

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；切勿為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為兩類：『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，後者則為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』為主；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為第三個總目標。學習態度、方法、及文化了解，皆屬於非語言層面的，應該列為初學階段（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』列為重心。

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為之。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』、『評分較為主觀』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，應以形成式評量為主，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。