

閱讀理解綜合評核系統的建構

薛寶嫦

澳門大學教育學院

本文首先闡釋教學與評核的緊密關係，帶出欲建構的萌發式故事閱讀理解綜合評核系統的重要性。接著，以Feldman（1994）的非普遍性理論為基礎，配合Palincsar & Brown（1984）的交互教學法為閱讀指導手段，運用故事書和故事板為評核工具，建構適用於評核幼兒萌發式故事閱讀理解的概念模型。當考慮綜合評核系統的原理後，再按概念模型發展出綜合評核系統實施架構，期望以此架構指導閱讀教學實驗。最後建議利用Messick（1989）的統一建構效度理論，闡釋評鑑系統信度和效度的程序和方法。

關鍵詞：萌發式故事閱讀、綜合評核系統、統一建構效度理論

萌發式故事教學與評核釋要

教學與評核的關係

近年來，「對學習的評核」(assessment of learning)和「促進學習的評核」(assessment for learning)是教育界熱切討論的議題(參考 Assessment Reform Group, 2002)。「對學習的評核」具有績效的功能，用以評核學習者在完成一件特定任務後的表現水平，或是教與學的達成程度，評核的目的通常用於報告教學結果。「促進學習的評核」則在教與學的過程中發生，評核活動所得的資訊能用於教與學過程的改進。這兩種評核取向各有優缺點，如果加以整合，組成一個兼具兩者優點的綜合評核系統(Integrated Assessment System, IAS)(Birenbaum et al., 2006)，這樣不但能將教學與評核緊密結合，還能同時兼顧總結性評核和形成性評核的功用。

為此，研究員期望建構一個適用於評核幼兒萌發式閱讀理解的綜合評核系統，評核目的並不是將幼兒分類為不同的閱讀水平，而是讓老師更能掌握幼兒現有的發展情況，以便因材施教(關於評核幼兒閱讀理解及閱讀理解策略，可參閱周兢，2001；陳淑琴，2001；黃瑞琴，1997；Flippo, 2003; Mariotti, 2005; McGee & Richgels, 2004; Paris & Stahl, 2005; Sarroub & Pearson, 1998)。礙於篇幅所限，本文只能集中探討評核系統概念模型和實施架構的建立方法，而未能運用真實案例報告評核系統實施和效度評鑑過程。

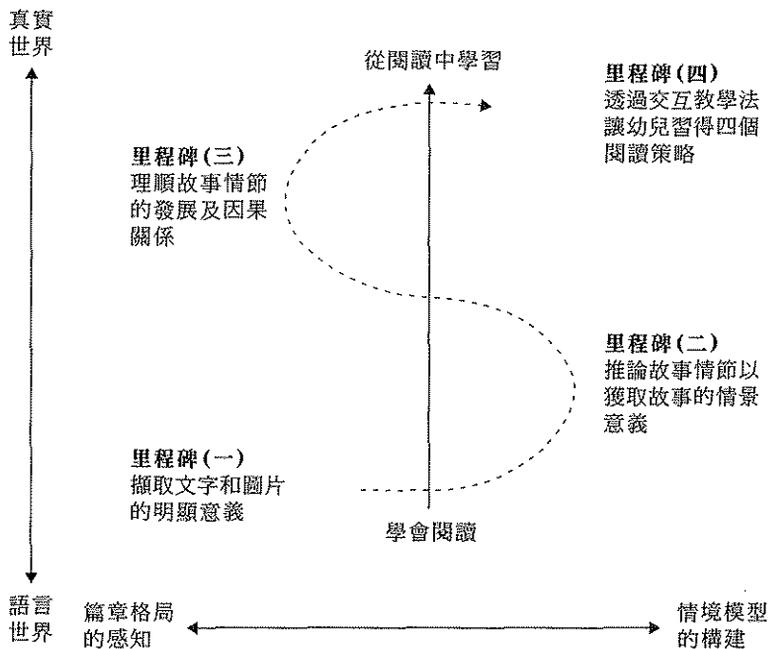
根據 Birenbaum等人(2006)的分析，綜合評核系統有需要體現以下七項基本原理：

1. 學習者會參與評核的過程；
2. 評核是情境化和回應學習者的需要，同時又和教學相關；
3. 綜合評核系統的重點，並不在於學習者知識和/或學習表現的不足之處，而應該是學習者所知道的，以及能夠做得到的；
4. 學習的過程和結果，皆是可以被評核的；
5. 無論對學習者還是教師而言，評核的準則都是顯而易見的；
6. 學習者和教師，都能夠從學習成果中得到回饋；

7. 綜合評核系統的關鍵目的，是讓教師和學習者得知如何進一步促進學習與發展，例如：綜合評核系統讓教師和學習者能夠規劃未來學習的路向。

研究員期望透過建構以故事書和故事板為評量工具的綜合評核系統，幫助教師鷹架（scaffold）幼兒，在課室的情境中學會閱讀預測性的圖書。此綜合評核系統需要將研究員所推薦使用的閱讀指導方法整合（即交互教學法），使其體現上述七項原理的精神，俾能用於評核幼兒的故事閱讀發展水平。在本評核系統中，幼兒的萌發式故事閱讀發展是用里程碑來標示，標誌著幼兒的萌發式故事閱讀是沿著特定路徑發展，而各里程碑之間是有著質的不同，圖一展示研究員所建議採用的萌發式故事閱讀理解概念模型（Sit, 2007），改寫自謝錫金、林偉業、林裕康、和羅嘉怡（2005，頁34）。

圖一 萌發式故事閱讀理解概念模型



萌發式故事閱讀指導與綜合評核系統的理論基礎

美國教育部 (U.S. Department of Education, 1996) 指出，零至八歲是幼兒學習閱讀的重要階段，唯有在此階段幼兒學會閱讀，日後才有可能從閱讀中學習。但謝錫金等人 (2005) 認為從「學會閱讀」到「從閱讀中學習」，是一個不斷積累的連續過程，兒童的閱讀發展路徑，並非於九歲時截然劃分為前後兩個階段，「學會閱讀」和「從閱讀中學習」兩者是不會一夜之間過渡，而是互有重疊。然而，幼兒畢竟年紀尚小，無論是讀圖片還是讀文字，均處於萌發的狀態，因此研究員所擬建構的綜合評核系統的焦點是幫助幼兒「學會閱讀」，為未來的「從閱讀中學習」做好準備。根據讀寫萌發的觀點，在幼稚園階段並不適宜直接教導幼兒閱讀策略，而是引導他們建構，幫助幼兒掌握閱讀策略。考慮幼兒的認知水平及閱歷有限，建構的策略宜簡不宜繁。為此，研究員建議運用交互教學法中的提問、澄清、預測、總結四種閱讀策略，協助幼兒進行閱讀策略的建構以達致閱讀理解 (參閱薛寶娣, 2006; Palincsar & Brown, 1984; Palincsar & Klenk, 1992)。

圖一的萌發式故事閱讀理解概念模型，是以Feldman (1994) 的非普遍性理論為理論基礎。Feldman對於發展心理學關於認知發展的四個假設，有以下的看法：

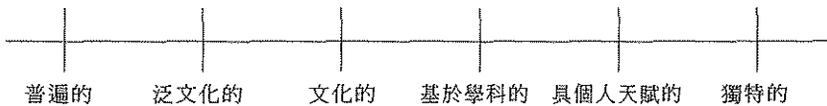
我們繼續接受順序性 (sequentiality) 和層級整合性 (hierarchical integration) 作為發展領域的必要特性，但我們需要限制使用普遍性 (universality) 和自發性 (spontaneousness) 作為區別不同發展現象的準則。(Feldman, 1994, 頁23)

所謂順序性，是指發展是循序向前進行的，不能逾越，亦不能後向發展。所謂層級整合性，是指由一個階段到下一個階段的發展，會經過層級性的累積與重組。前一個階段並非只是下一階段的基礎，而是會經過整合與重組，再轉換成為下一階段的一部分。所謂普遍性，是指發展是全人類共通的，不同文化的幼兒最終都能達致的共同水平。所謂自發性，是指發展是不受環境影響而發生的，個體習

得特定的認知功能和心理操作是無需特殊環境的配合，每一位兒童無需特別的干預，都能通過所有的認知階段及做到基本的心理操作。

Feldman大體上接受上述闡釋的第一和第二個假設，但認為有必要限制第三和第四個假設的意義與應用。在這個基礎上，他提出人類是在一個連續體之上按不同區域由「普遍的」到「獨特的」路徑逐步習得知識和發展能力。「普遍的」到「獨特的」路徑最初分為五個區域，後來加入「泛文化的」區域（參看Feldman & Fowler, 1997, 頁196；Krechevsky, 1998, 頁2）。

圖二 由普遍到獨特的發展範疇（轉引自Krechevsky, 1998, 頁2）



根據Feldman的觀點，閱讀是在文化環境中習得文化知識，屬於文化的範疇（見圖二）。每個孩子都被期待在上述選定的文化範疇中能達致一定的精通水平，但不表示所有孩子都必須達致該範疇的最高水平。此外，Feldman並未有刻劃文化及其他範疇的具體細節，而是簡略地勾勒出由「普遍的」往「獨特的」連續發展的主要里程碑，以說明發展過程中的大致狀況。

研究員採納Feldman的觀點以概念化閱讀理解概念模型的里程碑：

（1）確立擬建構的評核系統涉及Feldman提出的連續體中一個重要階段，閱讀是在文化環境中習得文化知識，屬於文化的範疇。在選取的文化範疇的學習路徑上，設定具體而且有研究基礎的故事閱讀里程碑，以說明幼兒故事閱讀理解的發展過程，讓學習目標變得具體，而且能夠得以讓教師知道幼兒故事閱讀理解達致的程度。

（2）採納順序性的觀點。幼兒的閱讀理解，是會逐漸由里程碑一向里程碑四發展的。此發展亦正如從「學會閱讀」到「從閱讀中學習」的發展是一個不斷積累的連續過程一樣，幼兒的四個閱讀理解發展里程碑，也是一個不斷積累的連續過程，並且每個里程碑之間明顯

地有著質的分別。(3)採納層級整合性的觀點。四個里程碑雖有著發展的先後順序，但各個里程碑在幼兒發展的過程中會互動互補。幼兒不一定要在里程碑一達到精熟的程度，才能進入里程碑二，而是里程碑一達致一定基礎水平時，里程碑二便可開始發展，不過達到的程度不同而已。(4)採納Feldman的自發性和普遍性觀點。幼兒的發展需要環境配合，教師有需要為幼兒創設合宜的閱讀環境。同時，每位幼兒都被期待能達致萌發式閱讀理解的最高水平，但這並不表示所有幼兒在各個里程碑必須達致同一水平。

在使用交互教學法作為閱讀指導策略後，可被識別的幼兒萌發式故事閱讀理解里程碑主要有四個：

- 里程碑一：擷取文字和圖片的明顯意義。根據讀寫萌發的觀點和故事圖書的特點，幼兒若能從文字和圖片擷取意義，即是理解的開始。以故事《小白兔送紅蘿蔔》為例，幼兒在看圖書時能認讀出故事中人物名稱，例如：小白兔、紅蘿蔔；或說出圖片中的重點，例如：兔爸爸、兔媽媽和小白兔一起用力把紅蘿蔔拔出來，即表示幼兒已達到里程碑一所要求的閱讀理解水平。
- 里程碑二：推論故事情節以獲取故事的情境意義。幼兒在擷取文字和圖片的明顯意義後，若能進一步推論故事情節，並將連續的數張圖片串連起來，藉此獲取故事的情境意義，即幼兒已掌握這些故事情節理解的關鍵。以《小白兔送紅蘿蔔》為例，幼兒能運用故事板重構故事場景，並根據自己的理解將其中一個重覆的情節重述或重演的話，即表示幼兒已能獲取故事的情境意義，達到里程碑二所要求的閱讀理解水平。
- 里程碑三：理順故事情節的發展及因果關係。以《小白兔送紅蘿蔔》為例，幼兒能運用故事板的人物、場景、道具，將故事情節的發展按順序重演，及簡略交代故事中的開始、經過和結果，以及當中涉及的因果關係，即表示幼兒已達到里程碑三所要求的閱讀理解水平。

- 里程碑四：透過交互教學法習得四個閱讀策略。此里程碑是萌發式閱讀理解的最高水平，每位幼兒都被期待透過交互教學法能習得四個達致閱讀理解的常用策略，即提問、總結、澄清和預測，並能用以閱讀預測性的圖書。

透過使用交互教學法進行閱讀指導，期望老師能帶領幼兒在「學會閱讀」的過程中沿著里程碑一往里程碑三的路向發展。最終，在長期的干預研究下，教師能有效發揮交互教學法中責任轉移的特點，幼兒成為熟練的讀者，無需教師的協助也可以獨自運用相關的閱讀策略閱讀預測性圖書，從而達致里程碑四所期許的閱讀理解水平。

這四個故事閱讀理解發展里程碑，由兩維所編織而成。水平的一維是由「篇章格局」(textbase)和「情境模型」(situation model)兩端組成(參閱謝錫金等，2005，頁34)。本模型所指的「篇章格局」，意謂幼兒把故事中擷取的文字和圖片意義一一聯繫起來，最終綜合成為連貫的整體，形成對故事情節的概念，這一端是屬於對故事體裁的把握。另一端的「情境模型」，是指幼兒能根據對世界的認識，把擷取的文字和圖片意義建構成有意義的故事情境，使故事成為一個可想像及推理的整體。值得注意的是：同一個故事，不同的讀者應該能夠讀出一致的「篇章格局」，但卻又根據各自的理解，建構出不盡相同的「情境模型」。模型中的里程碑一和三，皆是偏重於「篇章格局」的感知；而里程碑二和四，則是偏重於「情境模型」的構建。

垂直的一維是由「語言世界」和「真實世界」兩端組成。幼兒透過故事中的語言增加了對「真實世界」的瞭解，對「真實世界」瞭解的增加反過來也加深了幼兒對「語言世界」的認識。模型中的里程碑一和二，靠近「語言世界」多於「真實世界」；里程碑三及四則靠近「真實世界」多於「語言世界」。在幼兒從「學會閱讀」走向「從閱讀中學習」的發展旅途上，兒童會在這兩個世界來回往復。

萌發式故事閱讀理解概念模型所闡述的四個有進展特性的里程

碑，為研究員欲建立的綜合評核系統奠定堅實的理論基礎，這概念模型更是綜合評核系統的建構效度評鑑所倚仗的基石。

綜合評核系統的架構

在教師能熟練地運用交互教學法，透過對話與幼兒解釋及示範四項策略的運用，並有系統地創設教師協助下的萌發式閱讀環境以協助幼兒進行長期的閱讀指導後，教師開始在課堂的環境中對幼兒進行評核。萌發式閱讀指導綜合評核系統的實施架構共有四個步驟，各有不同的重點：

步驟一：

課堂觀察。教師於閱讀指導課時，觀察數名不同學習能力的幼兒在各教學環節的表現，以及將這些目標幼兒的學習情況在研究員設計的課堂觀察紀錄表上作出詳細的紀錄，教師同時需跟進目標幼兒往後步驟二、三的評核及步驟四的行動研究。課堂觀察紀錄表按教學環節分為八個部分，每部分由教師的評分和描述性紀錄兩部分組成，即教師針對幼兒在各教學環節的表現，以一至四的評分評定幼兒表現，並以描述性紀錄說明幼兒評分的依據。由於時間及資源所限，教師不可能同時對全班幼兒進行詳細的觀察和紀錄，故此建議教師選取高、中、低不同能力的幼兒進行課堂觀察，以期綜合評核系統日後逐步完善改良，更有可能適用於全體學生。

步驟二：

課後故事書評量。教師根據評核指引，利用故事書對目標幼兒進行個別化的口頭評核。由於故事書評量的主要目的是評核全班閱讀指導課的目標達成程度，因此故事書評量紀錄表會按照閱讀指導課所需達成的教學目標設計。評核的項目有：目標1.0——培養幼兒的閱讀興趣；目標2.1——能從故事書封面初步知道故事的主要角色和內容；目標2.2——能利用圖片加強對故事內容及情節的理解；

目標2.3——初步感知文字(例如：故事名稱)和口語(例如：故事名稱的含義)的對應關係；目標3.0——明白故事重點；目標4.0——理解故事難點。故事書評量紀錄表以李克特(Likert)四點量表評定幼兒表現，並以描述性紀錄說明評分的依據。在評核過程中，教師會進一步對目標幼兒進行閱讀指導，確保幼兒有機會對故事情節內容達致充分的理解。評核過程會進行錄影，並與評核紀錄一併存檔。基於評量信度和效度的考慮，除負責評核目標幼兒的教師外，還會邀請兩位教師一同參與評核工作，方便將評核結果進行三角檢驗。

步驟三：

課後故事板評量。在目標幼兒對故事書有一定理解的前提下，教師與幼兒運用故事板進行互動。教師會視幼兒表現提供鷹架，協助幼兒重述和重演故事的關鍵情節，以及創編故事。評核的重點考慮，是如何透過故事板來展示幼兒對故事情節的瞭解。故事板評量紀錄表同樣以李克特四點量表評定幼兒表現，並以描述性紀錄說明評分的依據。評核的項目有以下七項：項目1——參與程度；項目2——材料利用；項目3——師生互動；項目4——表達能力；項目5——故事結構完整性；項目6——主題一致性；項目7——創造力表現。評核過程會進行錄影，方便評核紀錄與錄像一併用作檔案評核之用。故事板評量紀錄表會配合圖一的萌發式閱讀理解概念模型進行分析，教師和研究員結合步驟一和步驟二所得數據，得以制定目標幼兒的閱讀理解進展分析表，以評定幼兒的閱讀理解水平。(具體作法可參考Krechevsky, 1998)。

步驟四：

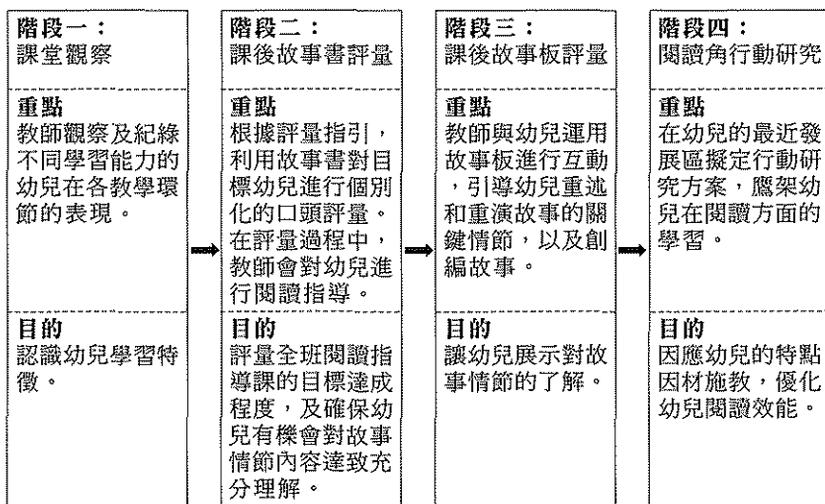
閱讀角行動研究。總結步驟一、二、三對幼兒進行閱讀時所遇到的問題，在幼兒的最近發展區擬定行動研究方案，希望在閱讀角處理好幼兒在閱讀時遇到的問題，優化閱讀效能，以及進一步促進幼兒的語言發展。在實施步驟四時，首先根據故事書評量和故事板評量的過程中，教師有沒有為幼兒提供鷹架，以及評量的評分結果，分

析幼兒閱讀效能，並結合圖一的萌發式故事閱讀理解概念模型，在幼兒的最近發展區擬定可行之行動研究方案。教師擬定行動研究方案後，會與其他教師進行內部討論和協調，最後於閱讀角將行動研究方案落實。行動研究過程會錄影存檔，供教師檢討和反思，以及作為擬定新一輪的行動研究方案之用。

研究員期望完成這四個步驟之後，教師能夠：（1）更具體地認識幼兒的學習特徵；（2）評鑑全班閱讀指導課的目標達成程度，並確保目標幼兒有機會對故事情節內容達致充分的理解；（3）確保目標幼兒有機會將這些理解顯而易見地展示出來；（4）因應幼兒的學習特點因材施教，優化幼兒閱讀效能。

以下是研究員所建議採用的「萌發式閱讀理解綜合評核系統」的具體實施架構：

圖三 萌發式閱讀指導綜合評核系統實施架構



由上可見，本綜合評核系統於設計時既將「對學習的評核」和「促進學習的評核」整合，兼具兩者的優點。研究員遵照萌發式閱讀指導綜合評核系統的實施架構，為澳門一所公立幼稚園四歲班的幼兒進行一項為期四個月的實驗研究，期間盡力確保教學和評核的

過程實現Birenbaum et al. (2006) 提出的七項原理，初步證實此實施架構能成為一個有教育功效的綜合評核系統（參閱薛寶嫦，2007）。

綜合評核系統的信效度考慮

任何評核系統在實際應用之前，都需要考慮信度和效度等議題。儘管現時與綜合評核系統相關的文獻，為如何建構綜合評核系統提供了指引，但對如何評鑑綜合評核系統的信度和效度，並未有提供具體的方法，這其實是有需要詳加說明的（Birenbaum et al., 2006）。

關於綜合評核系統的信度，體現於測量員在故事書和故事板評分的一致性上。由於實施時會選取數名而不是全體幼兒，目標是綜合反映幼兒的閱讀理解進展水平，故會集中評鑑測量員之間的一致性。具體的做法是每一位目標幼兒由一位指定的教師負責評量，另外兩位教師亦會分別對目標幼兒評分和紀錄評核的整個過程。這三位教師在評量後，會即時討論每一項評量項目的評分。評分如果有不一致的情況出現，他們會利用評分時所寫下的觀察紀錄，共同找出原因。這樣，不但評量項目內容的清晰度有所提高，同時，不同教師對不同能力的目標幼兒的評分亦有了共識，評分的標準變得更趨於一致。值得一提的是：透過上述的過程，不但提高了評量的信度，亦提高了評量的效度，原因是假設教師在討論後仍有不同的意見，他們可以參考研究員的評分，化解分歧並尋求共識。

以下將分兩部分具體說明評量工具信度和綜合評核系統效度的具體方法。由於綜合評核系統的核心是故事書評量和故事板評量的評分量表，如果收集大量數據的話，這兩個量表可以運用項目分析、Cronbach-alpha信度系數和因素分析等現有的方法檢測信度和效度。但由於擬選取數名不同能力的幼兒而不是全體幼兒進行評核，因此，評量教師與研究員之間的評分一致性遂成為本文焦點所在。為此，第一部分闡釋評量工具信度分析方法，重點檢視綜合評核系統中的故事書評量和故事板評量的評分一致性信度。由於現有文獻未有說明如何評鑑綜合評核系統的效度，因此第二部分闡釋綜合評核系統

的效度評鑑方法，主要說明如何引用Messick（1989）統一建構效度理論評鑑整個綜合評核系統的建構效度。

評量工具信度分析

綜合評核系統中利用故事書和故事板於課後對幼兒進行個別化評量。由於故事書評量的主要目的是評核全班閱讀指導課的目標達成程度，因此故事書評量紀錄表會按照閱讀指導課所需達成的教學目標設計。故事板評量的重點考慮，是如何透過故事板來展示幼兒對故事情節的瞭解，因此故事板評量紀錄表會重點紀錄幼兒對故事情節的理解。兩張評量紀錄表均以李克特四點量表評定幼兒表現，並以描述性紀錄說明評分的依據。以下將以故事書評量假設性評分為例，說明研究員如何檢視測量員之間評分的一致性。

從表一可以看到，教師一和教師二之評分以3和4為主，但兩位教師除了在目標2.3的評分是一致以外，其餘五個評量項目的評分各有不同，因此負責評量的教師（教師三）需要決定幼兒在不同評量項目的最後的評分。

表一 故事書評量的信度分析表

評量項目	故事書評量等級（四點李克特量尺）				一致性程度*
	教師一	教師二	教師三	研究員	
目標1.0	3	4	4	3	0.66
目標2.1	3	4	4	4	1.00
目標2.2	4	3	3	4	0.66
目標2.3	4	4	4	4	1.00
目標3.0	4	3	4	3	0.66
目標4.0**	—	—	—	—	—
一致性百分比 = 0.80					

* 反映教師三和研究員評分的一致性程度：當兩者的評分相同，一致性程度為1.00；當一人的評分是1，另一評分是4，兩人的評分相差三個等級，一致性程度為0.00；當兩者的評分相差一個等級（例如：1和2、2和3、3和4），一致性程度為0.66；當兩者的評分相差兩個等級（例如：1和3、2和4），一致性程度為0.33。

** 由於此次教學實驗的故事沒有難點，故此評量項目暫不適用。

教師三的評分是通過三位教師之間的協商和討論的過程而得到的。在表一的第四欄中，可以看到教師三在目標1.0、2.1、2.2三個的評量項目是同意教師二的評分，而在目標3.0的評量項目則同意教師一的評分。透過這個協商和討論的過程，教師三對故事書評量中各個項目的評分準則，有了更清楚的瞭解，因而更加能分辨各個項目之評分3和評分4的不同之處。

由於教師對幼兒進行評量時，除判定項目的評分外，還需詳述幼兒在進行評量時的表現，以此作為評分的依據。因此，在協商和討論的過程，評量的信度會增加。表一的第四和第五欄中，可以看出教師三和研究員的評分在目標2.1和目標2.3是一致的，而在目標1.0、目標2.2、和目標3.0則有所不同。兩人評分的不一致，信度自然欠佳，效度亦會存疑。為此，在第六欄計算了教師三和研究員評分的一致性程度，以檢視評量的信度。表一的第六欄中，可以看出一致性程度由0.66至1.00，平均值為0.80，說明這名目標幼兒的故事書評量有不錯的信度。雖然每個評量項目開始時是利用李克特量尺進行評分，但由於目標幼兒採樣自不同學習能力和水平，教師通過上述的協商和討論過程，應能理解每一個評量項目的不同評分（即李克特量尺中的1-4點）的評量準則。運用上述的方法，也能檢視故事板評量時測量員評分的一致性。同樣地，教師通過當中的協商和討論過程，應能理解故事板評量紀錄表上每一個評量項目的不同評分（即李克特量尺中的1-4點）的評量準則。

此外，不同能力的目標幼兒，評分的整體信度是否存在著差異，則需進一步進行檢視。在計算出每位幼兒故事書評量的一致性得分後，將其計算一致性程度之總體平均值，就能檢視評量工具的整體信度。假如評量項目的一致性程度之平均值低於0.7，便需要對相關的評量項目重新釐清內涵。

綜合評核系統的效度評鑑方法

研究員擬借助Messick（1989, 1995）的統一建構效度理論評鑑綜合評核系統的效度，評鑑的過程建議分四個步驟漸進地進行（參看表二）。

表二 綜合評核系統的效度分析步驟

	數據的解釋	評核的用途
證據搜集	<p>步驟一：建構效度(Construct validity: CV)</p> <p>結合不同數據(課堂觀察、故事書評量評分及故事板評量評分)，綜合解釋不同學習能力幼兒故事閱讀理解的情況，從而讓教師及研究員研判目標幼兒的閱讀里程碑的進展水平。</p>	<p>步驟二：CV + 適切性／實用性(Relevance/Utility: R/U)</p> <p>根據步驟一所判定的進展水平資料，教師為不同故事閱讀理解水平的幼兒擬定行動研究方案，鷹架幼兒在閱讀方面的學習。行動研究方案的擬定和實施，需搜集證據說明綜合評核系統評核用途的適切性和實用性。</p>
後果考慮	<p>步驟三：CV + 價值隱喻(Value implication: VI)</p> <p>根據步驟一所判定的進展水平資料，檢視不同學習能力幼兒的閱讀里程碑所依靠的理論背後的價值判斷準則，了解這些準則會為幼兒帶來怎麼樣的教育後果，分析該怎麼樣解釋數據來鷹架幼兒才會發揮綜合評核系統的功效。</p>	<p>步驟四：CV + R/U + VI + 社會後果(Social consequence: SC)</p> <p>綜合步驟一、二、三的結論，判定綜合評核系統的實施成效是否與預期的用途相符？有沒有值得關注的教育和社會後果需要考慮？</p>

步驟一：

在故事書和故事板評量得到的評分經過信度的計算後，結合課堂觀察，綜合解釋不同學習能力幼兒故事閱讀理解的情況，從而讓教師及研究員研判目標幼兒閱讀里程碑的進展水平，教師因此能夠清晰表述對達致不同里程碑的幼兒閱讀具體發展情況。此外，由於選取了一群學習能力不同的幼兒為目標幼兒，幼兒達致的閱讀里程碑的進展水平應有差異，讓研究員能對圖一的萌發式故事閱讀理解概念模型中各個不同里程碑有通盤檢視的機會。

步驟二：

根據綜合評核系統的實施架構，教師需在幼兒的最近發展區鷹架幼兒，以達致更高閱讀理解水平。教師以步驟一所判定的進展水平資

料為基礎，針對幼兒的真實表現，為不同故事閱讀理解水平的幼兒擬定行動研究方案，鷹架幼兒在閱讀方面的學習。行動研究方案的擬定和實施，需搜集證據說明綜合評核系統評核用途的適切性和實用性。假設目標幼兒讀圖不細緻，掌握故事的情節亦未夠完整，反映幼兒的里程碑一有一定基礎，但仍然需要鞏固，里程碑二發展亦有待提升。教師針對幼兒的表現，擬定相關的行動研究方案——利用圖片重組故事內容，過程中教師用提問方式，引發幼兒對圖片進行細心觀察，加強幼兒對故事完整情節的掌握，以提升幼兒的表現。故此行動研究方案的擬定和實施，需搜集行動研究成效的證據，以瞭解整個綜合評核系統的功效。

步驟三：

由於閱讀教學設計和行動研究方案是集體擬定，教師之間最初的不同意見，透過彼此的溝通和協商而消弭，並能釐清背後的教育理念的理解是否恰當。故事閱讀指導完結後，教師和研究員一起檢討成效，在大家集思廣益的過程中，教師需要知道從甚麼途徑可以幫助幼兒，以及分析教學成效之有利因素。教師和研究員在檢視行動研究方案背後的指導理念的同時，教師檢視不同學習能力幼兒的閱讀里程碑背後的價值判斷準則，思考這些準則會為幼兒帶來怎麼樣的教育後果，以及該怎麼樣解釋數據來鷹架幼兒才會發揮綜合評核系統的功效。例如前述的目標幼兒最後決定將重點放在加強幼兒對故事情節的掌握上，而不是放在鞏固幼兒里程碑一的發展。理由是目標幼兒的里程碑一已有一定基礎，只要教師鷹架得宜，再配合適當的環境，幼兒就能透過圖片加強對故事情節的掌握，幼兒里程碑二的發展應能提升；而提升里程碑二發展的同時，應能反過來加強幼兒對圖片的感知。

步驟四：

研究員需檢視綜合評核系統實施成效是否與預期的用途相符，以及有沒有值得關注的教育和社會後果需要考慮。例如：本綜合評核系

統實施時能否讓研究員及教師認識幼兒閱讀指導的教學成效？幼兒是否順利由一個里程碑邁向另一個里程碑，最終能學會閱讀？閱讀指導教學如何與現時的故事課結合？閱讀指導教學如何與主題活動教學配合？幼兒課程應如何重新設置，才能教導幼兒讀懂故事圖書？學校的資源如何重新調配，才能做好閱讀指導教學？透過綜合評核系統，可以為前四個問題帶來啟示；而後兩個問題，則是落實綜合評核系統時所必須考慮的。

總結

本文為回應教學評核要同時顧及「對學習的評核」和「促進學習的評核」的訴求，重點選擇了文化範疇中早期閱讀這個領域探討實踐建構綜合評核系統的可行性。本文的主要目的是建構一個萌發式故事閱讀理解綜合評核系統，以及提出一個教師可操作的實施架構。

這個實施架構主要由四個環節構成，包括：（1）課堂觀察、（2）課後故事書評量、（3）課後故事板評量，以及（4）閱讀角行動研究。評核系統希望能夠做到教學與評核緊密結合，教師將更能適時掌握學生的閱讀發展情況，以便做到因材施教，最終目的是鷹架學生運用閱讀策略讀懂預測性故事圖書。鑑於任何評核系統都有需要交代評核過程和結果的信度和效度。本文建議分析教師和研究員評分的一致性程度，以及借助Messick（1989, 1995）的統一建構效度理論分四個步驟對綜合評核系統的建構效度進行評鑑。

參考文獻

- 周兢(主編)(2001)。《幼兒園語言教育活動設計與組織》。北京：人民教育出版社。
- 陳淑琴(2001)。《幼兒語文教材教法》。台北：光佑文化。
- 黃瑞琴(1997)。《幼兒讀寫萌發課程》。台北：五南。
- 趙鈞鴻(1994)。《小白兔送紅蘿蔔》。香港：晶晶教育出版社。

- 薛寶嫦(2006)。〈幼兒萌發式閱讀環境的創設〉。《幼兒教育》(教育科學版)，第343期，頁20-23。
- 薛寶嫦(2007)。《透過課例研究協助教師發展幼兒的閱讀理解能力》。未出版博士論文，北京師範大學。
- 謝錫金、林偉業、林裕康、羅嘉怡(2005)。《兒童閱讀能力進展——香港與國際比較》。香港：香港大學出版社。
- Assessment Reform Group (2002). *Assessment for learning: Ten principles*. Retrieved May 22, 2006, from <http://www.assessment-reform-group.org.uk>
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., & Ridgway, J. (2006). A learning integrated assessment system. *Educational Research Review, 1* (1), 61-67.
- Feldman, D. H. (1994). *Beyond universals in cognitive development* (2nd ed.). Norwood, NJ: Ablex.
- Feldman, D. H., & Fowler, R. C. (1997). The nature(s) of developmental change: Piaget, Vygotsky, and the transition process. *New Ideas in Psychology, 15* (3), 195-210.
- Flippo, R. F. (2003). *Assessing readers: Qualitative diagnosis and instruction*. Portsmouth, NH: Heinemann.
- Krechevsky, M. (1998). *Project Spectrum: Preschool assessment handbook*. New York: Teachers College Press.
- Mariotti, A. S. (2005). *Linking reading assessment to instruction: An application worktext for elementary classroom teachers*. Retrieved May 22, 2006, from <http://site.ebrary.com/lib/umac/Doc?id=10110091&ppg=197>
- McGee, L. M., & Richgels, D. J. (2004). *Literacy's beginnings: Supporting young readers and writers* (4th ed.). Boston: Pearson.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117-175.
- Palincsar, A. S., & Klenk, K. (1992). Fostering literacy learning in supportive

- contexts. *Journal of learning disabilities*, 25(4), 211–225, 229.
- Paris, S. G., & Stahl, S. A. (2005). Current issues in reading comprehension and assessment. Mahwah, NJ: Lawrence Erlbaum.
- Sarroub, L., & Pearson, P. D. (1998) Two steps forward, three steps back: The stormy history of reading comprehension assessment. *The Clearing House*, 72(2), 97–105.
- Sit, P. S. (2007). Lesson study as a means to help kindergarten teachers develop children's reading comprehension ability. *Journal of Research on Elementary and Secondary Education*, 18, 145–164.
- U.S. Department of Education. (1996). *Reading: The first chapter in education*. Retrieved May 22, 2006, from <http://ericec.org/frstchap.html>

The Development of an Integrated Assessment System for Reading Comprehension

Pou-seong Sit

Abstract

This paper begins with an explanation of the inter-relationship between teaching and assessment as well as points out the importance of developing an integrated assessment system for emergent story reading comprehension. Based on Feldman's (1994) non-universal theory and Palincsar and Brown's (1984) reciprocal teaching methods as reading instruction strategies, this paper makes use of storybook and storyboard as assessment tools and constructs a conceptual model of an integrated assessment system for evaluating young children's emergent story reading comprehension. After considering the principles underlying an integrated assessment system, an implementation framework is formulated in accordance with the conceptual assessment model. Finally, with reference to Messick's (1989) theory of unified construct validity, the procedures for evaluating the validity of the integrated assessment system are presented.