度量，以便顯示其特性。

4. 解釋資料(interpreting data)：

運用可理解的、熟悉的語言與符號，闡明由觀察所獲得的資料，詮釋資料的涵義和特性。

5. 推斷(inferring)：

由觀察、度量、分類所得之事實與資料，做有依據的、邏輯的、歸納的提出結論。

6. 傳達(communicating)：

運用文字、插圖、表格、符號等有效的方式，將觀察、度量以及推斷所得，用有效的表示方法傳譯給其他人。

7. 控制變因(controlling variables)：

在諸多相關變因中，固定其他，只允一個變因改變，觀察它與依變項之間的因果及效應關係（陳文典，92年）。

第四節 IEA 的背景及組織運作

IEA (全名是 The International Association for the Evaluation of Educational Achievement, 中文叫做國際教育學習成就評鑑協會) 是作國際教育成就的一個組織, 本論文所引述的 TIMSS-R 研究報告正是此機構的公開發表研究。此機構多年致力於用科學方法來評鑑各國教育之優缺點, 很快的獲得國際間比較教育界的認可與支持 (戴曉霞, 1995)。以下要介紹 IEA 的背景、組織發展、運作及經費運用及至今的重要研究報告:

一、探討 IEA 的成立背景

IEA 的總部設在荷蘭 阿姆斯特丹。IEA 是在 1959 年由歐 美學者推動成立。IEA 最早在 1961 年 4 月首度出現在國際教科文組織 (UNESCO) 所屬的教育研究所的議程中, 有鑒於當時的比較教育仍偏重於描述性的研究, 這些歐、美學者覺得十分需要對教育系統共有之間的問題進行實證性的量化比較研究, 於是在 1964 年開始在英、法、美、日等十二個國家或地區以相同的測驗題目與問卷, 進行第一次國際數學成就調查 (The First International Mathematics Study, 簡寫為 FIMS), 經過 1960、1961 年多次會議之後, 在 1961 年正式命為「國際教育成就評鑑協會, 簡寫為 IEA」, 仍屬於 UNESCO 之下, 直到 1967 年才在比利時登記為法人機構 (戴曉霞, 1995)。

誠如曾經擔任 IEA 主席長達 17 年的 Torsten Husén 所說的, IEA 的成立與運作「不但代表各國緊密合作, 以建立經驗的比較教育學之努力; 也顯示學術研究反應時代精神 (Zeitgeist) 及盛行的研究典範之程度」(Husén, 1979, p.371)。IEA 的興起除了與當時比較教育的發展與需求有關, 還和二十世紀中葉的國際關係及社會經濟脈動有密不可分的關係 (戴曉霞, 1995)。

為了充分反映 1950、1960 年代的「時代精神」, IEA 自 1960 年代初期成立至今 40 餘年以來, 已完成的九次大規模的跨國研究 (最近一次是 2003 年), 這些年的研究都遵循一種簡單的「三段式」研究架構, 而不敢有所逾越。IEA 研究的架構包括下列三階段: 一、發展對世界各國都有效的研究工具, 以測量各國學生在主要學科領域之能力; 二、測量各國教育體系之輸入 (inputs), 包括經費、

教師、教材、教法等；三、測量各國教育體系之輸出 (outputs)，主要為學生之學習成就及態度，並進行輸入變項與輸出變項之相關研究研究(Husén ，1979 ， p.380) 在這樣的研究架構之下，IEA 各項研究所採取的是嚴格的量化研究典範，以科學方法來保證研究結果的有效性。

影響 IEA 最大的首推 Torsten Husén 。他從 IEA 的醞釀期就積極參與，並擔任 IEA 的主席達十七年 (1962-1978)。IEA 是一個沒有固定經費的非官方組織，能夠生存至今，他的領導能力功不可沒 (Postlethwaite ， 1986)。但是不可否認的，他個人的學術背景及經歷也主導了 IEA 的研究取向，在肯定與認同之外，也招致不少責難與批評。 Husén 在他的自傳中曾以專章記述他在 1941 到 1947 年之間，協助瑞典軍方進行心理測驗的經歷，而這段經歷對他日後學術研究去向顯然有相當深遠的影響，例如 Brian Holmes 即將他歸類為比較教育心理測驗派的代表 (戴曉霞 ， 1995)。

IEA 這些年來所作的跨國研究，除了 1980 年代「教室環境研究」兼採非參與觀察法之外，基本上都是採用調查研究法，並且將「教育成就評鑑」等同於「學業成就測驗」，充分地符應了心理測量客觀化、系統化、計量化、和標準化等的特徵。這種研究取向或許符合 1960 年代對比較教育的呼籲，但也窄化、簡化了教育成就評鑑的範圍和層次，限制了 IEA 研究的貢獻。

不過因為 IEA 的研究提供大量的資料，可以對參與國家的教育過程作很詳細的瞭解，這種循環式連續性的研究，可以讓每個國家去測量自己在數學、科學及閱讀能力上的教育成就；循環式的調查研究也可監控教育政策的執行成果，並可定義當代新的議題以作為改革努力的目標 (IEA ， n.d)。

二、IEA 的組織發展史

成立初期：因參加的國家有限，所以組織很單純。由各會員國組成的理事會 (Council for the International Evaluation of Educational Achievement) 是主要的決策單位，舉凡比較研究的學科、研究對象、研究進行的時間表、及經費的安排等都由其決定。理事會原則上一年召開一次，其下另設常務委員 (standing committee)，除執行理事會的決議之外，在理事會休會期間，代行必要之決策。

1964 年後：IEA 開始規劃包括科學、文學；公民等學科之「六科調查研究」(Subject Survey) 之後，開始有所改變。各會員國決定在各國增設「研究召集人」(national research coordinators, 簡稱 NRCs) 除統籌各會員國之國內會務之外，並擔任各會員國間之協調。此外，另設立「稽查委員會」(steering committee)，以審查日益增多的研究計畫。從此，IEA 的組織日益複雜。

1980 年代中期：由約 40 個會員國組成的大會 (general assembly) 是最高決策機構，其下設有常務委員會，之下再設立二類委員會：一為與研究計畫之審核、執行相關的各委員會；二為主管經費、未來活動、編輯等工作委員會 (如圖 2-4.1)。

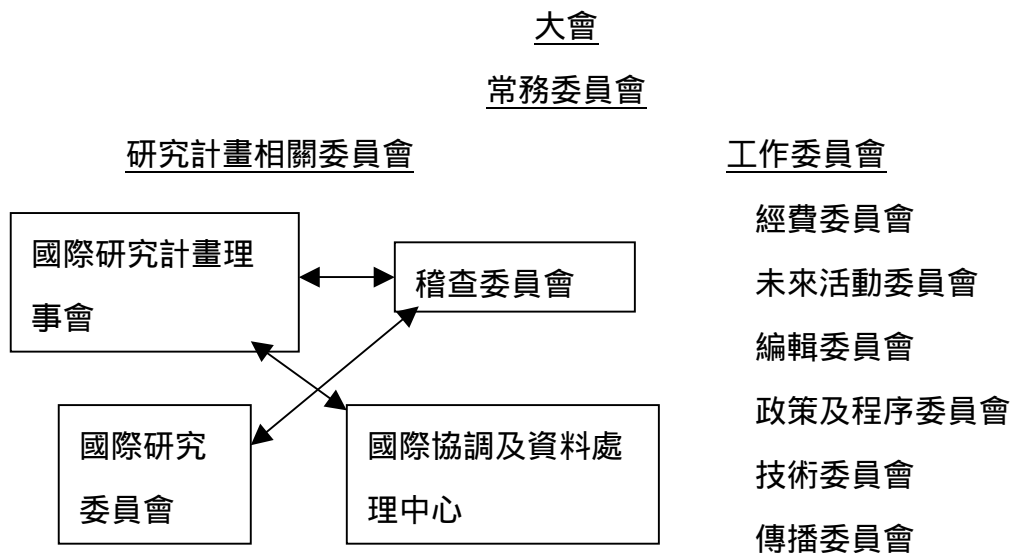


圖 2-4.1：IEA 組織結構圖 (Postlethwaite,1986,pp.129-130)

現在：IEA 仍是登記在比利時法人機構，享有比利時提供的捐款及募款得以免稅的身分。IEA 的執行秘書處是由委員會中的成員所組成，設在荷蘭的阿姆斯特丹，負責的工作為財務的管理和人員的雇用。IEA 的其他合作單位包括：IEA 的資料處理中心 (data Processing center, 簡稱 DPC); 加拿大國家統計局 (statistics canada), 還有教育測驗服務局 (educational testing service) 及部分知名大學協助等。例如 1998 年開始的 TIMSS-R 調查研究是由美國 Boston College 總主持，取樣工作由

加拿大國家統計局統籌，數據處理為德國漢堡大學負責，試題本籍問卷翻譯之核可由位在荷蘭阿姆斯特丹 IEA 總部辦理，就是由不少其他單位協助完成（陳昭地、沈青嵩，1998）。

三、IEA 的組織運作及經費運用

所有向 IEA 提出的研究計畫，先送交稽查委員會，就其價值、可行性及經費等方面加以評估。其次再送給由各國研究召集人（NRCs）組成之「國際研究委員會」（International Study Committee）特別就研究之概念，包括研究所涵蓋的變向及研究之模式等仔細檢討。研究計畫若能通過上述各委員會之審查即可展開資料蒐集等研究步驟。同時並由決定參與該項研究之會員國指派代表組成「國際研究計畫理事會」（International Project Council，簡稱 IPC），其任務是預算及時間之雙重限制之下，維護研究之品質。「國際協調及資料處理中心」（International Coordinating and Data Processing Center）負責確保所有的研究相關事宜都能按時進行。

IEA 各工作委員會屬任務編制，需要時由大會指定成立，工作完成後就解散。以經費而言，大致分為會員國和國際支出二項。和各國相關的活動，原則上由各會員國自行負擔。和國際性活動相關的經費，例如國際協調及資料處理中心、稽核委員會等支出，則由經費委會負責籌募。IEA 的經費來源有兩種特色：第一、沒有固定的、來自官方的收入；第二、IEA 不收會員國會費，以免有些會員國會因為無法支付會費，而造成會員國之間的差別地位，或妨礙各國參加研究活動的均等機會。以 1985 年為例，當時正在進行的四個研究計畫，每年共需要一百萬美元的國際性支出，若由會員國平均分攤，則每個會員國每年需支付二萬五千美元會費。並不是每個會員國都能負擔這筆費用（Postlethwaite，1986，p.135-136）。IEA 經費的來源主要是向歐美各國政府及私人基金會募款。這種募款方式使得 IEA 的研究經費並不穩定，常常影響研究的規劃與進行（Husén，1979，p.377）。但我國在 1992 年加入時，當時對於參加國家就有修正要繳交年費，但對於部份較貧窮的國家是得以申請免繳優惠。我國首度參加 TIMSS-R 的國際調查計畫是在 1998 年二月由台師大科教中心專家向國科會提出為期二年半的調查研究計畫，申請總經費約是新台幣七百五十萬元。（陳昭地、沈青嵩，1998）

除了研究結果的發表之外，IEA 對各國教育研究人員的訓練和研究能力的提昇也展現了相當的熱忱。IEA 不但協助技術不足的會員國進行研究，以提供其研究人員在職訓練的機會外，也舉辦各種和教育有關的研討會。例如在 1968 到 1972 年之間，IEA 以英語、法語、德語等為世界各地的研究者舉辦多場有關教育研究方法的研討會，也曾在 1971 年召開以課程發展與革新為主題的國際性研討會（Postlethwaite，1988，p.26-46）。

四、IEA 的重要研究報告

IEA 成立至今共作過二十多項國際研究，其中以完成的調查包括 1964 年的第一次國際數學研究（FIMS），1970 年到 1971 年的六科研究（FISS），1980-1982 的第二次國際科學研究及教室環境研究（SIMS），1980 年代末到 1990 年代的電腦在教育之應用調查（computer in education）、寫作能力研究（written composition）、閱讀素養調查（reading literacy）及教育中的語言研究（LES），還有 1995-1996 年的第三次國際數學、科學教育研究（TIMSS）、1999 年第三次國際數學、科學教育研究後續調查（TIMSS-R）（Martin *et al.*,2000），及最近才進行完成的六個調查研究，包括 1994-2004 的公民教育研究（CIVED）、1999-2003 年國際閱讀素養進階研究（PIRLS）、1986-2002 年第二次教育資訊科技教育第二模組（SITES-M2）、2000-2004 年國際數學、科學教育趨勢調查（TIMSS-2003）、1998-2002 年第三次國際數學、科學教育研究後續調查錄影帶計畫（TIMSS-R Viedo）等（Martin *et al.*,2000）。

綜合以上的分析可以發現，IEA 的成立除了反映 1950、1960 年代比較教育研究典範之轉換及建立比較教育科學之呼聲外，和當時的國際關係及各國經濟、社會現代化的浪潮有相當密切的關係。作為一個沒有固定財源、不時面對政治干預的國際組織，IEA 能夠展現其學術活動力長達 40 餘年，殊為不易，對比較教育研究風氣與水準之提昇，亦可為貢獻卓著（戴曉霞，1995）。

第五節 TIMSS-R 國際科學成就評量架構

本節將以三個方面來加以探討 TIMSS-R 的成就評量架構：首先先瞭解 TIMSS 和 TIMSS-R 的成立背景、研究目的；再深入探討其研究工具、評量概念架構和科學課程的設計架構；最後再探討其特殊的資料處理模式。

一、TIMSS 和 TIMSS-R 成立背景及研究目的

IEA 自 1994 年推動第三次國際數學與科學教育成就研究 (the Third International Mathematics and Science Study , 簡稱 TIMSS) , 是所有執行過的國際教育研究中研究範圍最廣的。學生的學習成就差異 , 可當成每個教育系統的差異 , 也可看作不同的經濟、狀況及在全球競爭下國家相關位置的指標。TIMSS 的研究模式是跨不同教育系統及以多量的變因數目研究學生學習的過程 , 使我們更了解所要研究的主題 (洪佳慧 , 2000) 。

TIMSS 是一個經由嚴謹的結構 , 作為長期性對於數學和科學教育循環研究的計畫 , 未來每四年舉辦一次。TIMSS 在 1995-1996 間舉行 , 有 42 個國家參加。藉由問卷、錄影帶、教材的分析來比較 42 個國家、三個年級在數學及科學上的成就。第一次 TIMSS 的結果報告是在 1996 年時公佈。當時即引起全球廣泛的討論及課程改革 , 國際間的學者也爭相出書發表其觀感及發現 , 而參與國對於自己國家的結果報告也紛紛出爐 (Martin *et al.*, 2000) 。

將國家作排名並非 TIMSS 的主要目的 , 排名的主要目的是將成功和教學方式拿來作為其他國家的參考。因為過去 IEA 所作的研究對於數學和科學都有重大的發現 , 於是國際間各國課以利用 TIMSS 的研究結果來改善在學校的教學及學生的學習 (Robitaille , 1993) 。

TIMSS 會選擇數學和科學這兩個主題 , 是因為世界上很多國家認為國民的科學素養和經濟生產力是十分相關的 , 而數學是現代科技中的通用語言。數學和科學是每個學校課程中的重要成分而國際上的很多國家也正面臨對於該國數學與科學課程的重要改革 , 許多國家將會參考 TIMSS 的成果來引領教育改革計畫 , 因為大多數的國家都有一個共識 , 那就是要提供該國的學生一個是世界級的課程 (world class) , 為進入二十一世紀作準備 (Robitaille , 1993) 。

TIMSS-R (the Third International Mathematics and Science Study - Repeat) 是指在 1999 年所作的第三次國際數學、科學教育研究後續調查。因為在 1995 年舉辦 TIMSS 之後，許多國家紛紛表示希望能有後續的調查，因為不僅可以取得有關學生在數學和科學科目中的是知識和能力的資訊，以及有關文化環境、教學實施狀況、課程目標和教育上的安排等伴隨學生成就的資訊外，最重要的還是了解當年在 TIMSS 研究中的九歲群的學生，經過四年升上八年級（十三歲群）之後的學習成就，更可比較這兩群相隔四年八年級學生的學習成就是否有差異（羅珮華，2000）。

我國是在 1988 年 2 月 1 日正式執行並參加此國際教育成就調查計畫，配合國際測驗時程，於 1998 年 4 月 28 日舉行試測，抽測 25 校 30 班 1205 位學生；1999 年 5 月 17-28 日舉行正式施測，抽測 150 校，每校一班，共 5889 位學生（陳昭地、沈青嵩，1998）。

二、TIMSS-R 的研究工具、評量概念架構及科學課程的設計架構

（一）研究工具：包括問卷和科學成就測驗試題兩種。

1. 問卷：

共有三種：包括學生問卷、教師問卷及學校問卷。每個參與的學生都要回答學生問卷；施測學生班級的科學教師及數學教師都要填寫教師問卷；施測學生學校的校長要填寫學校問卷。此問卷可用來瞭解不同的教學安排與學生成就相關的訊息。

2. 科學成就測驗試題：

包括數學和科學方面共有七個領域分類，即數學、地球科學、生命科學、物理、化學、環境與資源議題、科學探究與科學本質（Garden & Smith，2000）。

TIMSS 的測驗品質也和抽樣的品質一樣受到相當地監控，各國的 NCR 必須一起合作並全程參與 TIMSS 測驗的發展。經過一連串的努力，除了各國提交經過科學學科專家檢視的題目之外，另外還加入一些題目以能夠包括所有的科學主題，這些題目首先需經過試測與檢視試測的結果，最後加入新的題目並再次進行試測。

TIMSS的測驗結束後，公佈大約三分之二的試題，剩下的三分之一試題則為了能夠使用於下次測驗以便測量趨勢，因此保密。在TIMSS-R的測驗中這些公佈的試題被新的試題所取代，新的試題則是根據已釋出試題之內容、形式與難度重新編寫而來。科學與數學試題更換委員會(Science and Mathematics Item Replacement Committee)，由參與國所提名的國際知名數學與科學教育家組成，他們的工作就是對測驗的學科議題提出建議，一共有超過300題的數學與科學題目被發展出來作為可能的替換試題。新的試題同樣需要經過科學學科專家檢視以及試測，因此幾乎所有參與TIMSS-R的國家都參與了新試題的試測，而且所有的國家聯絡人也有檢視試題與評分標準的機會。試測與檢視的結果，數學有 114 題，科學有98題被選為替換試題納入TIMSS-R的測驗之中(Martin, Mullis Gonzales, Gregory, Smith, Chrostowski, Garden & O'Connor ,2000 ; Gonzalez&Miles, 2001)。

TIMSS-R 科學試題共有 146 題，呈現一系列的科學主題與技能。有四分之一的試題屬於簡答題(short answer)與申論題(extended response)題型，需要學生產生並寫下自己的答案，其餘的試題都是選擇題題型。表 2-5.1 呈現各科學內容的題數以及總分，由於有些需要學生較長時間來組織回答的問題採部分給分，全對給兩分，對一部分給一分，而其它試題給分都是一分，所以科學部分試題的總分會越過試題數。

表 2-5.1 各科學內容的題數以及總分 (Gonzalez & Miles , 2001)

內容	題數	分數
地球科學	22	23
生命科學	40	42
物理	39	39
化學	20	22
環境與資源議題	13	14
科學探就與科學本質	12	13
合計	146	153

(二) 評量概念架構

TIMSS-R 的概念架構基本上是延續 TIMSS 的研究架構，以課程為焦點來瞭解組成教育環境的種種因素。TIMSS 將課程分為三個不同的層次去解讀：社會想要的教學（意圖的課程）、實際進行的教學（實施的課程）以及學生學到的課程（完成的課程），此三個層次說明學生成就的主要影響因素（Robitaille, 1993；Martin, 1996；Gonzalez & Miles, 2001）。如圖 2-3.1，TIMSS 的概念模式把影響教育的因素視為處於一系列鑲嵌的情境背景中，這些情境從最「總體的（global）」到最「個人的」狹窄的情境鑲嵌於較廣的情境背景中，它們也會受到較廣情境的影響，但是它們不只是較廣情境的從屬部分而已（Robitaille, 1993）。

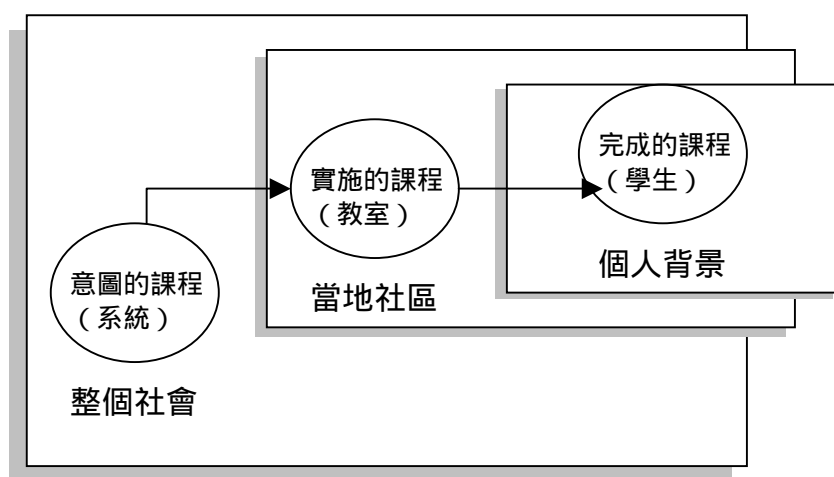


圖 2-5.1：TIMSS 的概念架構（譯自 Robitaille, 1993, p.26）

1. 意圖的課程：「學生預期要學到什麼？」

意圖的課程是在國家或教育系上階層所界定的數學與科學內容，存在於教科書、課程綱要、考試內容以及用來指導教育體系的政策、規章與其他官方聲明之中，學生被預期要學習的課程可能以數學與科學概念、過程與態度的形式來描述。意圖的課程處於一個特定的教育背景，此背景包含了對於教育系統所作的制度安排(institutional arrangements)，在系統階層所作的教育決策會被下決策的結構所影響，這些決策則包括了學校組織的模式、教學的分派以及財政與人力資源的配置。意圖的課程也位於一個較大的社會背景中，社會的情境包含那些影響系統層級制度安排與課程內容的因素(Robitaille,1993)，它說明了社會對教與學的目的

標，這些目標反映出社會的想法與傳統，並且被教育體系的資源所限制。此課程的研究是 TIMSS 計畫初期的主要部分，課程分析包括課程綱要與教科書的內容以及問卷，其目標在詳細地描繪出參與國的課程目的(Martin,1996；Gonzalez & Miles,2001)-

2. 實施的課程：「誰提供教學?教學如何被組織?」

實施的課程所指為經由教師詮釋並將其傳達給學生的數學與科學內容，雖然會受到意圖課程的引導，但是並非和意圖的課程一模一樣，它也可以用概念、過程以及態度來描述。實際在課堂進行的教學活動大部分通常由教師決定，教師的行為極可能受到本身所受教育、訓練與經驗的影響，也會受到學校的性質與組織架構、與教師同儕之間的互動以及學生主體的組成所影響(Gonzalez & Miles，2001)。實施的課程所處的教育背景是由在學校與班級階層作的制度安排所構成，它們會受到系統階層的制度安排所影響，而且在某個程度上來說它們取決於系統階層的安排;此層次的制度安排包括教學工作、班級經營、資源的使用、教師態度以及教師的背景。此課程處於一個較廣的地方社區背景之中，在某些方面它反映了整個社會，從另一些方面來說，在同一個教育權限範疇內的各個地方社區彼此會有很大的不同(Robitaille,1993)。

3. 完成的課程：「學生學到了什麼?」

完成的課程由學校教育的成果所構成，為學生經過求學過程所獲得之數學與科學的概念、過程與態度，學生所學會受到他/她們被期望的學習所影響，也會受到可得機會的品質與種類所影響。完成的課程處於一種個體學生的教育情境，由學生決定的制度安排所構成，包括學生作了多少作業、學生的努力以及學生在教室的行為模式等等，這些因素受到系統與教室階層的安排大大地影響。此課程所處的情境為學生的個人背景，然而學生的背景並不一定都相同，學生的成就有部分與實施的課程以及社會、教育的背景有關，有大部分則是與個體學生的特質有關，這些特質包括能力、態度、興趣及努力（Robitaille，1993；Gonzalez & Miles，2001）。

關於內容、制度安排以及社會情境背景的邊界並不是那麼地卻明確，重要的是在三個層次內容中的變數需要在三個不同制度安排階層中考慮，並且在三個

不同的社會情境背景之下討論 (Robitaille, 1993)。

(三) 科學課程的設計

TIMSS 為數學與科學所發展的課程架構是其研究的第一步，此架構作為 TIMSS 設計成就測驗部分的指引以及課程分析部分的基礎，呈現數學與科學課程的結構，是一種能夠比較與對照各國課程的有力工具 (Robitaille, 1993)。

TIMSS-R 沿用 TIMSS 的課程架構，包括三個向度：「內容 (content)」向度表示學校數學與科學的學科內容；「預期表現 (performance expectations)」向度描述許多預期的學生表現與行為；「觀點 (perspectives)」向度則是把焦點放在學生對於科的態度、興趣與動機之發展 (張一誠, 2002)。

每個向度都分成數個類別，每個類別又分成幾個子類別，視需要可再往下細分，圖 2-3.2 為 TIMSS 課程架構的主要類別。不過因為此架構是為了從學生開始接受教育一直到中學教育完成的課程而發展的，所以並非所有的面向都能表現於 TIMSS 的八年級測驗之中 (Robitaille, 1993 ; Gonzalez & Miles, 2001)。

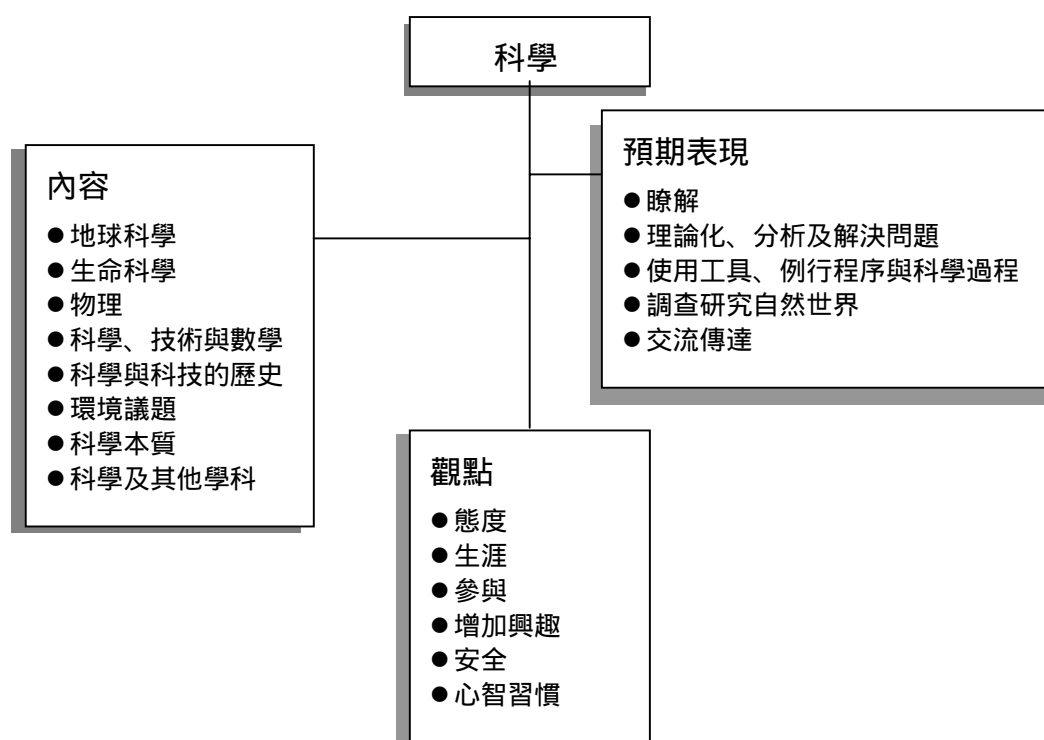


圖 2-5.2 TIMSS 課程架構的主要類別

(譯自 Robitaille, 1993.p.46 ; Martin, 1996.p.1-7)

TIMSS-R 的科學測驗包含了六個內容範圍：

1. 地球科學：包括地球的特徵、地球過程以及宇宙中的地球。
2. 生命科學：包括生物的多樣性、組織與構造；生命過程與維生系統；生命螺旋 遺傳的連續性與多樣性；生物的互動；以及人類的生活規律與健康。
3. 物理：包括物理的屬性與變化、能量與物理的過程以及力學與運動。
4. 化學：包括物質的分類與結構、化學的屬性以及化學的變化。
5. 環境與資源議題：包括污染問題；土地、水與海洋資源的保護；物質與能量資源的保護；以及自然災害的效應。
6. 科學探究與科學本質：包括科學知識的本質；科學事業；科學、技學、數學與社會的互動；以及進行科學研究所使用的工具、程序與過程。

三、資料處理模式

非選擇題部分的評分方法為使用2-數字(2-digits)的編碼，並且每一試題都有自己的評分規準(rubrics)。第一個數字代表回答的正確程度，有些題目答對給一分，有些題目則是全對給兩分，部分答對給一分；第二個數字代表得分或沒有得分的理由，紀錄學生的解題方法、策略或常見錯誤、迷思概念，這一部分可以提供資訊幫助學生更了解數學與科學的概念以及解決問題的方法。

為了確定資料的品質與可比較性以供分析，TIMSS 以嚴格的品質管制步驟成立了國際資料庫(International Database)，並準備手冊與軟體讓參與國輸入他們的資料，因此這些資料才能以國際標準的形式送到 IEA 的資料處理中心納入國際資料庫。資料送達資料處理中心時，必須經過整理：鑑定、紀錄並修正由國際工具、檔案的結構以及編碼系統所產生的誤差；這些過程還包括強調國家內資料的一致性，以及許多學生、教師與學校其資料檔案之間的適切連結。整個過程經由 IEA 資料處理中心、ISC (the International Science Center) 以及各國國家中心檢查，國家中心許多機會可以檢視自己國家的資料，而 ISC 則結合 IEA 資料處理中心檢視每個國家中每一認知試題的統計資料以確認表現不佳的試題。

TIMSS 成就資料的報告主要是以「試題反應理論(Item Response Theory，簡稱 IRT)的計分方法為基礎；以對或錯來給分的試題使用的是 2-參數(2-parameter)

與 3-參數(3-parameter)IRT 模式，而 2-參數將部分得分模式(partial credit models)一般說來用於得分可以超過一分的試題。IRT 計分法藉由平均每個學生對於題本內試題的回答來產生分數，每一試題的難度與鑑別率都有考慮在內。TIMSS 使用的方法使得即使每個學生分配到的題本只有包含全部試題中的一小部份，也能夠產生可靠的分數。因為每個學生所分配到的題本是入本中一本，作答的試題也不盡相同，所以 IRT 方法適合用在此研究以方便比較所有學生的成就 (Gonzalez & Miles,2001)。

第六節 TIMSS-R 國際科學成就報告內容

TIMSS-R 國際成就報告內容大致分成五部分：第一部份當然是學生的科學測驗題本的量尺分數統計分析，比較各國學生的科學成就結果；第二部分是學生問卷結果分析，研究學生家庭背景環境和科學態度等因子的影響分析報告；第三部分各國教育政策的分析，比較各國學校的科學課程安排有何程度的影響；第四部份是教師問卷部分，包括教師年齡、性別、課前作何準備等因子的分析探討；第五部分是學校問卷部分有關教學設備的供應與師資安排、教學進度等因子的影響分析 (Martin, Mullis, Greogry, Chrostowski, Garden & O'Connor, 2000)。

茲將就前三部份做詳細的分析和研究：

一、學生科學測驗題本的量尺分數統計結果分析：

1. 我國和新加坡的測驗平均成績表現最好，緊接著的是匈牙利、日本以及韓國。其他表現良好的國家還有荷蘭、澳大利亞、捷克以及英國；低成就的國家包括菲律賓、摩洛哥以及南非。我國學生的科學測驗排名第一，平均分數是 569 分；南非是最後一名，平均分數是 243 分；而國際的測驗平均分數是 488 分。有 19 個國家的科學成就成績在測驗平均分數之上，包括首度參加測驗的台灣和芬蘭兩國；13 個國家的科學成就成績在測驗平均分數之下，包括首度參加測驗的摩爾達維亞、馬其頓共和國、約旦、印尼、土耳其、突尼西亞、智利、菲律賓和摩洛哥等九個國家。
2. 在大部分國家的中間段學生 (the middle group)，意指測驗成績百分等級的 25 % 到 75 % 的這 50 % 學生，高、低分者的測驗分數相距約 100 ~ 150 分的差距；相對的，測驗成績百分等級 5 % 到 95 % 的這 90 % 的學生高、低分的測驗分數差距卻拉大到 250 ~ 300 分的差距。此次參加測驗時的學生平均年齡從 13.8 歲 (塞浦路斯、芬蘭) 到 15.5 歲之間 (南非)。
3. TIMSS 的研究分析將學生的成就表現分成四種層級，各代表學生在科學成就上有不同的學習成果：

- ◇ 第一級 (the top 10 % benchmark) : 指在 TIMSS-R 的科學測驗成就表現百分位九十以上的學生，也就是測驗成績在 616 分以上的學生。這些學生懂得一些複雜和抽象的科學概念，在六大科學測驗的每一個領域中。
- ◇ 第二級 (the upper quarter benchmark) : 指在 TIMSS-R 的科學測驗成就表現百分位七十五以上的學生，也就是測驗成績在 558 分以上的學生。這些學生能證明科學的循環、系統和定理上的一些概念知識。
- ◇ 第三級 (the median benchmark) : 指在 TIMSS-R 的科學測驗成就表現百分位五十以上的學生，也就是測驗成績在 488 分以上的學生。這些學生能認識和了解不同學習主題範圍的基本科學訊息。
- ◇ 第四級 (the lower quarter benchmark) : 指在 TIMSS-R 的科學測驗成就表現百分位二十五以上的學生，也就是測驗成績在 410 分以上的學生。這些學生懂得基本的科學事實在地球、生命和物理方面。

研究這些學生的背景問卷資料作一系列的了解，本論文也是依照這些指標性的測驗分數作分級分析研究。在本論文的第四章第二節中之(表 2-4.2) 為各國在不同成就層級的學生分佈比例。

4. 在這次測驗中 38 個國家中有 16 個國家，男生的測驗平均分數明顯的高於女生的測驗平均分數；而且男學生在物理、地球科學、化學和環境與資源議題上的科學成就成績明顯表現較好，尤其在高分層級 (top quarter) 的學生表現中，這種性別的差異更形顯著：在測驗成績中男生有 29 % 達到前四分之一的，女生則只有 21 % 達到前四分之一；而這種現象在香港、斯洛維尼亞和伊朗等國家或地區較不明顯。在本論文的第四章第二節中之(表 2-4.3) 為各國男、女生的科學成就表現。

二、學生問卷結果分析：

在背景問卷分析上，將學生依科學成就測驗成績分成四級：

- (1) 百分等級 90 --the top 10 %
- (2) 百分等級 75 —top quarter

(3) 百分等級 50 --top half

(4) 百分等級 25 – lower quarter

做各種不同的背景因素分析：

1. 雖然各國的家庭教育資源皆不大相同，但是擁有較多的教育資源的學生在國際測驗成就的分數上表現仍高於缺乏家庭教育資源的學生。這裡所指的家庭教育資源的指標因子項目有：家中藏書、三種讀書工具(電腦、個人書桌及字典工具書)和家長學歷。所謂的有較多教育資源是指家中藏書超過 100 本、擁有三種讀書工具及父、母至少有一人有大學畢業學歷；缺乏家庭教育資源是指家中書本少於 25 本、沒有讀書工具及父、母的學歷較低。
2. 調查發現：發現國二的學生幾乎對自己未來的教育程度期望都很高。超過一半的學生都希望能完成大學學歷，且測驗分數高者期望越高。
3. 一般來說，國二的男生比女生在科學方面較有興趣，特別是在自然分科教學的國家中較明顯。在這些國家中女生較喜歡的自然科目是生物，而男生則較喜歡物理，較不喜歡地球科學和化學。
4. 雖然在自然合科教學下的學生普遍比分科教學下的學生對科學有較正面的態度，但這些國家的學生仍較喜歡生物、地球科學領域的單元，較不喜歡物理、化學領域的單元。且男生還是比女較喜歡科學，特別是在物理、化學和地球科學領域。女生則是在生物領域中有較高的興趣。

三、各國教育政策的分析：

1. 38 個國家中有 35 個國家在全國的課程發展目標中有特別明定學生的科學課程目標，但澳大利亞、加拿大及美國三個國家除外。有二十一國家是將科學領域合科教學，其他國家則是分科教學。
2. 在各國科學教育中，測驗和評量仍廣泛的在課程實施中使用。只有比利時(法語區)和台灣兩個國家沒有使用官方的科學測驗來證實或篩選學生是否有能力上大學或是學院。一般說來，三分之二的國家有作二至三級的科學分級測驗以確定學生有無能力學習更進階的課程。
3. 一般的國家，科學的授課時間比例從小學四年級的 11 % 到國二的 16 % ，

都有增加的趨勢；反之，數學的授課時間比例卻是減少。

4. 在大部分的國家中，課程設計的重點是熟悉基本的自然事實和科學的概念，甚至強調科學概念的應用；只有少數國家強調實驗室的各項操作能力和進行科學實驗。