

國立臺灣師範大學特殊教育學系
特殊教育研究學刊，民 94，29 期，73-94 頁

國小課程本位閱讀測量方法之信度與效度 研究

王梅軒

台北市立敦化國小

黃瑞珍

台北市立師範學院

本研究主要以「課程本位測量」(Curriculum-Based Measurement, 簡稱 CBM) 理論為基礎，嘗試研究中文課程本位測量方法在閱讀領域的信度與效度，以作為評量學生閱讀能力的依據。先用「國語文成就測驗」從台北市兩所國小二年級全體學生中，選取低、中、高閱讀能力組各 36 人，進行研究者自編之「課程本位閱讀測量」，包含「口語朗讀流暢度 (Oral Reading Fluency, 簡稱 ORF)」與「選字測驗 (maze task, 簡稱 Maze)」各八篇。研究顯示：1. 整體學生在口語朗讀流暢度、選字測驗上的表現與效標國語文成就測驗間有顯著的相關；2. 口語朗讀流暢度與選字測驗能有效區辨不同閱讀能力水準 (高、中、低) 兒童，且學生在相同版本，四篇不同文本，不論在口語朗讀流暢度、選字測驗之表現，均達到內部一致性顯著相關；3. 學生在康軒版與翰林版之口語朗讀流暢度與選字測驗之表現上一致性很高；4. 學生經由四週後之重測時，在口語朗讀流暢度與選字測驗整體之表現上均呈現進步情形，並達到顯著相關；5. 不同計分者使用指導手冊計分結果，口語朗讀流暢度與選字測驗均呈現計分者間一致性顯著相關。

本研究的研究結果顯示口語朗讀流暢度與選字測驗可以有效評量國小二年級學生之閱讀能力，這種省時、容易施測的課程本位評量具有良好的信度與效度。

關鍵字：課程本位測量、閱讀能力、口語朗讀流暢度、選字測驗

研究動機

「閱讀」是人類必備的基本能力，因為閱讀是獲得知識的重要技能，也是學習的主要媒介，因此，閱讀能力為個人獨立與否的一大關鍵（洪月女，民 87; McLoughlin & Lewis, 1994; Polloway & Patton, 1996）。閱讀雖然重要，卻有許多兒童在閱讀上有顯著的困難，根據美國 1997 年的官方資料指出，學習障礙者人數是所有身心障礙者之冠，佔所有身心障礙者 51.10%，有 80% 以上的學習障礙者有閱讀上的困難（Lerner, 2000）。相較於我國，教育部九十三年度特殊教育統計年報中指出，我國的學習障礙者佔所有身心障礙者的 18%，共 12,820 人；若以 80% 的學習障礙者有閱讀困難來推估，估計國內有閱讀困難的學習障礙者佔所有身心障礙者的 14.48%，約 10,256 人（教育部，民 93）。因此，如何評量閱讀困難的學生，並提供適當、有效的教學，是特殊教育的首要工作。

傳統用於鑑定或評量學生學業成就的主要工具是標準化、常模參照的成就測驗，但有許多研究者批評這種測驗方式，因為標準化測驗無法在短期間內重複施測，而無法靈敏地反映出學生本身的進步情形；測驗內容與課程無關，教師不能以標準化評量的結果作為教學的參考依據。此外，標準化測驗結果無法探究學生是否掌握特定學業技巧，亦未考慮學生的熟練度等。（Deno, Fuchs, Marston & Shin, 2001; Fuchs, Fuchs & Hamlett, 1989a; Fuchs, Fuchs, Hamlett & Ferguson, 1992; Marston, 1989; Shinn, 1989; Shinn, 1998）

因此，有些學者（Deno, 1985; Deno & Fuchs, 1987; Fuchs, Fuchs & Hamlett, 1989b）開始研究一種替代性的評量系統——「課程本位測量」（Curriculum-Based Measurement，以下簡稱 CBM）。CBM 非常重視將測量直接與課程和教學聯結在一起，其目的在於建立一種以學校

使用的課程為基礎的測驗，這類評量方式簡短、可經常性地實施，並能基於學生實際上課的內容，評估學生的基礎學業能力。自 1980 年起許多研究已顯示實施 CBM 的確有助於提升學生的學業成就以及提高老師的教學效能。CBM 在閱讀領域中最常使用的方法是口語朗讀流暢度（Oral Reading Fluency，以下簡稱 ORF）、克漏字（Cloze）與選字測驗（Maze Task，以下簡稱 Maze）等三種評量方式，許多研究證實它們具有良好的信度與效度（Deno, 1985; Fuchs, Deno, & Markin, 1984; Fuchs & Fuchs, 1986b; Fuchs, Fuchs, & Hamlett, 1989a; Fuchs, Fuchs & Stecker, 1989）。

近年來，美國學障的鑑定的取向有所改變，學障的評準從原本的「差距模式」逐漸變為監控學生進步情形的「教學反應模式」（Response-to-Instruction），美國教育部特殊教育司（U.S. Office of Special Education Programs）下所設置的國立學生進步監控中心（National Center on Student Progress Monitoring），就是以課程本位測量研究結果作為理論依據，經過持續追蹤兒童學習進步狀況來作為教育決定的重要依據（<http://studentprogress.org>），更可說明 CBM 之重要性。

相較於國外之研究，國內關於閱讀的研究較為缺乏，王瓊珠（民 90）在臺灣地區讀寫障礙研究回顧與展望一文中指出，在 1995 年之前國內讀寫障礙的研究十分零星，1995 年以後才突然增加許多研究篇數，較多的研究主題是以探討後設認知教學、識字教學和與閱讀能力直接或間接相關的認知因素研究為主。而在閱讀評量方面之相關研究亦明顯不足，僅有洪碧霞、江秋坪、蕭淳元（民 84）研究閱讀障礙篩選工具的發展；周台傑（民 86）研究國小閱讀障礙錯誤類型分析及有效教學；洪碧霞、邱上真（民 86）研究發展國民小學國語文低成就學童篩選工具系列；黃秀霜（民 88）研究中文識

字量表之編製及國語文低成就學童認字困難之診斷。國小低年級學童屬於閱讀的初學階段，若能及時瞭解學生的閱讀能力，針對閱讀困難學生進行輔導，或許更具意義，加上目前國內相關研究不足，因而引發研究者探討中文課程本位閱讀測量方法（ORF 與 Maze）之信度與效度，並建立中文版課程本位測量方法之指導手冊，以作為評量學生閱讀能力之方法。

根據上述之研究動機，本研究主要的目的如下：

一、進行課程本位閱讀測量方法（口語朗讀流暢度與選字測驗）之效度分析。

二、進行課程本位閱讀測量方法（口語朗讀流暢度與選字測驗）之信度分析。

三、考驗課程本位閱讀測量方法在康軒版與翰林版間不同文本之一致性。

四、綜合研究結果，發展一套完整課程本位測量方法之指導手冊，以作為學生閱讀能力之參照指標。

文獻探討

一、課程本位測量之意義與特色

課程本位測量是由 Deno 為首的美國明尼蘇達大學學習障礙中心研究群在八〇年代發展出來的，它是一套系統化的測量程序，其最主要的目的是發展經常性施測的程序，讓教師能夠作為決定是否（whether）或何時（when）該作教學上的改變並促成學生最大幅度的進步（Deno, 1985）。也就是將學校的課程當作評量的工具，提供客觀的數據資料做為教育決策的基礎。

課程本位測量針對標準化評量的缺失而設計，其優點有：1.施測過程短暫：只需要花費一分鐘到三分鐘，能夠重複地評量學生的表現；2.省時、省錢：內容取自學生之教材，結合教學與評量，不需要額外的測驗材料，也有

助於教師節省選擇評量題目的時間；3.藉由經常性的評量，瞭解學生的表現：它比一般成就測驗更能看出學生基礎學業能力的表現，發現學生學習的困境或是進步情形、評鑑教師的教學績效，做為改進的依據；4.個別參照與同儕參照：以圖表來記錄學生的進步情形，經由圖表能清楚瞭解兒童本身的能力狀況及兒童與同儕能力的落差，並可建立當地的常模來提供決定學生是否合於接受特殊教育服務；5.能增進溝通：評量的結果為明確的數據，且利用圖表說明學生之表現情形，很容易讓教師、家長、學生瞭解（邢敏華，民 84；Deno, 1985; Fuchs, et al., 1984; Fuchs & Fuchs, 1986a; Fuchs, Fuchs, & Hamlett, 1989b; Shinn, 1989）。

根據以上的原則，國外研究 CBM 的學者逐漸發展出閱讀、書寫、拼寫及數學等基礎學科領域的課程本位測量方式，目前在這四個基本學科領域已累積了相當豐富的實證資料。這些研究不斷指出實施 CBM 的確有助於提升學生的學業成就以及提高老師的教學效能（Deno, 1985; Fuchs, et al., 1984; Fuchs & Fuchs, 1986b; Fuchs, Fuchs & Hamlett, 1989a; Fuchs, Fuchs, & Stecker, 1989），其在發展初期是針對小學階段學習障礙學生之需求所設計的，近年來 CBM 適用對象已拓展到輕度至中度智能不足、嚴重情緒障礙、學業低成就和普通班學生，適用的年齡亦自小學往上下拓展至中學和學前階段（Faykus & McCurdy, 1998; Madelaine & Wheldall, 1999; VanDerHeyden, 2001）。

二、課程本位測量在閱讀評量上之研究

美國 CBM 研究自 1977 年即開始，至今仍有許多研究者依據不同的向度持續對 CBM 作研究，尤其在閱讀領域的研究成果更是豐碩（Deno, 2003）。CBM 發展初期，國外的研究者研究不同的測量方式，期望能找出信度效度皆高的閱讀測量方式，主要的測量方式有字詞朗讀、回答問題、事後回想、口語朗讀流暢度、

克漏字、選字測驗等方式，研究結果說明 ORF 和 Maze 是 CBM 當中最佳的兩種閱讀測量方式，能快速、有效地預測閱讀理解能力（Deno, 1985；Deno, Markin & Chiang, 1982；Fuchs、Fuchs & Maxwell, 1988；Marston, 1989；Shinn, Good, Knutson, Tilly, & Collins, 1992），以下針對 ORF 與 Maze 之相關研究分做說明。

（一）口語朗讀流暢度（ORF）之相關研究

閱讀能力在概念上包含了所有閱讀技巧及閱讀行為，口語朗讀流暢度和理解恰符合此領域，流暢及自動化解碼可視為基礎的閱讀理解能力（Shinn *et al.*, 1992），當讀者分配較少的認知成分於解碼時，就能將焦點置於意義的理解，ORF 與理解是交互作用的而非單向的，讀者必須閱讀得流暢才能增進理解，相對地，理解閱讀的內容才能讀得流暢（Allinder, Dunse, & Brunken, 2001），由此可知 ORF 與閱讀能力間有高度的相關。

Fuchs、Fuchs 與 Maxwell（1988）研究四種非正式測驗評量的效度：回答問題、事後回想、口語朗讀流暢度、克漏字，研究對象為 70 個輕度障礙的 9-15 歲男生，平均的閱讀能力在 3.3 年級，障礙類別為學障、情障、智障，研究結果顯示 ORF 有最強的效度，而事後回想的效度其次。Shinn、Good、Knutson、Tilly 和 Collins（1992）探討不同的閱讀評量之效度，其研究對象為三、五年級學生，以語音解碼、非字解碼、口語朗讀流暢度、文本理解、推論理解、克漏字、重述故事（書寫）等七項測驗來測量他們的能力，研究結果顯示 ORF 是一般閱讀能力很好的指標。

關於 ORF 的研究除了上述 Deno 等人（1982）、Fuchs、Fuchs 與 Maxwell（1988）、Shinn 等人（1992）以外，之後有許多類似的效標關聯效度研究，它們皆以學生在標準化閱讀測驗上的表現作為效標，Marston（1989）曾在其研究中記錄這些研究的發現，結果顯示 ORF

與效標之相關係數介於.70 到.95，大部分在.80 以上，肯定 ORF 可作為評量整體閱讀能力的指標。

Fewster 與 MacMillan（2002）研究 CBM 閱讀及書寫的指標，研究結果指出小學的閱讀及書寫表現與國中英文及社會課程的 CBM 分數有極高度的正相關，且一分鐘念正確的字數比拼正確的字數有較高的預測性，兩者皆可預測學業的表現，只是閱讀仍較書寫有較強的預測力。研究的結果支持了以下觀點：CBM 的分數，尤其是 ORF 可作為特殊教育的篩選及安置決定之用。

以上研究顯示 ORF 是 CBM 閱讀測量很有效度、省時方便的測量方式。因為 ORF 兼顧閱讀的正確性及流暢性，讀者必須閱讀得流暢才能增進理解；相對地，理解閱讀的內容才能讀得流暢（Deno, 2003）。

（二）選字測驗（Maze）之相關研究

許多研究已顯示 ORF 是評量學生閱讀成長很好的指標，可是 ORF 卻無法有效測量閱讀理解部分（Deno, 1985；Fuchs, Fuchs & Maxwell, 1988；Shinn *et al.*, 1992），因此，Fuchs 與 Fuchs（1992）開始尋找替代的閱讀監控評量，以回答問題、回想、克漏字測驗及選字測驗四項為研究內容，研究結果顯示 Maze 的成績是回想和克漏字的兩倍，且教師對 Maze 的滿意度高，而 Maze 與 ORF 在許多方面的成績具有一致性，可作為評量閱讀的指標之一。

Espin, Deno, Maruyama 和 Cohen（1989）曾研究 163 個三到五年級學生在 Maze 與 ORF 表現的相關，研究結果指出一分鐘朗讀的正確字數與選字測驗得分之相關為.85，在每個年級間的相關介於.78 至.86，顯示這兩種測量方式的相關性很高。Fuchs 和 Fuchs（1990），測量 153 個二到三年級普通與障礙學生，Maze 與 ORF 的相關為.83，Maze 與史丹福成就測驗的相關為.77。Jenkins 和 Jewell（1993）將研究擴

展至測量一年級學生 Maze 與 ORF、閱讀能力測驗 (Gates MacGinitie Reading Tests) 與成就測驗 (Metropolitan Achievement Tests) 間的相關, 研究對象為兩所國小共 326 個學童, 研究結果顯示 Maze 與 ORF 的相關為 .80; Maze 與 Gates MacGinitie 閱讀測驗的相關為 .85; Maze 與 Metropolitan 成就測驗的相關為 .80, 本結果顯示 ORF 與 Maze 皆有很高的信度與效度。

Faykus 和 McCurdy (1998) 研究 6 個伴隨情緒問題的嚴重閱讀障礙者, 使用 Maze 來預測教學的有效性, 利用 ORF 與 Maze 兩個方式來比較學生的進步情形。結果顯示兩者皆能評量學生的閱讀進步, 但 ORF 是評量閱讀進步較有效的方法, 也是學生較喜歡的方式; 而 Maze 較省時, 教師願意推薦給其他教師使用。

Shin、Deno 與 Espin (2000) 研究使用 Maze 來預測學生閱讀能力的成長, 有 43 個二年級學生參與研究, 結果顯示選字測驗除了可以測出團體的進步, 也可以反映出學生一學年的成長及個人成長速率的差異。研究的結果支持 Maze 是評量閱讀成長之有效、可信的測量方法。

由以上的研究結果說明 ORF 和 Maze 是 CBM 當中最優的兩種閱讀測量方式, 能快速、有效地預測閱讀理解能力, 支持本研究選擇 ORF 及 Maze 作為中文 CBM 閱讀測量之方法, 以瞭解 ORF 與 Maze 在中文語系上是否也能有高度的信度與效度。

三、國內課程本位閱讀測量之研究

相較於國外豐富的研究, 中文課程本位測量研究僅有以下幾篇: 王慧豐、陸正威 (民 90) 研究國小資源班數學科解決問題課程本位評量之應用; 黃瑞珍、黃玉凡 (民 90) 研究國小五年級 CBM 寫作測驗顯著性指標; 葉靖雲 (民 82) 研究課程本位閱讀測驗的效度; 葉靖雲 (民 83) 研究課程本位中、英文閱讀與英文書寫表達測驗的效度; 葉靖雲 (民 84) 研究課程本位

閱讀理解測驗的效度; 葉靖雲 (民 85) 研究三種課程本位數學評量模式的效度; 葉靖雲 (民 87) 再次研究課程本位閱讀測驗的效度; 葉靖雲 (民 89) 研究國中英文讀寫與內容科目表現流暢性指標的效度等, 而其中針對國小階段閱讀方面之研究僅有葉靖雲所做的幾篇研究。

葉靖雲 (民 82) 以國小 152 學生為樣本, 研究 CBM 閱讀測驗四種測驗形式: 「口語朗讀」、「故事重述」、「問題回答」和「填字」測驗之效度, 並以標準化測驗分數、學期成績以及級任教師之判斷作為學生實際閱讀水準的效標, 結果顯示各種測驗之計分方式以正確表現的分數最能反映學生的閱讀能力, 而在小學高年級填字的效標效度顯著高於其他三種測驗。作者認為填字法和口語朗讀為適當的中文 CBM 閱讀測量的方法。

葉靖雲 (民 87) 研究建立流暢性課程本位測量在閱讀領域的實證資料, 共分為兩個階段, 除了在民國 84 年研究「字詞釋義」、「字詞朗讀」、「文本朗讀」、「克漏字測驗」、「故事重述」、「問題回答」與「畫分有意義段落」等七種課程本位閱讀測量的效標關聯效度。由上述七種方法找出效度較佳的測量方式, 然後再比較閱讀障礙學生與普通閱讀能力學生在這些測量上成長速率之差異。研究結果顯示: 二、四、六年級皆以字詞朗讀、文本朗讀與克漏字測驗之效標關聯效度最佳, 且有隨年級增加而逐漸提昇之趨勢。

由上述中文之初期研究, 再由豐富之英語研究結果可以看出 ORF 與 Maze 是 CBM 當中最優的兩種閱讀測量方式, 能快速、有效地預測閱讀理解能力, 支持本研究選擇 ORF 及 Maze 作為中文 CBM 閱讀測量之方法。此外, 由於九年一貫的推行、教科書的開放, 民間許多出版社相繼編製了教科書, 而每個學校、年級所使用的教科書版本不盡相同, 由原本單一國編版教科書增加為非常多種不同書局所編的教科

書可供選擇，因此研究者想瞭解以常用字彙來規範的各版本教科書，是否能夠在使用課程本位閱讀測量時輪替使用，而不只限於目前使用的版本。本研究以上述中文研究為基礎，再次驗證 ORF 與 Maze 在中文語系上是否也能有高度的信度與效度，並編製一份中文版課程本位閱讀測量施測及計分手冊，以供教師實際使用的參考。

研究方法

一、研究對象

本研究以九十一學年度就讀於台北市中正區兩所國小的二年級全體 420 位學生為研究對象，兩校皆使用康軒版做為國語科教材。由於本研究的需要，選取學生的方法如下：

(一) 兩校二年級學生共 420 名，施以「國語文成就測驗」(洪碧霞、邱上真，民 88)。

(二) 由測驗結果，選出分數最高的 36 名學生為高閱讀能力組、分數最低的 36 名學生為低閱讀能力組及最中間的 36 名學生為中閱讀能力組，108 個學生中有 51 個男生，57 個女生。研究對象排除視障、聽障、智障學生。

二、研究工具

(一) 國語文成就測驗(洪碧霞、邱上真，民 88)

國語文成就測驗係一評量學生識字及閱讀理解的測驗，每年級的測驗分為甲、乙兩式，每式中含有聽力及書面測驗兩部分，本測驗的內部一致性係數在 0.86-0.95 之間，二年級題庫試題平均通過率為 0.74，識字與理解能力的相關為 0.62，與學校國語成績的相關為 0.62。本測驗以團體方式進行，適用年齡為國小一到六年級，每個年級有不同的題本。

本測驗是基於閱讀成分編製，其題數、難度適中，因此選擇本測驗二年級之書面測驗作為評量兒童閱讀能力之工具。

(二) 自編課程本位閱讀測量指導手冊

由研究者自編課程本位閱讀測量方法，測驗內容選自國小二年級國語科教科書。由於國內小學之教科書有七種不同版本，為了瞭解同年級不同版本的國語科教科書使用 CBM 來評量學生能力時是否有一致性，本研究使用系統抽樣的方式，根據台北市學校聯絡箱號碼每隔 5 號選出一所學校，共抽出 30 所學校，調查台北市國民小學二年級語文科使用版本的情形後，發現使用康軒版的有二十所，其次為翰林版共六所，因此選擇使用率最普遍的康軒及翰林兩種版本作為研究內容，來驗證不同版本間的差異。研究者根據康軒出版社及翰林出版社出版之國小二年級下學期尚未學過之教材，作為課程本位閱讀測量之材料。測量方法分為兩種：

1. 口語朗讀流暢度 (ORF)

選取康軒版及翰林版國小二年級下學期的文本各四篇為 ORF 之材料，共八篇。編製程序是由教材當中選出一篇不是詩、對話或劇本的課文；將課文選出後，配合國語課本編排的方式，將課文中的注音、插圖去掉，重新以直式由上而下、由右至左、標楷體 24 號字的方式繕打呈現。

ORF 的測驗方式是請學生由第一個字開始大聲朗讀，計算學生每篇文本在一分鐘時間內唸的總正確字數(總字數減總錯誤字數)，本測驗以個別方式進行。

2. 選字測驗 (Maze)

中文的字又稱為一個語素，是構詞的最小單位(程祥徵、田小琳，民 82)，張顯達(2000)引用了湯廷池(1992)文獻中提及對詞的界定方法，認為詞在中文不易界定，尤其是中文複合詞與詞組的界限有相當程度的不確定性與模糊性。呂芳菁(民 88)、程祥徵、田小琳(民 82)比較字與詞之差別，「字」指的是一個個的漢字，是書寫形式的概念，未必具有意義，如：

日、月、風、雨、矛等；「詞」是最小的、能夠自由運用的形體和意義都固定化的造句單位，如：日、月、快樂、婀娜、直昇機、阿斯匹靈、狐假虎威、引狼入室等。比較字詞間的不同發現：「字」在中文的構詞中是一個最易、最方便計算的單位，它或許無法具有意義，但是比起中文詞的劃分相當困難且複雜，而各語言學家間有歧異的情形下，因此研究者建議以「字」為編製 Maze 的單位，以免增添困擾。而且由於中文文字的特性，學生的錯誤常是由於相似字與同音異形字，因此 Maze 編製的方式編題方式與國外的標準較為不同，干擾選項編題之條件除了不可以符合上下文文意之外，可以選擇字形相似字、同音異形字、字音相似字等作為干擾之選項，因為加入了干擾選項，所以雖然只去掉雙字詞中的一字，如「寄□」，學生仍需要依據前後文的意思才能選出正確答案。例如：「你媽媽有事出（來夫去）了，她把鑰匙寄（信放養）在我家。」。

本研究由康軒版及翰林版國小二下教材中另外選取共八篇文本為 Maze 之材料，編製程序是除了首句與末句之外，從第二句起，每間隔七個字刪除一字，改以空格代替，在空格中加入三個選項，要學生圈選出其中一個正確的答案，這三個答案中只有一個是符合文意的正確答案，正確答案以隨機的順序排放於三個選項中。經由上述編製程序，學生一篇文本在兩分鐘內答對的總數即是 Maze 的分數，本測驗以團體方式進行。

另外，為了進一步提高指導手冊的信度，請 30 位參加教育部溝通障礙學分班的資深國小特教教師，在經過一小時之計分說明及訓練之後，對三個樣本的朗讀過程進行計分，每個樣本錄音檔皆播放三次，經檢定結果，Kendall 和諧係數 $W = 1.000$ ，已達到 .001 顯著水準，顯示計分者間有高度的信度，再由計分者之結果及計分過程的建議修改指導手冊使信度更

高。課程本位閱讀測量指導手冊（參見附錄一），手冊中說明「口語朗讀流暢度」與「選字測驗」的施測程序及計分原則，可供教師參考。

三、資料蒐集

本研究之研究工具有「國語文成就測驗」、「口語朗讀流暢度」、「選字測驗」三種。

1. 國語文成就測驗：以團體施測方式施測「二年級國語文成就測驗」，一次測驗一個班級，每次團測約需 40-50 分鐘（一節課）。並根據國語文成就測驗的結果選出閱讀能力高、中、低三組學生各 36 人。

2. 課程本位閱讀測量：根據國語文成就測驗的結果選出閱讀能力高、中、低三組學生各 36 人，得到家長同意之後，以個別方式測量 ORF 並錄音，將康軒版文本按照課數排列（第四課、七課、九課、動態閱讀）及翰林版文本按照課數排列（第二課、四課、六課、十一課），之後再將康軒版與翰林版文本交錯排列，一次共施測八篇文本，每篇文本朗讀一分鐘，教師於學生朗讀時同時以教師計分表記錄，再於全部測驗完之後，重新聆聽錄音檔計分。

以團體方式施測選字測驗，選字測驗亦如上述排列方式，將康軒版及翰林版的文本交錯排列，一次施測八篇文本，每篇文本作答時間為兩分鐘。

在施測四週後，隨機抽取高、中、低三組各 10 名學生，用相同的 ORF 文本及 Maze 各四篇，再以同樣的施測程序進行重測。

四、資料處理與分析

（一）計分者的訓練

計分者為一位受過計分訓練之特教教師，研究者清楚列出計分方法，並講解計分原則。計分人員練習許多例子，在熟悉計分方法後，研究者將一份 Maze 結果及 ORF 之錄音帶交由計分員評分，最後研究者與計分員比較結果，正確率達 .99 為合格之計分員。

(二) 計分者間信度

從總數 108 個樣本中抽出 12 位作為信度考驗樣本，將計分者的計分結果與研究者計分結果核對，以確認計分者間信度的一致性。

(三) 資料分析

本研究主要使用的統計方法說明如下：

1. 在效度方面，以 Pearson 積差相關來考驗 ORF 與 Maze 之效標關聯效度、構念效度；以 Cronbach α 考驗內部一致性；以單因子變異數分析及 Ω 指數來考驗團體差異分析，並以 Scheff'e 法進行事後多重比較。

2. 在信度方面，以 Pearson 積差相關來考驗重測信度、複本信度；以 Pearson 積差相關及 Kendall 和諧係數考驗計分者間一致性。

研究結果與討論

本研究結果依據研究目的，將依序討論 ORF 與 Maze 之信度與效度。本研究討論效標關聯效度與構念效度，將分析 ORF、Maze 與效標國語文成就測驗間的關係、構念效度中團體差異分析及內部一致性分析；在信度方面，本研究討論計分者間一致性信度、複本信度與重測信度分析，將分述如下：

一、ORF 與 Maze 之效度分析

ORF 所選取的文本分別為康軒版二下(1)第四課「烏鴉喝水」、(2)第七課「母親節日記」、(3)第九課「種子的旅行」及(4)動態閱讀「小書迷來演戲」四課。每課的總字數分別為 251 字、256 字、208 字及 222 字。由表一顯示 ORF 的各篇平均數介於每分鐘朗讀 119.84 至 133.43 字，標準差介於 47.27 至 53.43，四篇的總平均數為 128.52，標準差為 50.19。Maze 所選取的文本分別為康軒版二下(1)第三課「嗨！你好」、(2)第五課「黃狗生蛋」、(3)第六課「千人

糕」及(4)第十二課「張媽媽的鑰匙盒」四課，依據每課課文的長度不同，每篇的總分分別為：25 分、36 分、32 分、38 分。表一顯示測驗結果各篇平均數介於 16.81 至 19.25 分，標準差介 6.53 至 7.25，四篇的總平均數為 18.02，標準差為 6.50。

(一) 效標關聯效度分析

為了瞭解 ORF 與 Maze 評量閱讀能力的效度，本研究以四篇康軒版本與效標國語文成就測驗結果進行相關分析，由表二顯示，ORF 與國語文成就測驗有高度的相關，相關係數為 .745，達到 .001 顯著水準；Maze 與國語文成就測驗亦有高度的相關，相關係數為 .724，達到 .001 顯著水準，此項結果說明 ORF 與 Maze 兩者都是相當有效度的測量方法，與國外的相關研究之結果相似 (Deno, 1985; Deno, Markin & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Jenkins & Jewell, 1993; Shinn *et al.*, 1992)。

(二) 構念效度分析

1. 口語朗讀流暢度與選字測驗之內部一致性分析

由表三所示 ORF 各篇分數間的積差相關係數介於 .896 到 .951，皆達 .001 的顯著水準。ORF 之 Cronbach α 值為 .982，內部一致性 α 值越高，表示測量整體的內部一致性越高，說明同一位學生即使朗讀不同的文本，其閱讀表現仍大致相同，不因文本的不同而影響或導致其口語朗讀流暢度有太大的差異，由此可以說明 ORF 是個很穩定的測量方式，同一個學生在朗讀不同文本時，能有穩定的表現，亦即有極佳之構念效度。

表四說明選字測驗各篇分數間的積差相關係數介於 .748 到 .843，皆達 .001 的顯著水準。Maze 之 Cronbach α 值為 .942，其內部一致性信度極高。

表一 學生在康軒版 ORF 字數與 Maze 得分之描述統計

	人 數	ORF		Maze	
		平均字數	標 準 差	平均數／總分	標 準 差
第一篇	108	133.43	53.43	16.81／25	6.53
第二篇	108	127.72	47.27	16.97／36	7.25
第三篇	108	119.84	53.75	19.25／32	7.20
第四篇	108	133.10	51.68	19.04／38	7.16
四篇平均	108	128.52	50.19	18.02／32	6.50

表二 ORF、Maze 與國語文成就測驗之相關矩陣

	國語文成就測驗	ORF	Maze
國語文成就測驗	—		
ORF	.745***	—	
Maze	.724***	.848***	—

*** p < .001

表三 ORF 四篇得分之相關矩陣

	第 一 篇	第 二 篇	第 三 篇	第 四 篇
第一篇	—			
第二篇	.932***	—		
第三篇	.896***	.939***	—	
第四篇	.925***	.951***	.949***	—

*** p < .001

表四 Maze 四篇得分之相關矩陣

	第 一 篇	第 二 篇	第 三 篇	第 四 篇
第一篇	—			
第二篇	.786***	—		
第三篇	.790***	.837***	—	
第四篇	.748***	.827***	.843***	—

*** p < .001

2. ORF 與 Maze 之團體差異分析

由表五顯示，ORF 在低、中、高三組學生之平均數分別是 81.11、126.20、178.26 字，標準差為 27.84、32.77、31.45 字，再由單因子變異數分析結果說明 ORF 之測量方法能有效區辨不同閱讀能力水準（高、中、低）兒童。

在 Maze 方面低、中、高三組學生平均數分別是 12.32、17.41、24.32 分，標準差為 4.38、3.51、4.81，單因子變異數分析結果說明 Maze 能有效區辨不同閱讀能力水準（高、中、低）兒童，亦即閱讀能力高組學生，在每 7 個字空格中選填正確字之答對數顯著高於中組與低

組。

若是探討 ORF 與 Maze 效度之比較，依 Faykus 和 McCurdy (1998) 研究 ORF 與 Maze 的效度，研究結果指出兩種測驗方式皆能評量學生閱讀進步情形，而 ORF 的效果較 Maze 為佳，本研究結果與前述國外的研究結果相當一致。表五中的 F 值及 Ω 指數再次顯示 ORF 的效度高於 Maze，無論在 F 值或 Ω 指數部分，皆是 ORF 對學生有較高的區辨力。由以上的說明可以知道 ORF 的效度較高，若是時間允許，教師可以 ORF 為主來評量學生閱讀能力，再輔以 Maze 來監控學生的表現。

若是再經由 Scheff'e 法進行事後比較，亦顯示三組閱讀能力兒童之間有顯著差異，ORF 與 Maze 結果皆為高組 > 中組 > 低組，亦即 ORF 及 Maze 具有優異的構念效度。

除此，由表六顯示的信賴區間來看成績分佈，三組學生在 ORF 的得分並未重疊，很明顯地看出三組學生有不同的表現，低組學生朗讀的字數為 71-90 個字之間，中組學生朗讀的字數為 115-137 個字之間，高組的學生朗讀的字數為 167-188 個字之間，顯示閱讀能力越高的學生，每分鐘朗讀的正確字數越多。同樣地，Maze 的結果亦相同，三組閱讀能力兒童的表現有明顯的區隔，沒有重疊，更顯示 ORF 與 Maze 對不同程度學生有很高的區辨力。

4. ORF 與 Maze 表現之相關分析

本研究根據 108 位學生在康軒版 ORF 與 Maze 的得分，ORF 與 Maze 之相關係數為 .848，達到 .001 之顯著水準；其內部一致性 Cronbach α 值為 .894，標準化的 α 係數值為 .971，由於 ORF 與 Maze 之測量單位不相同，為了避免因測量單位不同造成變異量的不同，因此應採取標準化係數較為適當，所以內部一致性係數為 .971。由此結果顯示同一位學生有好的 ORF

也會有好的 Maze，相對地，ORF 不佳的 Maze 也相對不佳，這種互為高度相關的結果，表示這兩種測量方法均是非常好的閱讀評量方法。

本研究結果與許多研究的結論相似，Espin 等人 (1989) 曾研究 163 個三到五年級學生在 Maze 與 ORF 表現的相關，研究結果指出 ORF 與 Maze 得分之相關為 .848，在每個年級間的相關介於 .78 至 .86；Fuchs 和 Fuchs (1990) 的研究結果顯示 Maze 與 ORF 的相關為 .83，Maze 與史丹福成就測驗的相關為 .77；Jenkins 和 Jewell (1993) 的研究結果顯示 Maze 與 ORF 的相關為 .80。由以上的研究可以發現，本研究與國外的研究得到相似的結果，Maze 與 ORF 的相關都在 .80 以上，說明同一位學生即使用這兩種不同的測驗方式，其閱讀表現仍相同。

由以上的效標關聯效度及構念效度分析，說明 ORF 與 Maze 具有極高的效度，不僅能夠高度預測學生目前的閱讀能力，也能區辨不同閱讀能力兒童，因此十分值得推薦給國小教師使用。

二、ORF 與 Maze 之信度分析

(一) 計分者間一致性分析

本研究藉由明確之施測程序、詳細的計分說明及練習等方式來提高計分者間的信度。從總數 108 個樣本中隨機抽取高、中、低三組各 12 名學生，共 36 位作計分者間一致性信度之考驗，將研究者計分結果與另一計分者的計分結果核對，以確認計分者間信度的一致性。由表七結果顯示，在 ORF 部分，播放樣本朗讀之錄音檔讓研究者及計分者計分，計分結果顯示兩個計分者間相關係數為 .98，達 .001 的顯著水準，顯示計分者之間的計分上相當一致，有良好的信度。Maze 為選項式，因此計分結果則完全相同。

表五 三組閱讀能力學童在 ORF 及 Maze 的平均值（標準差）之比較

	低 (N=36)	中 (N=36)	高 (N=36)	F	Scheff'e 考驗	Ω
ORF	81.11(27.84)	126.20(32.77)	178.26(31.45)	89.94***	高>中>低	.63
Maze	12.32(4.38)	17.41(3.51)	24.32(4.81)	71.64***	高>中>低	.58

表六 三組閱讀能力學童在 ORF 及 Maze 的信賴區間比較

	低 (N=36)	中 (N=36)	高 (N=36)	全體 (N=108)
ORF	71.69-90.53	115.1-137.29	167.6-188.90	118.9-138.10
Maze	10.84-13.80	16.22-18.60	22.69-25.95	16.78-19.26

表七 計分者間一致性信度描述統計摘要表

	人 數	平均 數	標 準 差	Pearson 相關係數
ORF 研究者計分	36	129.71	49.32	.98***
ORF 計分者計分	36	128.31	48.12	
Maze 研究者計分	36	20.43	6.58	1.00***
Maze 計分者計分	36	20.43	6.58	

*** p<.001

(二) 不同版本（康軒版與翰林版）在 ORF 與 Maze 表現之一致性分析

根據國外的研究，雖然他們有非常多不同版本可供教師選擇，亦有版本不相同的問題，所以他們在選擇文本時，已經有一套統一的考驗標準，以文本的「難易度」作為選擇的標準。在考驗難易度時，通常使用 Spache (1953) 訂定的閱讀指數 (readability score) 規則、Dale-Chall (1948) 指數及 Fry Readability Index (1968) 來計算文本的難易度，並依此選擇需要的文本。我國過去幾十年都是使用國立編譯館主編的課本，直到近幾年才開放教科書多元化，因此沒有一個文本難易度的計算標準。雖然教育部制訂了「九年一貫課程綱要」，將國語文教材編選原則以常用字的識字量加以規範，但各個版本之間所編輯出來的課程，是否在同一個年級有相同的難易度，並無相關的研

究及驗證，因此，針對不同版本教科書的一致性考驗是非常重要的。研究者想進一步瞭解以常用字彙來規範的各版本教科書，是否能夠在使用課程本位閱讀測量時輪替使用，使教師在評量學生閱讀能力時可有不同的選擇，並能讓接受不同版本教學的學生皆能適用此套評量方式，而不只限於目前使用的版本，因此進一步檢證 ORF 與 Maze 使用不同版本學生是否有一致性的表現。

依表八顯示學生在使用康軒版及翰林版 ORF 之相關係數為 .983，達到 .001 的顯著水準，此研究結果說明兩種版本之 ORF 有顯著相關；康軒版及翰林版 Maze 之相關係數為 .936，達到 .001 的顯著水準。這兩個版本可視為內容相似的複本，在相同的受試者及施測程序下，所求出的複本信度很高，顯示康軒及翰林兩個版本學生之表現具有一致性。

表八 不同版本課程本位閱讀測量之相關係數摘要

	人 數	Pearson 相關係數
康軒 ORF-翰林 ORF	108	.983***
康軒 Maze-翰林 Maze	108	.936***

*** $p < .001$

表九 重測信度描述統計摘要表

	第一次測驗 平均數 (標準差)	第二次測驗 平均數 (標準差)	Pearson 相關係數
ORF	128.52 (50.19)	148.64 (44.95)	.918***
Maze	18.02 (6.50)	21.66 (8.36)	.884***

*** $p < .001$

(三) 重測信度分析

在施測四週後，隨機抽取高、中、低三組各 10 名學生，用相同的文本、施測程序進行重測信度考驗。由表九顯示，ORF 重測之平均字數為 148.64，標準差為 44.95，其重測信度係數為 .918，達到 .001 的顯著水準。Maze 重測之平均數為 21.66，標準差為 8.36，其重測信度係數為 .884，達到 .001 的顯著水準。

此外，四週以後的重測結果與四週前的結果相較，一分鐘口語朗讀流暢度的正確字數由 128.52 個字提升至 148.64 個字；Maze 的分數亦由 18.02 分上升至 21.66 分，顯示樣本經由四週的教學之後，閱讀能力的提升。本研究的主要目的雖然並未監控學生的閱讀進步情形，然而從許多國外的研究中顯示 CBM 可以監控學生的進步情形，同時亦可評鑑教師的教學績效，Shinn 等人 (1989) 研究使用不同難度教材的課程本位評量對學生進步情形測量，結果顯示兩種難度教材皆能有效評量學生的進步；Faykus 和 McCurdy (1998) 研究 ORF 與 Maze 評量嚴重閱讀障礙學生的進步情形，結果顯示兩者皆能有效監控學生的閱讀成長狀況，而 ORF 比 Maze 的效果更好。雖然本研究的目的並未包含學生進步情形的監控，但經由四週後

的重測發現學生在 ORF 與 Maze 兩者皆有明顯的進步，顯示 ORF 與 Maze 兩者是動態的評量，可以在短時間內監控學生的進步情形，教師經由學生進步的狀況，來決定目前的教學方法是否合適。如果學生的表現未達目標，教師則應調整教學方式或修改學年目標，因此可以由學生的進步情形進一步評鑑教師的教學績效。

結論與建議

一、結論

本研究目的是以目前已具有完整規模的英語「課程本位測量」方式，嘗試研究中文課程本位閱讀測量方法之信度與效度，編製一份指導及計分手冊，並考驗九年一貫課程不同版本間之差異，以作為評量學生閱讀能力的依據。

茲將研究結果陳述如下：

(一) ORF 與 Maze 具有良好的效度

420 名普通班學生在 ORF 與 Maze 上的表現與效標「國語文成就測驗」間有很顯著的相關，顯示 ORF、Maze 具有良好的效標關聯效度。

ORF 與 Maze 能有效區辨不同閱讀能力水

準（高、中、低）兒童，且學生在相同版本，四篇不同文本，不論在 ORF、Maze 之表現，均達到內部一致性顯著相關。此外，同一位學生在不同文本的測量時，其表現仍大致相同，不因文本的不同而影響或導致其在四篇文本的作答結果有太大的差異；同一學生在兩種測量方式（ORF 與 Maze）之表現亦呈現顯著相關，此結果顯示 ORF 與 Maze 具有良好之構念效度。

（二）ORF 與 Maze 具有良好之信度

學生在康軒版與翰林版之 ORF 與 Maze 之表現上一致性很高，亦即 ORF 與 Maze 具有良好之複本信度。此說明教師在用 CBM 時可以不受版本限制，亦即教學可以某一版本為主，但是學生若是轉學或是因為需要而變課程內容時，或是選擇降低課程難度時，只要紀錄起點能力，選擇合適之難度課程，均可以一邊教學一邊記錄監控學生學生的表現，不受課程版本的限制。

學生經由四週後之重測時，在 ORF 與 Maze 整體之表現上均呈現進步情形，並達到顯著相關，此結果顯示 ORF 與 Maze 兩者均是相當穩定的測量閱讀的方法，亦即學生能力佳的，在相同的施測方式，不因為時間因素而呈現不穩定的表現。

ORF 與 Maze 具有良好之計分者信度，不同計分者使用指導手冊計分結果 ORF 與 Maze 均呈現計分者間一致性顯著相關。計分者一致性信度代表計分手冊是否具有規範性，計分者間一致性越高，表示計分說明清楚、詳細，因此，本測驗之計分者信度相當高，顯示課程本位閱讀測量施測及計分手冊具有良好之規範。

（三）本研究的目的是為提供教師一個簡單、省時、容易施測、方便、具有信度與效度的評量方式，研究結果顯示 ORF 與 Maze 可以有效評量學生之閱讀能力。

二、研究建議

（一）教育上的應用

根據本研究結果，提出下列幾項建議：

1. 根據本研究，發現中文之課程本位閱讀測量有非常高的信度與效度，具有省時、簡便、容易施測與計分等精神，能讓教師在標準化成就測驗之外，有另外一個更好的選擇，其結果除了能初步篩選出低閱讀能力的兒童，以供進一步鑑定外，還能經由每週的重複施測，監控學生的進步情形。此為未來鑑定閱讀障礙兒童一項可行的方法（Stecker,2005），因此，本測量方式實為一個實用的測量方式。

2. 教師可以將學生在課程本位閱讀測量的表現繪成進步曲線圖，並根據學生一分鐘朗讀的字數與選字測驗的得分，擬定學生個別化教育計畫中的學習目標。例如：將閱讀之學期目標訂為「學生的閱讀表現在康軒版第四冊，由一分鐘朗讀 38 字進步至一分鐘朗讀 52 字；選字測驗的得分為 15 分」。

3. 課程本位閱讀測量主要以「速度」及「正確性」兩個向度為主，是評量自動性注意力的兩項重要指標，因此在施測過程中，教師可以很明顯地觀察學生是否能夠專注於朗讀，藉此診斷學生是否有注意力不足的問題。

4. 在 ORF 施測過程中，可以同時觀察學生的學習動機、自信心及問題解決能力，並經由學童的錯誤類型，分析其閱讀的能力。

（二）未來研究上的建議

1. 本研究只探討課程本位測量中之 ORF 與 Maze，研究對象為二年級學童，因此未來可擴充至其他測量方式之信度與效度研究及其他年級之研究。

2. 本研究屬於初步的信效度驗證研究，然而課程本位閱讀測量之目的主要在監控學生閱讀的進步情形，因此，經由本研究之驗證，說明中文課程本位閱讀測量有非常高的信度與效度，未來可以針對學生的閱讀進步情形，建立

一套可行、明確的監控方法。

3. 未來研究可以根據本研究結果，將研究對象擴展至障礙兒童與一般兒童之比較。

4. 課程本位測量提出區域性常模的理念，其中區域性的定義可指城市、某學區、某學校、某一年級或某一班，未來教師可以利用課程本位測量區域性常模的理念，建立班上、班級間、年級間，在短時間就可以快速、簡易地對照出低閱讀能力學童，可以用來作為低閱讀能力學生的篩選與轉介。未來並可以建立全國性的常模，以供教師對照使用。

參考書目

一、中文部分

- 王慧豐、陸正威(民 90): 國小資源班數學科解決問題課程本位評量應用之研究。東台灣特殊教育學報, 3 期, 261-291 頁。
- 王瓊珠(民 90): 台灣地區讀寫障礙研究回顧與展望。國家科學委員會研究彙刊: 人文及社會科學, 11(4), 331-344 頁。
- 邢敏華(民 84): 評介課程本位測量在特殊教育上的應用。特殊教育季刊, 54, 1-6。
- 李瑩均(民 90): 國小寫字困難學童與普通學童寫字相關認知能力之分析研究。國立臺灣師範大學特殊教育學系碩士論文(未出版)。
- 呂芳菁(民 88): 中文的字、詞、詞素語詞組。聽語會刊, 19, 19~24。
- 周台傑(民 86): 國民小學閱讀障礙學生閱讀錯誤類型分析及有效教學策略之研究。國科會專案報告。NSC 87-2413-H-018-002-F10。
- 林寶貴、錡寶香(民 91): 中文閱讀理解測驗。教育部出版。
- 林寶貴、錡寶香(民 89): 中文閱讀理解測驗編製。特殊教育學報, 19 期, 79-104 頁。
- 洪月女譯(民 87) (Kenneth S. Goodman 原著): 談閱讀。台北: 心理。
- 洪碧霞、江秋坪、蕭淳元(民 84): 閱讀障礙篩選工具的發展。行政院國家科學委員會專題研究計畫成果報告(NSC 83-0301-H024-010), 台北: 行政院國家科學委員會。
- 洪碧霞、邱上真(民 86): 國民小學國語文低成就學童篩選工具系列發展之研究。特殊教育研究學刊, 15 期, 83-107 頁。
- 洪碧霞、邱上真(民 88): 國民中小學國語文成就測驗。教育部特教小組編印。
- 柯華葳(民 88): 閱讀理解困難篩選測驗施測說明。教育部特殊教育工作小組。
- 郭生玉(民 93): 教育測驗與評量。台北: 精華書局。
- 教育部(民 93): 特殊教育統計年報—93 年度。台北市: 教育部編印。
- 程祥徵、田小琳(民 82): 現代漢語。台北市, 書林。
- 張顯達(民 89): 平均語句長度在中文的應用。聽語會刊, 39, 36-47。
- 黃秀霜(民 88): 中文識字量表之編製及國語文低成就學童認字困難之診斷(II)。行政院國家科學委員會專題研究計畫成果報告。台北: 行政院國家科學委員會。
- 黃瑞珍(民 91): 新竹縣身心障礙兒童溝通訓練研習講義。新竹縣特殊教育資源中心。
- 黃瑞珍、黃玉凡(民 90): 課程本位測量寫作測驗之顯著性指標研究。東台灣特殊教育學報, 3, 1-36 頁。
- 湯廷池(民 81): 漢語的「字」、「詞」與「語素」。漢語詞法句法, 3, 1-57。
- 葉靖雲(民 82): 課程本位閱讀測驗的效度研究。特殊教育學報, 8 期, 273-303 頁。
- 葉靖雲(民 83): 課程本位中、英文閱讀與英文書寫表達測驗的效度研究。特殊教育學報, 9 期, 257-287 頁。
- 葉靖雲(民 84): 課程本位閱讀理解測驗的效

- 度研究。國科會專案報告。NSC 84-2421-H-018-002-4401。
- 葉靖雲 (民 85)：三種課程本位數學評量模式的效度研究。*特殊教育學報*，11 期，35-77 頁。
- 葉靖雲 (民 87)：課程本位閱讀測驗的效度研究。*特殊教育與復健學報*，6 期，239-260 頁。
- 葉靖雲 (民 89)：國中英文讀寫與內容科目表現流暢性指標的效度研究。國科會專題研究計畫成果報告，研究編號 NSC-89-2413-H-018-006。
- 葉靖雲 (民 92)：課程本位測量。載於「2003 特殊教育學術研討會」會議手冊。國立臺灣師範大學。
- ## 二、英文部分
- Allinder, R.M, Dunse, L., & Brunken, C.D. (2001). Improving fluency in at-risk readers and students with learning disabilities. *Remedial and Special Education, 22*(1), 48-54.
- Dale, E., & Chall, J. (1948). A formula for predicting readability. *Educational Research Bulletin, 27*, 11-20.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternatives. *Exceptional Children, 52*(3), 219-232.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *Remedial and Special Education, 37*, 184-192.
- Deno, S. L., & Fuchs, L.S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children, 19*(8), 1-16.
- Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J (2001). Using curriculum-based measurements to establish growth standards for students with learning disabilities. *School Psychology Review, 30*(4), 507-524.
- Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49*(1), 36-45.
- Dunn, E. K. & Eckert, T. L. (2002). Curriculum-based measurement in reading: A comparison of similar versus challenging material. *School Psychology Quarterly, 17*(1), 24-34.
- Espin, C., Deno, S. L., Maruyama, G., & Cohen, C. (1989). *The basic academic skills samples: An instrument for the screening and identification of children at risk for failure in regular classrooms*. Paper presented at the annual meeting of the American Educational Research Association.
- Faykus, S. P. & McCurdy, B. L. (1998). Evaluating the sensitivity of the maze as an index of reading proficiency for students who are severely deficient in reading. *Education and Treatment of Children, 21*(1), 1-21.
- Fewster, S. & MacMillan, P. D. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education, 23*(3), 149-156.
- Fry, E. (1968). A readability formula that saves time. *Journal of Reading, 11*, 513-516, 575-578.
- Fry, E. (1968). Graph for estimating readability. *Journal of Reading, 577*.
- Fuchs, L. S., & Deno, S. L (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58*, 232-243.
- Fuchs, L. S., & Fuchs, D. (1986a). Effects of systematic formative evaluation on student achievement: A meta-analysis. *Exceptional*

- Children*, 53, 199-208.
- Fuchs, L. S., & Fuchs, D. (1986b). Curriculum-based assessment of progress toward long- and short-term goals. *Journal of Special Education*, 20, 69-82.
- Fuchs, L. S., & Fuchs, D. (1990). Relations among the reading comprehension subtests of the Stanford Achievement Test and maze, recall, Question-answering, and fluency measures. (Unpublished)
- Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*, 21, 45-58.
- Fuchs, L. S. & Fuchs, D. (1997). Use of curriculum-based measurement in identifying students with disabilities. *Focus on Exceptional Children*, 30(3), 1-16.
- Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on student achievement, pedagogy, and student awareness of learning. *American Educational Research Journal*, 21, 449-460.
- Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989a). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children*, 55, 429-438.
- Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (1989b). Effects of instructional use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education*, 10(2), 43-52.
- Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, 9(2), 20-29.
- Fuchs, L. S., Fuchs, D., & Stecker, P. M. (1989). Effects of curriculum-based measurement on teachers' instructional planning. *Journal of Learning Disabilities*, 22, 51-59.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement: Using a reading maze task. *Exceptional Children*, 58, 436-450.
- Hintze, J. M., Shapiro, E. S., & Lutz, J. G. (1994). The effects of curriculum on the sensitivity of curriculum-based measurement in reading. *The Journal of Special Education*, 28, 188-202.
- Hintze, J. M., Owen, S. V., Shapiro, E. J. & Daly III, E. J. (2000). Generalizability of oral reading fluency measures: Application of G theory to curriculum-based measurement. *School Psychology Quarterly*, 15(1), 52-63.
- Hintze, J. M., Daly III, E. J., Shapiro, E. S. (1998). An investigation of the effects of passage difficulty level on outcomes of oral reading fluency progress monitoring. *School Psychology Review*, 27(3), 433-448.
<http://studentprogress.org>
- Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children*, 59, 421-432.
- Kranzler, J. H., Miller, M. D. & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading. (ERIC Document EJ No.602976)
- Lerner, J. W. (2000). *Learning disabilities: Theories, diagnosis, and teaching strategies* (8th ed.). Boston: Houghton Mifflin.
- Madelaine, A & Wheldall, K. (1999). Curriculum-based

- measurement of reading : a critical review. *International Journal of Disability*, *46*(1), 71-85.
- Marston, D. (1989). Curriculum-based measurement: What is it and why do it? In M. R. Shinn (Ed.), *Curriculum based measurement: Assessing special children* (pp. 18-78). New York: Guilford.
- Marston, D., Mirkin, P., & Deno, S. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *Journal of Special Education*, *18*, 109-117.
- McCurdy, B. L., Shapiro, Edward S. A (1992). Comparison of teacher-, peer-, and self-monitoring with curriculum-based measurement in reading among students with learning disabilities. *Journal of Special Education*, *26*(2), 160-180.
- Mcloughlin, J. A. & Lewis, R. B. (1994). *Assessing special students* (4th ed). Prentice-Hall, Inc.
- Parker, R., Hasbrouck, J. E., & Tindal, G. (1992). The maze as a classroom-based reading measure: Construction methods, reliability, and validity. *The Journal of Special Education*, *26*, 195-218.
- Polloway, E. A. & Patton, J. R. (1996). *Strategies for teaching learners with special need*. (6th ed.) New York: Macmillan Publishing Company.
- Shanker, J. L., & Ekwall, E. E. (1998). *Locating and correcting reading difficulties* (7th ed.). Upper Saddle River, NJ: Merrill.
- Shin, J., Deno, S. L. & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *The Journal of Special Education*, *34*(3), 164-172.
- Shinn, M. R. (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford.
- Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relations to reading. *School Psychology Review*, *21*, 459-479.
- Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York: Guilford.
- Shinn, M. R., Gleason, M. M., & Tindal, G. (1989). Varying the difficulty of testing materials: Implications for curriculum-based measurement. *The Journal of Special Education*, *23*, 223-233.
- Spache, G. (1953). A new readability formula for primary-grade reading materials. *Elementary School Journal*, *53*, 410-413.
- VanDerHeyden, A. M. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review*, *30*(3), 363-383.

附錄：國小課程本位閱讀測量指導手冊

國小課程本位閱讀測量指導手冊

王梅軒 編製

施測注意事項

- 一、本測驗使用對象為國小學童。
- 二、測驗場所最好選擇受試者較熟悉的地方，要空氣流通，光線充足，並避免干擾。施測時，主試者與受試者面對面而坐，桌椅高度要適合。
- 三、主試者在施測過程中，應儘量維持受試者的注意力。若有疲倦、鬧情緒、缺乏動機的狀況，則不宜施測。
- 四、本量表「口語朗讀流暢度」與「選字測驗」皆有時間限制，需用計時器（如馬錶）計時。「口語朗讀流暢度」每篇文本朗讀時間為一分鐘，「選字測驗」每篇文本作答時間為兩分鐘。
- 五、為使測量順利的進行，主試者應熟悉兩種測驗的施測程序、材料及記分標準。

「口語朗讀流暢度」施測說明與計分方法

一、材料：

- （一）朗讀材料
- （二）施測者記錄紙
- （三）碼表
- （四）錄音機或錄影機

二、測驗編製程序

- （一）由教材當中選出一篇文本，文本的選擇原則為文本的內容不可以是詩、對話或劇本。
- （二）將課文選出後，去掉注音、圖片及插畫，並與課文中相同大小及字形呈現，由於中文之國語課本編排的方式為直式加注音、字型約為標楷體 24 號字，因此，將課文中的注音、插圖去掉，重新以直式由上而下、由右至左、標楷體 24 號字的方式繕打呈現。

三、施測說明：

- （一）將朗讀材料呈現在學生面前。
- （二）使用相同備份以為記分，但以不影響學生之情緒及注意力為原則。
- （三）指導語如下：

當我說「開始」時，請開始大聲朗讀，由第一個字開始順序念下去（用手指），試著念每個字。如果有任何看不懂之處，我會告訴你，怎麼念。請盡力而為，努力去念。

- (四) 說「開始」，當學生開始念時，教師隨即按下碼表並全程錄音。若學生不會念文本的第一個字而停頓超過3秒時，教師應停止計時並告知學生正確讀法，再恢復計時，並算一個錯誤。
- (五) 教師可以一邊紀錄學生錯誤之處，以斜線（/）表示。
- (六) 當一個字停頓超過3秒時，教師應即時告知學生正確讀法，並要學生繼續往下讀，算一個錯誤。
- (七) 跳行則算一個錯誤，並指導學生重新念此行。
- (八) 一分鐘到時，說「停止」，並以【】表示結束。

四、計分方法：

計算學生一分鐘內朗讀的總正確字數（Words Read Correctly, WRC）為幾個字，計分方法如下：

(一) 正確字

1. 發音符合上下文文意。
2. 原本念錯，但在三秒內自行糾正成正確發音。
3. 重複字及添加字不計算為錯誤字，但亦不計算在總正確字數中。
4. 方言或一般通俗用語。

(二) 不正確字

1. 發音不符合上下文文意。
2. 替代。
3. 省略。
4. 停頓超過三秒鐘的字。
5. 字序顛倒。
6. 文本中只要念錯的地方，無論錯幾次，就要算幾個錯誤。

(三) 計分範例

1. 發音正確字計算正確字一個。中文有許多破音字，需依上下文用法讀出正確語音。

例如：(1)字裏行間→ㄉㄤ ㄩ ㄉㄨ ㄛ。 WRC=4

(2)可行方法→ㄒㄧㄥ ㄉㄨ ㄉㄨ ㄛ。 WRC=4

(3)各種美味→ㄅㄛ ㄨ ㄉㄨ ㄛ。 WRC=4

(4)種田的人→ㄅㄛ ㄨ ㄉㄨ ㄛ。 WRC=4

（這裡請用標楷體，否則聲調符號無法顯示）

2. 在三秒內自行糾正發音的字，以正確字計算。例如：

(1)卻沒有蕭條的氣息。

→卻沒有蕭（ㄉㄨ ㄨ ㄛ）條……（2秒）……蕭（ㄉㄨ ㄨ ㄛ）條的氣息。 WRC=8

(2)這正是所謂的天涯若比（ㄉㄨ ㄨ ㄛ）鄰。

→這正是所謂的天涯若比（ㄉㄨ ㄨ ㄛ）……比（ㄉㄨ ㄨ ㄛ）鄰。 WRC=11

3. 重複字及添加不算錯，而重複及添加的字數也不算在總字數中。

例如：(1)距離遙遠的兩地方→距離遙遠的兩個地方（添加一字不算錯） WRC=8

(2)科技的進步和距離間的拉近（重複不算錯）

→科技的進步和距離……距離間的拉近 WRC=12

3. 方言或一般通俗用語以正確字計算。

例如：(1) 垃圾 (ㄉㄞ ㄐㄞ)。 $WRC=2$

(2) 媽媽 (ㄇㄞ ㄇㄞ) → 媽媽 (ㄇㄞ ㄇㄞ) 或 (ㄇㄞˊ ㄇㄞˊ) $WRC=2$

(四) 不正確字

1. 念錯字或不符合上下文文意者以不正確字計算。

例如：(1) 天涯 (ㄧㄞˊ) 若比 (ㄅㄞˊ) 鄰。

→ 天涯 (ㄧㄞˊ) 若比 (ㄅㄞˊ) 鄰。(念錯二字) $WRC=3$

(2) 等到一覺 (ㄉㄞˊ) 醒來 → 等到一覺 (ㄉㄞˊ) 醒來 $WRC=5$

2. 替代字以不正確字計算。

例如：(1) 說話缺乏誠意。→ 說話欠缺誠意 (替代 2 字) $WRC=4$

(2) 在一百年之前。→ 在一百年之久 (替代 1 字) $WRC=5$

3. 省略字以不正確計算。

例如：記得我讀一年級的時候。→ 記得我一年級時候 (省略 2 字) $WRC=8$

4. 停頓超過 3 秒鐘時，應提示正確字，並計不正確字一個。

例如：(1) 天涯若比鄰。→ 天…… (3 秒) …… (施測者提示 ㄧㄞˊ) 涯若比鄰。 $WRC=4$

5. 字序顛倒均以不正確字計算。

例如：1. 到了秋天，楓葉紅，菊花黃。

→ 到了秋天，紅楓葉，黃菊花 (顛倒 6 字) $WRC=4$

2. 文化的交流和融合。→ 文化的交流和合融。(顛倒 2 字) $WRC=6$

6. 文本中只要念錯的地方，無論錯幾次，就要算幾個錯誤。

例如：1. 大烏鴉和小烏鴉終於開開心心的喝到水了。

→ 大烏鴉和小烏鴉終於開開心心的喝到水了。 $WRC=16$

2. 垃圾本來沒有家，有了垃圾桶，垃圾才有家。

→ 垃圾 (ㄉㄞˊ) 本來沒有家，有了垃圾 (ㄉㄞˊ) 桶，垃圾 (ㄉㄞˊ) 才有家。 $WRC=14$

參考資料：

黃瑞珍 (民 91)：新竹縣身心障礙兒童溝通訓練研習講義。新竹縣特殊教育資源中心。

Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children.* (p.239-240). New York: Guilford.

「選字測驗」施測說明及計分方法

一、材料：

- (一) 選字測驗題本（共有八篇文本）
- (二) 碼錶

二、選字測驗編製程序：

由教材中選擇一篇內容不是詩、對話或劇本文本。

將文本中每間隔七個字刪除一字，改以空格代替，在空格中加入三個選項讓兒童圈出正確的答案。若每間隔七個字剛好遇到虛字或無法經由上下文看出答案的字時，則跳過此字不刪除，並繼續往下第七個字才再刪除一字。

三個答案中只有一個是符合文意的正確答案，正確答案以隨機的順序排放於三個選項中。

兩個干擾選項可依據上下文選出兒童在選詞時較容易犯的錯誤編選，例如：字形相似字、同音異形字、字音相似字等作為干擾之選項。

將課文中的注音、插圖去掉，重新以直式由上而下、由右至左、標楷體 24 號字的方式繕打呈現。

三、施測說明：

- (一) 將選字測驗題本發給學生。
- (二) 指導語如下：

當我說「開始」時，請由第一個字開始閱讀，依照文本的意思選出括號中符合文意的字，並將正確的答案圈起來，請盡力而為，努力作答，當我說「停止」時，請將筆放下，停止作答。作答的時間為兩分鐘。

- (三) 練習例題，確定所有學生皆瞭解測驗的方式後，開始正式施測。
- (四) 說「開始」後教師隨即按下碼表，並於兩分鐘時說「停止」。
- (五) 將八篇文本以相同方式施測完後，收回全班選字測驗題本，測驗結束。

四、計分方法：

選字測驗之計分方式為兩分鐘內答對的總數，答對一個算一分。

Bulletin of Special Education, 2005, 29, 73-94
National Taiwan Normal University, Taiwan, R.O.C.

The Reliability and Validity of Curriculum-Based Reading Measures on Elementary School Students

Mei-Hsuan Wang

Taipei Dun-Hwa Elementary School

Rei-Jane Huang

Taipei Municipal Teachers' College

ABSTRACT

The purpose of this study was to examine the reliability and validity of curriculum-based reading measurement. We used Chinese Academic Achievement Test to sample 108 students from 420 second-grade elementary students, whose reading abilities are high, medium, and low. Then apply curriculum-based reading measurement which include Oral Reading Fluency (ORF) and Maze Task (Maze) to those 108 sampling students. The main findings of this research were stated as follows: 1. There are significant correlation among ORF, Maze and the Chinese Academic Achievement Test. 2. ORF and Maze can effectively distinguish three levels of reading capability. Therefore, this supports that ORF and Maze present good construct validity. 3. The results of ORF and Maze from different versions of textbooks are highly consistent. 4. After test-retest reliability, alternate-forms reliability, and scorer reliability, the result reveals that ORF and Maze are reliable measurements. 5. The results of ORF and Maze from different raters are highly consistent.

The goal of the study is to provide teachers with a simple, efficient, easy-to-adopt and credible measurement. The result indicates that elementary school teachers can utilize ORF and Maze to evaluate students' reading ability.

Keyword: Curriculum-Based Measurement (CBM), reading ability, Oral Reading Fluency, Maze Task