

第3章 多元化評量 20-41

教學與評量是現代教育系統中的兩大部分，兩者之間有互動，但其關係仍是教學為主，評量為輔。評量為教學的一項後續 (follow-up) 活動，用以了解學生學習成果，及調整或擬定日後教學策略。但由於學生、家長、甚至許多老師未能了解評量的真正目的及評量結果所代表的意義，而僅強調其中所呈現的片面資料，例如成績、在全班的位序或排名等；再加上傳統評量方式設計不周，不夠多元化，未能涵蓋教學內容的各個層面，形成教學與評量之間的落差，使得評量的角色有機會易客為主，對教學產生反向的導引作用 (backwash)，亦即所謂的「考試領導教學」。所幸，這種反向的導引作用，雖然可能為阻力，亦可能成為助力 (Hughes, 1989)，端看如何因應。以國內大學聯考為例，將英文作文納入考試，進而帶動國內英語老師及學生對英文作文的重視，便是因勢利導，使一個反向的導引作用成為助力的最佳例子。基於多元評量的重要性，本章將釐清一些多元評量的基本理念，並探討如何將其落實在九年一貫新課程的英語科中。

3.1 評量的定義：評鑑、評量與測驗 20

許多外語教學的學者如 Valette (1994)，對 evaluation (評鑑/評量)、assessment (評量)、testing (測驗) 並不作特別區分；更多的學者僅作二分法，一類是層次較低的 testing，另一類則層次較高，有的學者用 assessment 一詞，如 Bailey (1998)，有的用 evaluation，如 Genesee and Upshur (1996)。

事實上，這三者之目的及範圍上，皆有所不同 (Brown, 1989; Lynch, 1996)。testing 範圍最小、也最明確，其試題皆有“正確”或“最恰當”的答案，如段考、隨堂考等皆屬此類。assessment 範圍較廣，以測驗、面談 (interview)、實作或工作計畫 (project)、檔案評量 (portfolio assessment) 等方式，診斷學生的語文能力。evaluation 層次更高，係“有系統地蒐集相關資料以改善課程、評估其成效、了解學生的態度” (Brown, 1989)，甚至可包括“評估測驗試題的品質” (Alderson, Clapham, and Wall, 1995)。在中文裡，此三詞可分別以『測驗』『評量』『評鑑』翻譯，界定明確，不至於有所混淆。不過，如上所述，有一些常用詞語如 formative evaluation 及 summative evaluation，因許多外國學者沿用已久，故其本質雖為評量，英文部份雖然有時仍沿用 evaluation，中文則以「評量」稱之。

3.2 評量的目的

21-22

Bailey (1998) 指出正式的語言評量之目的有以下八種：(1) 評估語言性向 (aptitude test)、(2) 檢定雙語使用者的優勢語言 (language dominance test)、(3) 語言能力檢定 (proficiency test)、(4) 入學篩選 (admission/screening test)、(5) 編班 (placement test)、(6) 診斷學習困難 (diagnostic test)、(7) 了解學習進展 (progress test)、(8) 檢定學習成就 (achievement test) (註一)。她甚至將評量與課程設計連接，在教材設計編選的過程中，採用不同目的的評量，參考其結果，以找出最適合自己學生的教材。以此觀之，不妨可將“編選／修正教材”列為第九項目的。

這些目的在學校正規的外語課程裡，僅第 (3) 至第 (9) 項適用；而我國中小學採常態分班，因此，第 (3) 至第 (5) 項又不適用，而僅餘“診斷學習困難”、“了解學習進展”、“檢定學習成就”、“編選／修正教材”等四項符合現況。事實上，國內英文老師大多已於日常以小考及非正式的評量，評估學生能力、診斷其學習困難，符合第 (6) 項目的；而進一步以此結果，作為增刪、編選教材的依據，則符合第 (9) 項目的。若將學習前、學習中的各個階段、及學習後等任何兩個時間定點上的評量結果相比較，尚可獲知學生的進步情形，例如語言能力的增長、困難的克服等，則符合第 (7) 項。至於中小學的段考，則是以該階段為單位所作的總結式評量，用以評定學生學習成果，給予成績，屬於第 (8) 項。

傳統上，評量的主要目的在於以此作為學生升級、留級的判斷標準。但現代的評量觀念則認為評量所產生的一些資訊不應再僅留存於老師處，而宜讓學生全程參與或被告知，以了解個人進步情形。換句話說，這些資料不僅可作為老師對學生能力評定的依據，更可作為學生對自己學習過程的一種自我檢視，以修正自己的學習方向或策略，而其反應或所採取的行動，便可能決定其未來學習的成敗。無論是升級與否，抑或調整學習態度或方法，對學生的影響皆既深且鉅。

學生固然能由評量中獲益，老師亦能由評量所呈現的資料中，獲得極寶貴的回饋 (feedback)，了解其教學上哪些部份最能引起學生的興趣，哪些部份學生最感困難，哪種教學方法最為有效，哪種教學技巧最有效率，以期在教學上

(註一) 許多學者採簡易的歸類，例如 Alderson, Clapham and Wall (1995) 便將評量依其目的分為五類：1) placement tests, 2) progress tests, 3) achievement tests, 4) proficiency tests, 5) diagnostic tests。

精益求精。Sutherland (1999)即指出，評量能讓教師評鑑其（1）教學是否有效（effectiveness of teaching）（2）教材是否適宜（appropriateness of the textbook）（3）課程設計是否有效（effectiveness of the syllabus）（註二）。

此外，評量的結果亦提供學生家長極重要的資料，讓其了解孩子在學校的學習狀況（Sutherland, 1999）。事實上，國內許多老師請家長在學生的成績單、考卷、或批改過的作業上簽名，其目的即在於提供機會，讓家長了解子弟在學校的學習情形。

3.3 多元化評量的基本理念

鑑於評量本身的重要，及其多元化的必要，教育部於民國八十七年八月二十六日修訂公佈「國民中學學生成績考察辦法」，規定：學校對國中學生成績之考查，應視學生身心發展與個別差異，以獎勵及輔導為原則，並依各學科及活動性質，就下列十五項評量方式選擇辦理：紙筆測驗、口試、表演、實作、作業、設計製作、報告、資料蒐集整理、鑑賞、晤談、自我評量、同儕互評、校外實習、實踐、及其他。隨後，又在民國八十七年九月三十日公佈的「國民教育階段九年一貫課程總綱綱要」中特別提及，各科的“評鑑方法應採多元方式實施”。

配合溝通式教學觀念，在進行多元化語言評量時，有許多原則性的理念必須注意，方能落實評量的真正功能，此處僅就其中主要的八項基本理念討論如後。

一、評量範圍應配合教學目標

David Nunan (1990) 將評量定義為“評估學生知識、技能程度的一連串步驟”。在外語學習中，所評估的確切知識及技能則視該語言課程的目標而定，例如「文法翻譯教學法」以閱讀外國文學作品為目的，因此，其所評量的範圍多局限於學生對語言的閱讀能力，並多以翻譯方式進行，學生只要對文法結構與字彙充分掌握，即可獲取高分。

語言學習的重心在語言能力之培養，理所當然地評量的重點應置於學生語言能力之發展。但事實上，溝通式教學法旨在培養學生運用所學的語言（Larsen-Freeman, 1988），因此，語言評量時除了各種語言成份

（註二） Sutherland 所談的教學效能與 Bailey 的第(6)至(8)項概念相似；教材與課程設計兩項則與 Bailey 所提及的第九項相呼應。

(字母、發音、字彙與句型結構)外,更應兼顧聽、說、讀、寫四種語言技能。我國九年一貫課程英文科的教育目標包含三方面:(1)語言能力;(2)學習態度與方法;(3)文化習俗。其中,“學習態度與方法”及“文化習俗”雖屬副學習(associate learning),但是完整的評量仍應涵蓋觀察學生的學習態度是否積極,學習方法是否正確,並評估其對外國文化習俗的了解是否有所增長^(註三)。

二、評量方式須多樣化

現代的語言評量必須多元化,所謂“多元”可以從不同的角度來看。就評量時間來看,評量可以是課程完畢時才實施的期末考試(亦即“總結式評量”(summative assessment)),亦可能是過程中經常性的評量(亦即“形成式評量”(formative assessment))。“多元”也可以指評量形式上的多元,因此不僅有傳統的“測驗式評量”(test-based assessment),也有以活動、遊戲方式進行的“活動式評量”(task-based assessment)。“多元”也可以指內容上的多元,因此不僅評量各項語言成份,亦評量聽、說、讀、寫四種語言技能。“多元”也可以指評量工具的多元,除了傳統的紙筆測驗,亦可以多媒體(如電腦、錄影機、錄音機等)進行。評量者的“多元”,指的是傳統的老師評量外,並納入“自我評量”(self-assessment)甚至“同儕評量”(peer-assessment)。此外,在測驗式評量中,單項式語言成份與整合型語言運用兩種性質的試題兼容並蓄;強調以學習者參與決定評量內容及著重學習過程的「檔案評量」(portfolio assessment),等等亦皆為多元化評量的體現^(註四)。

三、講求效度與信度

一如其他學科的評量,英語科的評量亦須講求信度與效度。

(一) 信度

信度指的是評量工具的精確、穩定程度。信度高的試題,若對相同的受試者先後進行兩次測驗,其結果應該是一樣或極為相近(兩者的相關係數即為其信度)。但受到某些因素的影響,兩次評量結果完全一樣(亦即信度為1.0)的情形不易達成,以小規模的課堂評量而言,.70的信度已屬相當不錯(Cohen, 1994)。

語言評量工具的信度取決於三個因素:評量工具因素、情境因素、個

(註三) 現代國民中小學注重德、智、體、群、美五育均衡發展,評量的範圍應擴大至教育的全面。因此,成績考察的範圍應不僅包括主學習(primary learning),亦應涵蓋副學習(簡茂發, 1999)。

(註四) 外語初學者對該語言的掌握能力不足,無法給他人的語文程度精確合宜的評估,同儕評量的方式較不適用。因此,九年一貫新課程英文科的評量,尤其是國小階段,較不宜採行。

人因素(Cohen, 1994)^(註五)。其中，評量工具因素包括『學習目標的涵蓋程度(或取樣情形)』、『評量項目的明確與否』、『學生的回答是否受到限制』、『作答說明是否明確』、『試題的版面或呈現方式的品質』、『學生對評量方式的熟悉與否』、『整套評量的長度』等(Hughes, 1989)。情境因素指的是評量場所的情況，如照明設備、音響效果、桌椅舒適度、室外噪音等。個人因素則包括個人可變性的生理及心理特質(如動機、與施測者的關係等)及個人永久性的因素如 I.Q.、學習能力等。

因此，爲了確保評量工具的品質，國內國民中小學英語科的評量方式，不管是傳統的測驗或我們提倡的活動式評量，皆應均衡地涵蓋絕大多數的學習目標，就此而言，一成不變的傳統測驗已然不敷所需，評量方式的多元化是必然的趨勢。以學生熟悉的題型讓學生有充分的空間表現其能力、並提供明確的作答說明，亦有絕對的必要；此外，評量應力求一個舒適、無引發焦慮之虞的施測環境，同時儘量激發學生的動機去完成該活動，以此觀之，本研究所提倡的活動式評量，正能提高評量的信度，因爲讓教學活動及遊戲兼具評量的角色，學生此時的焦慮遠低於正式的測驗情境，較能充分表現能力。

(二) 效度

信度固然重要，仍須效度的配合。若一份試題若所測者，並非我們所要的語言能力，即使其信度再高，亦是枉然。簡而言之，一個評量活動所真正評量出的語言能力，應與當初所擬評量的語言能力種類及項目一致，這種一致性便是效度。檢驗某一評量工具的效度，方法有好幾種，也因此有許多種的效度，例如：『內容效度』(content validity)旨在檢驗是否涵蓋足夠的學習內容或目標、『一致性效度』(concurrent validity)以另一業已建立效度的類似評量工具作爲比較的對象得出相關性、『建構效度』(construct validity)關心的則是所評量者是否真的代表了語言能力抑或其他的能力(如猜題能力、記憶能力)、『表面效度』(face validity)則僅就評量的類型或方式判斷其是否真能測到所欲測的能力等等，不一而足(Cohen, 1994)。

在九年一貫英語科的評量中，我們較關心的是『建構效度』與『內容效度』。就前者而言，我們必須確定，在溝通式的教學法中，真正的英語學習是學會使用英語(use of English)而非對英語的這個語言的了解而已(knowledge of English)。因此，評量應該是在真實的溝通情境中進行，重點則在看學生是否藉由英語達成溝通的目的，而非傳統的測驗

(註五) Cohen 的原文雖僅使用“測驗”兩字，而未提及評量一詞，但他此處所提的原則，不僅侷限於傳統的測驗，亦適用於其他類型的評量方式。

文法知識。就此點來看，本文後面所提的活動是評量正符合此需求。『內容效度』亦是我們在國小英語評量時應該特別注意的，隨時提醒自己，試題必須反映這個階段的學習與評量目標。不要忘了評量的目的在了解學生是否已經學會，學習困難在何處；切勿爲了將學生的程度區分高下，而偏離評量目標，在試題的難度上大費周章。

四、配合學生的年齡層

Halliwell (1992) 將外語課程的學習目標分爲兩類：『課程內容目標』(content goals) 與『態度目標』(attitude goals)，前者指的是發音、字彙及文法等各種語言成份及聽說讀寫四種語言技能的養成，後者則爲正確學習態度的培養。她並主張在小學的初學階段，應以『態度目標』爲主；隨著年齡的增加，逐漸加重『課程內容目標』的比重。

國內九年一貫課程英文科的教育總目標，不僅包含了這兩項，同時在第二個總目標中，將正確的學習方法與學習態度並列；此外，更增列了對中西文化的了解作爲第三個總目標。學習態度、方法、及文化了解，皆屬於非語言層面的，應該列爲初學階段（例如：國小五、六年級）的重點，俟學習經驗較豐富後（如國中一、二、年級），正確的學習方法與態度業已略有概念，此時方宜將『課程內容目標』列爲重心。

除了評量目標，評量方式亦應隨著學生的年齡不同而調整。國小階段五、六年級，每週上課時間相當有限，用以教授新教材、複習及進行綜合活動等已嫌不足，實不宜浪費太多時間於正式的評量中，與國中相較，更應該以活動式評量爲之。總結式評量『一次定生死』及『重在學習成果的評定而對學習的改進已無意義』的缺點，讓我們覺得，應該儘可能使用形成式評量；然而，形成式評量雖可能招致『試題較無整體性規劃』、『評分較爲主觀』等批評，但這在『計分多寡影響排名、事關將來升學』的國中較有影響，在小學則無此顧慮，因此，形成式評量的實施，應該較無阻力。總而言之，國小階段，應以形成式評量爲主，且儘量採用活動式的評量方式；到了國中階段，才略微增加總結式評量。

五、形成式與總結式評量並重，正式測驗與非正式評量兼用

多元化的教學評量觀念主張，除了學期末或教學單元完畢後驗收學習成果的總結式測驗，更應於平日進行形成式的評量，觀察學生在學習過程中的表現及其進步的情形。

總結式的評量一般大多以傳統的紙筆測驗、面談、聽力搭配紙筆等正

式測驗的方式進行，事實上，亦可以較活潑、強調語言運用的活動式評量為之，目前許多學校已然實施之全班或全校性的「過五關」評量活動便是一例，在該項活動中，學生必須以所學的英語在“實際生活的情境”中進行問候、點菜、問路、唱歌等語言溝通活動。

形成式評量的資料來源不再是正式測驗的成績或結果，而是學生學習歷程中的種種表現。動態方面，於課堂中進行各種觀察，包括對老師上課時非正式問題的回答、練習活動中語言能力的呈現、甚至其上課提問題及回答問題時所表現出的學習態度與方法等皆可納入；靜態方面，收集各種書面資料，除了傳統的評量項目習作，學生的學習日誌（learning journal）、學習問卷等都在收集之列，其中最具特色的是學生學習檔案的建立，這種『檔案評量』是英語科新課程中新增列的，目的在收集更多的資訊，以期更詳實地勾勒個別學生的學習歷程、進步情形、及成果。

六、老師評量為主，自我評量及同儕評量為輔

傳統上，評量皆是由老師進行。但是，部份的評量亦可由學生本人及同學來作，其好處在於提供另一個角度的資訊，特別是一些非課堂內的語言學習活動或語言表現；學生作自我評量，讓學生有機會反思自己的學習過程及成果，對自己的學習負責，在改進學習態度及調整學習策略上頗有助益；此外，讓學生參與評量，對自己的成績有部份的主導權，且以一種非考試、非測驗式的方式進行，可降低學生的不必要焦慮。

但受限於外語初學者所學仍相當有限，不易對彼此的語言能力作正確的評估，同儕評量以評量同學的學習態度、方法及文化了解為原則，且次數宜少、比重宜低，例如一學期一次。自我評量宜以較輕鬆、非正式的方式進行，次數可稍多，評量的範圍可涵蓋語言能力、學習態度方法、及文化的瞭解，評量表可馬上收回或置於學生的個別檔案中。必須注意的是，自我評量與一般評量一樣，診斷學習困難的功能應大於評定成績的作用。

七、語言的流利重於語言的正確性

在溝通式教學中，學習外語的目的是希望能以此語言與他人溝通，與傳統的其他教學法強調對該語言有正確的知識與了解有很大的不同（Larsen-Freeman, 1988）。一個成功的外語學習者，當然應該不僅能精確地遣詞用字，更能流利順暢地表情達意。但是，初學者經常無法兩者兼顧，此時，傳統的教學方式經常因處處要求合乎語法、用字精確、

發音標準等，而犧牲了說話時的流暢；溝通式教學，則對學生語言使用的一些錯誤採取寬容的態度，以讓學生充分表情達意為首要考量。對於學生所犯的錯誤，除非是一再出現或因此會嚴重影響上下語意了解者，才於事後予以糾正，以免造成學生心生畏懼不敢表達。

基於這種精神，九年一貫英語科評量的重點，閱讀與聽力部份應以文意、語意的了解 (comprehension) 為主，口語表達與簡單的寫作則以是否順暢地表情達意、達成溝通目的為主，初學階段小學五、六年級的英文科評量，尤其是如此。

『語言流利重於語言正確』的原則，固然適用於整個外語學習的教學過程與評量過程。不過，這並不意味著語言的正確便可因此棄而不顧。當教學與評量重點為字母、發音、字彙、語法等語言成份時，學生表現的精確與否仍非常重要；其中，發音在國小英語的階段，更應儘量要求準確。許多研究 (Ellis, 1986; 楊懿麗, 1988) 都顯示，兒童學習外語很大的一項好處在於較容易習得較道地 (native-like) 的發音，而若要發音道地，在初學時便應要求相當程度的精確。前面第二章亦曾提及，Yalden (1987) 指出，溝通式課程包含語言形式 (linguistic form) 與溝通功能 (communicative function) 兩者。而要教授或評量語言形式，便不得不將語言形式的正確與否列入考量。

八、評量結果旨在了解學習困難與進步而非獲致學期分數

依 3.2 中所列的各種評量目的來看，九年一貫英語科的評量旨在診斷學生的學習困難與進步情形，而不是檢定這個學生是否達到升等或通過的標準，或需不需要留級。因此，若要評量某一項目（如：利用『字母拼讀法』念出字母“a”的長母音與短母音、是否會使用 yes/no 問句，等等），應該多次使用不同的評量方式或活動，以確定學生已然學會或其困難所在，這也是多元評量的精神所在；至於分數的評定，不妨就其中選取最佳的一次給予分數。

這種『多次多元方式』的評量，除了給學生充分的機會表現能力以外，依 Cohen (1994) 所言，藉由比較各次、各種評量的結果或成績，尚可同時增加評量工具的效度。評量方式的多元與次數的增加，已非傳統測驗方式所能負荷，唯有多採用形成式評量，並以不同活動、遊戲的方式進行，方有可能成功。

3.4 多元化評量的實施方式 28-39

綜合上述各項評量原則，九年一貫英語科的評量方式建議如下：

表 3-1 英語科評量類別

類別	項目	評量重點	比重**
總結式評量	1. 活動式評量	語言能力	國小：30%
	2. 測驗式評量 (例如:段考、期考)	語言能力	國中：50%
形成式評量	1. 活動式評量	語言能力	國小：70%
	2. 課堂觀察*	學習態度與方法、文化了解	
	3. 檔案評量	語言能力、學習態度與方法、文化了解	國中：50%
	4. 作業評量	語言能力、學習態度與方法、文化了解	
	5. 進步情形評量	語言能力、學習態度與方法、文化了解	
	6. 測驗式評量 (例如：小考)	語言能力 (僅國中階段酌予採用；國小階段儘量不採測驗式)	

* 課堂觀察，廣義來說，可以觀察學生上課過程中靜態與動態、語言及非語言的各種行為表現 (Genesee and Hamayan, 1994)。但在語言課堂中，仍以學生的語言表現為重心，因此，本文特別將此單獨列為“活動式評量”，而課堂觀察僅著重於觀察非語言的“學習態度與方法”及“文化了解”。

** 依此模式評量的成績可登錄於如本文附錄一所示之成績登錄表中。

九年一貫英語科的評量應由總結式及形成式兩種評量方式為之。其中，總結式評量為正式評量 (formal assessment)，因為不論是傳統的測驗式評量，抑或如上述「過五關」的活動式評量，皆為學期末或單元結束時，老師為評估學生語言能力而特別安排的評量活動。形成式評量以各種非正式的評量活動 (informal assessment) 為主，例如，利用教學過程中的活動來評量學生能力及學習情形的「活動式評量」、以學習態度為重點的「課堂觀察」、及以呈現學習過程為目的的「檔案評量」皆是；值得注意的是，形成式評量固然可以較傳統的測驗方式進行 (如日常小考)，為減少學生考試的焦慮，應儘量少採用；國小

階段，以培養學生學習英語的興趣為主，尤其應該避免實施小考。

兩類評量所佔學期總成績比重，則依教學目標的重心及階段性評量目的之不同而做適度地調整。國小階段，原則上，以形成式的評量為主體，佔學期總成績百分之七十，總結式評量結果佔百分之三十。國中階段，因學習目標中「語言能力」部份的加重，更重要的是，評量結果將作為日後高中升學的參考，因此，可考慮適度加重客觀性高的總結式評量，提高至百分之四十或百分之五十，藉以提高整體成績的公信力，減少不必要的困擾。若將本表第三欄的「比重」改為「得分」，並加上學生個別的姓名，即可依設定之比重給分，作為學生個人學期成績表。

此模式中之「執行方式」所列的各總評量活動說明如後：

一、測驗式語言評量：

雖然形成式評量與非正式評量近年來在外語教學界備受青睞，許多學者亦大力推廣，並不表示我們教學尚便可自此貶抑、甚至排拒正式的測驗。形成式與非正式評量重視學生學習過程中的（每一步）進展，使得教學與評量的界限模糊了，Brown (1994) 指出，兩者之間的界限仍有適度釐清的必要，而且是兩者各司其職。我們雖然不贊成教學與評量涇渭分明，但適量的評量成績是由“純評量的方式”（亦即為了評量而刻意實施的活動）取得，仍有其必要。從學生的內發性學習動機（intrinsic motivation）與情意因素來看，為了提供最佳學習環境，應該鼓勵學生在學習過程中盡量嘗試使用所學，讓其了解犯錯無妨，目前的錯誤不僅無礙日後英語溝通能力的培養，反倒是一項助力；而事先排定的正式（尤其是總結式的）測驗，正可以讓學生明瞭何時將自己調整成最佳狀態即可，之前可盡情地以各種方式或內容練習、運用所學，而不必步步為營，擔心犯錯會被扣分。

測驗式評量不論是以紙筆測驗方式或以口試進行，皆應依上述所提評量效度、信度之原則設計。此外，除了單項（discrete-point）式試題外，更應包括整合（integrative）型試題，愈是總結式的評量，以及學生的程度愈高時，後者所佔的比重應愈高。換句話說，即使是在初學階段，教學以單項的語言成份為重心，評量時仍應提供較多的情境(context)，以整合型的試題為之。

二、活動式評量

活動式評量顧名思義，指的是以活動的方式進行評量，與測驗式評量強調「個人智能表現」形成強烈對比的，此處的活動具有較多的師生及同學間的互動，運用較多的肢體動作，涉及較多的團隊合作。簡易

者，如辨認語音時要學生以舉牌的方式或起立坐下來表示異同；繁複者，如讓學生利用英語完成某項任務或解決問題(problem-solving task)，其他如角色扮演、韻文歌曲吟唱、遊戲等皆是常用的活動式評量。如表 3-1 所示，活動式評量可以是為了總結觀察學生的學習成果而安排，但此類的活動更常扮演教學（學習）的角色，讓學生練習或複習所學。活動過程中，學生呈現各個階段的語言能力，老師若加以評量並記錄，則可同時了解每個學生的語言發展，這種以教學過程中的活動來評量學生的語言學習成果，以彌補傳統測驗「一次定生死」的缺憾，這正是形成式評量最大的特色。

如表 3-2 所示，活動式的語言評量，可針對語言成份與語言能力兩部份進行，前者包括字母、發音、字彙、文法句型，後者則包括聽、說、讀、寫四種能力及四者的綜合運用。語言成份的評量，應以語言的“形式”（form）及“精確”（accuracy）為重；聽說讀寫語言能力的評量，則應以語言的“意義”（meaning）及“流暢”（fluency）為第一要務。

此類活動若是用於形成式評量，一方面是因不便要求每位或每組學生重複相同的活動，一方面是因為老師同時要兼顧教學與評量兩項工作，不易作大規模的評量，比較務實的做法是一堂課只觀察一部份（例如五、六位）學生（Heaton, 1990），但基本原則是在同一單元裡的數堂課中，儘量至少評量到每一位學生一次。

評量時，應以事先準備的觀察表（checklist）為之，較有效率，依條列的觀察項目，逐項地於「優、可、差」之中圈選一項即可。必須注意的是，表中所列目標，若在這個教學活動中無法觀察到，或非此次評量重點，則可略而不評。此表可用於大多數的綜合活動中，對學生的語言表現作一般性的評估；若活動重心或評量重點在於某單項語言成份或能力時，表 3-2 顯然不敷使用，此時，應使用各單項語言成份或能力的專用表格（請參見第四至第九章）。

表 3-2 活動式評量表

學生姓名：_____

活動名稱：_____

評量得分：_____

評量日期： / /

評量目標	評量項目	表現
語言成份—字母	能辨識、唸、及書寫	優 可 差
—發音	能辨識及唸出單音、字句重音及句子語調	優 可 差
—字彙	能辨識及唸出字詞之音、形、義	優 可 差
—句型文法	能判斷、使用句型及文法結構	優 可 差
語言能力—聽	能了解句意及簡易對話，並做適當回應	優 可 差
—說	能依對象、場合有效地表達、溝通	優 可 差
—讀	能辨識及朗讀字詞，書寫句子、短文或信函	優 可 差
—寫	能摹寫字詞，書寫句子及短文、信函	優 可 差
—綜合能力	能利用字母拼讀法看字發音、聽音拼字	優 可 差
—綜合能力	能看懂並填寫簡易表格、資料	優 可 差
—綜合能力	能聽懂簡易對話，並以簡單的句子述說或寫出大意	優 可 差
—綜合能力	能看懂並填寫簡易表格、資料	優 可 差

三、課堂觀察

Halliwell (1992) 指出，教學目標中的「態度目標」(attitude goals) 與「內容目標」(content goals) 的比重，初學階段應以前者為主，隨著兒童年齡及學習經驗的增長，後者慢慢開始加重。事實上，我國九年一貫英語科中，當初國小階段增列英語科的初衷之一便為培養學生學習英語的興趣，其重要性不下於（甚至高於）英語能力之訓練，而態度目標的達成與否，難以測驗方式為之，須仰賴平日課堂內的觀察。老師可以相隔較長的時間（如每兩週、每個月或每個兩月）進行一次，實施時，可仿上述語言能力的評量的方式，上課前先選定一至二組的

學生，於上課中進行較專注的觀察，下課後再以回想的方式，依學生的整體表現於表 3-3 或類似的評量表上分別給分^(註六)。

表 3-3 課堂觀察評量表

學生姓名：_____ 活動名稱：_____

評量得分：_____ 評量日期： / /

評量目標	評量項目	表現
學習態度、方法	1.上課專心、認真	優 可 差
	2.熱心參與課堂討論	優 可 差
	3.樂於發問及自願示範、做活動	優 可 差
	4.不畏犯錯，樂於溝通	優 可 差
	5.準時繳交作業	優 可 差
	6.將英語與日常生活相結合，搜尋生活中常用之英語	優 可 差
	7.善用預習、複習、歸納整理以增進學習成效	優 可 差
	8.體會學英語的最終目的在溝通，而非文法文法	優 可 差
文化習俗的了解	1.能認識外國之主要節慶習俗	優 可 差
	2.能了解一些國際上的生活基本禮儀	優 可 差
	3.能了解一些英語交談時的基本禮儀	優 可 差
	4.能以簡易的英語介紹中外風土民情	優 可 差
	5.能尊重不同的文化習俗	優 可 差

四、檔案評量

傳統測驗評量的項目多侷限於語言能力，其結果常受學生考試焦慮(test anxiety)、學生考試當日身體狀況等因素的干擾。因此，近年來，國外許多學者積極發展新的評量方式，檔案評量便是其中深受矚目的一種，並廣泛地使用於各個教育領域中，如中小學生評量、大學生評量、甚至職前教師評量與現任教師證照換証等（張秀雄，1995）。而許多英語教學的學者與教師亦發現，檔案評量在呈現學生的成就、努力與進步方面，有極大的優點；同時亦能讓學生有自我評量的機會（Chen, 1999）

(註六) 課堂觀察評量表可稍加修改，而成爲小組評量表，一次評量五至六個學生。

(註七)。

Richards (1996) 更明確指出，學生的個別檔案提供了一項持續累積的語言發展記錄、勾勒學習過程的全貌、呈現進步的情形、提供機會師生一起設定目標一起評量，並呈現具體成果與家長、其他學生、教師一起分享。因此，檔案評量具有以下的諸多好處，值得於外語課堂內廣泛使用。

- 學生參與評量
- 學生負起自評 (self-assessment) 的責任
- 增進師、生、家長之間的互動，了解學生的學習
- 賦予學生對自己學習過程的主導性
- 增加學習的樂趣
- 培養學生針對學校功課獨立思考的能力
- 營造合作、分享的教室氣氛

總而言之，檔案評量最大的優點為賦予受評學生更多的主導權，讓其積極地參與、決定自己的學習過程與評量 (Newman and Smolen, 1993; Chen, 1999)，對學生學習興趣的提升有極大的助益。

實施檔案評量，大致可以下列方式進行 (Genesee and Upshur, 1996)：

- (1) 老師發給每人一份檔案夾 (或小盒子)，並決定存放地點。
為提高學生興趣，加強個人對檔案的重視，可鼓勵學生加以美化成一份具有個人特色的檔案。地點則以教室裡專用檔案櫃或學生個人儲物櫃最，老師可隨時調閱；若教室內無此設備，方由學生自行保管，定期攜帶至學校，由老師查看。
- (2) 師生共同決定檔案內容與格式
老師與學生共同討論檔案的主題為何，裡面應放些什麼，例如：所寫過的作文或信件、臨摹過的大小寫字母、在家練習英語發音或會話時的錄音、閱讀過的英文課外讀物、所購買的家電之英文說明、出國旅遊時各地的英文簡介、自行製作的英文聖誕卡等。學生在討論主題時，充分地參與，選擇作品置入檔案時，更是有絕對的自主權。不過，若學生的年紀較小、經驗不足時，在過程中老師應給予較多的建議。許多學者主張，凡是學生所挑選的，及學習過程中的作品如考試卷、作業等，皆應納入。如此一來，內容急速膨脹，加上未經去蕪存菁的

(註七) Chen (1999)一文指出，相關研究包括 Allaei (1992), French (1992), Gottlieb (1995), Hamp-Lyons, Voltmer, Pany, Davison, Campbell, and Davidson (1992), Markstein, Withrow, Prookes, and Price (1992), Moya and O'Malley (1990)。

篩選，有些作品可能與語言學習並無直接的關係，每份檔案可能便成 Genesee and Upshur (1996) 所稱的「垃圾櫃」了。一方面為避免誤蹈此窘境，一方面因國內許多英語教師對檔案評量初次接觸，不宜驟然增加師生太多負擔，九年一貫英語科的檔案評量，應與傳統上老師指定的作業或練習分隔，所納入者均以學生自發性的、課堂外所蒐集或製作的作品為原則；此外，學習檔案旨在呈現學生的學習結果，因此，檔案所蒐集的資料雖鼓勵具應用性、與生活有關者，但仍應儘量以與課堂所學有關為原則。事實上，評量與教學密不可分，兩者若能配合，檔案評量尚有提供老師改進教學的功效。

(3) 學生定期或不定期置入新作或抽換舊作

學生了解如何建立、灌溉自己的檔案後，便可開始，隨時擷取自己最得意的成果或優良表現的記錄，置入檔案中。如上所述，為避免過於龐雜，鼓勵學生求精進，亦可以新作抽換舊作，在同一時期內相同或類似的作品僅留一至兩份。而所謂的時期，在小學階段，建議大約為二至三週，中學階段則為一至二週；換句話說，扣除期初、期末較繁忙時期，小學階段每學期約六至九項（次），國中則八至十二項（次）。

(4) 老師檢視或師生一起討論檔案內作品

一旦個別檔案累積了數份成品後，老師便可開始檢視並加以評量。評量的項目固然包括所呈現的英語能力，更應注意學習態度、方法、投入程度等非語言方面的表現。評量的重點在於平日測驗、活動式評量、課堂觀察未納入者，例如學生對所學語言的歸納整理、於實際生活中的運用、對語言使用的敏感與創意等。成績的評定應兼重成果的表現與其過程中的成長。

檔案評量的個別化（individuality）特性除了表現在資料蒐集時學生的自主權外，亦表現在評量時學生的參與。老師可以個別談話(individual conference) 的方式進行，了解每件成品背後的相關資料（例如：呈現這件作品的動機、製作的過程、對此成品的自我評價或省思等），與學生共同探討其優點及如何改進。

學生的自我評量或省思不僅是在個別談話時進行，更可顧鼓勵學生於平日作，並置入個別檔案中。此類自我評量可依學生年齡以圖示、文字選擇、或問答式的文字敘述等方式進行。例如：

A、對初學字母及字詞書寫的國小五年級學生，老師可發給以下自評表，讓其根據自己的狀況填寫。

表 3-4 學生自我評量表（圖繪式）

姓名：_____ 日期：__年__月__日	
說明：如果你覺得以下的敘述與自己的狀況相符，請在右邊一欄劃上笑臉；如果不符，則請劃上哭臉。	
項 目	笑臉 / 哭臉
1. 依照格子的大小書寫英文字母	
2. 字母主要筆畫的寬度與高度比例與範例相似	
3. 不同字母之間的大小比例適宜、好看	
4. 字母間的間隔小而一致；字與字間的間隔大而一致	
5. 寫連續體時，字母的傾斜度一致	
6. 寫的字母／字詞美觀、易讀	

- B. 在學了相當數量的字母拼讀法後，可以下面選擇式與填空式混用的表格進行自評。

表 3-5 學生自我評量表（選擇／填空式）

姓名：_____ 日期：__年__月__日
說明：請依下面提供的選項，勾選（或填入）與你自己的狀況相符者。
1. 本週／月老師所教的單音，我能夠看到字母便唸出： 都不會 一小部份會 大部分會 全部會
2. 本週／月老師所教的單音，我能夠聽到語音便拼寫出： 都不會 一小部份會 大部分會 全部會
3. 本週／月老師所教的單音，看到字母時，我比較不會念的音是 _____
4. 本週／月老師所教的單音，聽到音時，我比較不會拼寫的是 _____

- C. 對較高年級的學生，不妨以下面問答題的方式，讓學生就聽、說、讀、寫的能力進行自評（改寫自 Sutherland, 1999）。

表 3-6 學生自我評量表（問答式）

姓名：_____	日期：__年__月__日
說明：假設今天你是老師，你會對你所交上去的作業或目前的上課表現做什麼樣的評語？	
1. 字體方面（美觀與否）	
2. 拼字方面	
3. 聽力	
4. 口語能力	
5. 閱讀能力	
6. 寫課堂討論結果／摘要	
7. 您對這一個學生，還有什麼要補充說明的？	

使用以上三種表格時，老師應先詳細說明填寫方式；尤其是填寫表 3-6 時，應先對學生說明每一單項的含意與重點，並提供範例，以提高學生自我評量的效度。此外，每一種自我評量最好能作兩次，中間相隔相當的時日。因為，評量時，將兩張相同的表比對一下，便可了解學生的進步情形。

老師對學生的個別檔案，可每半學期、一學期或大單元教學完畢後檢視，就其外觀、個件品質、內容類別、整體組織等加以評量，以如表 3-7 之評量表為之。並可比較各次結果，得知學生進步情形。

表 3-7 個別檔案評量表

班級／組別： _____

學生姓名： _____

項 目	第 一 次 日期： / /	第 二 次 日期： / /	第 三 次 日期： / /
1. 整潔美觀	優 可 差	優 可 差	優 可 差
2. 作品的排放組織	優 可 差	優 可 差	優 可 差
3. 作品的數量	優 可 差	優 可 差	優 可 差
4. 作品的品質	優 可 差	優 可 差	優 可 差
5. 作品中英語文的活用性	優 可 差	優 可 差	優 可 差
6. 作品中運用英語文的創意	優 可 差	優 可 差	優 可 差
7. 內容的多樣化	優 可 差	優 可 差	優 可 差
8. 自評表的填寫	優 可 差	優 可 差	優 可 差

五、作業評量

九年一貫英語科的作業形式應多元化，而不僅限於紙張式的文字抄寫練習。而且，除了個別作業外，亦可偶而採兩兩一對小組合作的方式進行。作業評量的範圍，除了傳統的習作，亦應儘量採研究計畫 (project) 等活動式的作業，例如：將今天課堂討論的結果用英文寫成簡短的摘要；以該組課堂討論的結論徵詢班上其他同學（或家人親戚）的意見（同意或反對），並將調查結果以英文製作簡易的圖表；每一組同學挑定一項今天所教的食物，蒐集同一類別的其他食物圖片及其英文名稱，正確搭配後繳交；撰寫簡易英文短句或故事；兩兩一對，錄製簡易的英語對話等。

評量的標準視該項作業的性質而定。若是屬於統一規格的制式作業，則僅就語言成份與語言能力的優劣標準評之；若是屬於活動式、或可以不同風貌呈現者，則除了語言成份與語言能力之外，其他項目亦應

列入評量；這些項目可參照評量個別檔案時所考量的指標即可，例如：創意、活用性、品質等。

六、進步情形評量

老師對學生的學習成就，往往是以其實際所呈現的能力來評斷，而未考慮到課程開始時各個學生的差異。這種起跑點的不同，扭曲了真正的學習成效，對來自社經地位較差家庭的孩子，是極大的不公平。這種情形，在國內的國小英語教學更是嚴重，許多家長在這種怕輸在起跑點的心態下，紛紛“偷跑”，提前將孩子送往兒童美語補習班上課。

面對此亂象，老師在評量時作法有二。第一，老師應將所評量的結果減去上課前學生既有的能力，所得的數值便是學生的進步幅度，並將此列為學期成績重要的考量，屆時，若學生的最後一次評量仍未達標準，但其進步幅度頗大，仍應給予相當的肯定。第二，若老師採用 mastery learning 的觀念，只要學生在這學期中學會課程目標即可，何時學會皆無妨；因此，只要在第一、第二次或第三次評量中有一次展現此能力，學生在此單項能力的表現即為滿分。兩種做法皆可以下列『進步情形評量表』記錄學生的表現。

表 3-8 進步情形評量表

班級：____年____班

學生姓名：_____

評量目標	評量項目	第一次評量 / /	第二次評量 / /	第三次評量 / /
語言成份—字母	能辨識、唸、及書寫	優可差	優可差	優可差
—發音	能辨識及唸出單音、字句重音及句子語調	優可差	優可差	優可差
—字彙	能辨識及唸出字詞之音、形、義	優可差	優可差	優可差
—句型文法	能判斷、使用句型及文法結構	優可差	優可差	優可差
語言能力 — 聽	能了解句意及簡易對話，並做適當回應	優可差	優可差	優可差
— 說	能依對象、場合有效地表達、溝通	優可差	優可差	優可差
— 讀	能辨識及朗讀字詞，書寫句子、短文或信函	優可差	優可差	優可差
— 寫	能摹寫字詞，書寫句子及短文、信函	優可差	優可差	優可差
—綜合能力	能利用字母拼讀法看字發音、聽音拼字	優可差	優可差	優可差
—綜合能力	能看懂並填寫簡易表格、資料	優可差	優可差	優可差
—綜合能力	能聽懂簡易對話，並以簡單的句子述說或寫出大意	優可差	優可差	優可差
—綜合能力	能看懂並填寫簡易表格、資料	優可差	優可差	優可差
評 量 得 分				

3.5 結語 39-41

限於篇幅，本文僅就多元化評量的各個層面中較常見、實用者（例如：總結式與形成式、測驗式與活動式、教師與自我評量、檔案評量等）擇要簡述。

雖為“簡述”，文中提及的各種評量理念與評量方式，倒也林林總總，可能讓初次接觸英語評量的教師一時不知所措，爲了即將面臨的各種“決定／抉擇”（例如：該用何種原則、方式、表格以及何時用、如何用）及伴隨而來的龐大評量資料發愁。

事實上，本文旨在提供一些有別於傳統而較理想、合宜的評量方式，讓英語老師們從中選用，更希望藉以激發老師們因時、因地制宜，發展或修訂成更適合自己學生的評量方式；而評量類別或次數的增加，目的是讓學生（尤其是某次評量不合理想的學生）有不只一次的機會或層面去展現自己的能力，因此，如果此次評量，有三分之一的學生上不能確切掌握某個語言成份或能力，則下次同樣重點的評量，僅需針對這些學生實施即可，如此一來，應可大量減少評量的負擔。此外，若能減少測驗式而增加活動式的評量，老師在評分時較易採行整體式 (holistic) 而非分析式 (analytic) 的給分方式，亦可節省不少時間。最重要的是，如前所述，語言課堂裡的評量旨在協助教學，用以檢視教法、教材的成效，老師毋需、也不應將評量當成英語課的重心，反客爲主，鎮日實施各種評量活動，造成學生學習與老師教學上的重大負擔。這點認識，在國小階段更是重要，當初將英語教育向前延伸至國小五、六年級開始，是寄望在較無升學壓力的學習環境下，能培養出學童對英語學習的濃厚興趣，若因老師過度強調評量或測驗，讓學童視英語課爲畏途，便有違九年一貫英語科課程的初衷了。

附錄一

九年一貫英語科成績登錄表範例 (國民小學)

學生姓名	形成式評量 (70%)													總結式評量 (30%)				總分(100%)
	活動式評量				課堂觀察		個別檔案		作業		進步	小考		活動式評量		期、段考		
	(1)	(2)	(3)	(4)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(1)	(2)	(1)	(2)	(1)	(2)	

- 說明：1. 國民中學部份，形成式評量與總結式評量各佔 50%。
 2. 各類評量的次數（如：總結式評量四次、課堂觀察兩次），可依實際需要增減。