

# 閱讀理解成長測驗之編製研究

蘇宜芬\*

國立臺灣師範大學  
教育心理與輔導學系

洪儷瑜

國立臺灣師範大學  
特殊教育學系

陳柏熹

國立臺灣師範大學  
教育心理與輔導學系

陳心怡

國立臺灣師範大學  
特殊教育學系

本研究主要目的為編製適用於小學四、五、六年級學生的多複本閱讀理解測驗，提供一套可瞭解閱讀理解能力發展情形的工具，也可用於閱讀低成就學生的補救教學成效評估，作為介入反應之決定。題目型式為測量文章理解的題組。每套複本都有四個題組，每個題組都有一篇文章及 10 個閱讀理解題目。四個題組的文章分別是短記敘文、短說明文、長記敘文、長說明文。40 道題目中有兩題是不計分的，主要用以確認受試者是否認真作答。每個題組中的題目所測量的閱讀理解成份為字彙觸接、字面理解、摘取大意、及推論理解。在信度考驗方面，四、五、六年級每個複本的 Cronbach's  $\alpha$  信度及折半信度多在 .8 以上。在效度方面，以「國民小學閱讀理解篩選測驗版本 A」為效標，所測得的效標關聯效度介於 .59~.75 之間。本測驗除了提供百分等級常模外，也提供能力值轉換常模，以便於繪出學生的閱讀理解能力成長曲線及計算成長係數。本研究也對於教育應用上的建議與限制提供說明。

關鍵詞：成長曲線、能力值、國小學生、複本測驗、閱讀理解

---

\* 本篇論文通訊作者：蘇宜芬，通訊方式：yifensu@ntnu.edu.tw。

本研究感謝國立臺灣師範大學教育研究與評鑑中心在命題階段的經費補助（師大研發字第 098000240 號），也承蒙科技部在預試及建立常模階段之經費補助（計畫編號：NSC 99-2420-H-003-003，NSC 100-2420-H-003-003-MY2）。研究過程中，感謝何緬翎、張璣勻、張祐瑄、徐珮筠在命題上的協助，袁瑞伶、賴彥文在施測聯繫與資料登錄上的幫忙，以及邱于真、謝佳真、張倫睿、薛人華、楊元傑等在資料分析上的協助。最後，也感謝兩位審查委員提供精闢的修改意見。

閱讀理解是閱讀能力發展的最終目標，也是各學科學習的基礎，學生若具備良好的閱讀理解能力，即掌握了自學的方法（Burns, Griffin, & Snow, 1999），這在知識發展快速，強調終身學習的二十一世紀尤其重要。由於閱讀能力的習得是學習的重要任務，那麼透過適當的測驗工具，瞭解學生閱讀理解能力成長的狀況就頗為重要。

就閱讀能力的發展而言，根據 Chall（1996）所提出的閱讀發展階段論，其中，與中小學有關的閱讀階段為：初始閱讀階段（initial reading or decoding stage），流暢閱讀階段（confirmation, fluency, ungluing from print），閱讀新知階段（reading for the new），及多元觀點階段（multiple viewpoints）。「初始閱讀階段」約 6~7 歲，相當於一至二年級。此時的發展重點在識字，學習者能覺察文字與讀音之間的對應關係，此階段在閱讀時容易發生認字上的錯誤。「流暢閱讀階段」約 7~8 歲，相當於二至三年級。此時的發展重點在識字技能的自動化，亦即能流暢地閱讀，此階段比較能夠精確地辨識文字，以及建構文字與意義間的連結。「閱讀新知階段」約 9~14 歲，相當於四至八年級。此時的發展重點為透過閱讀學習新知，此階段學生不僅藉由大量閱讀吸收知識，也開始發展閱讀策略。「多元觀點階段」約 14~18 歲，相當於國中的八年級至高中三年級。此時的發展重點為豐富觀點，此階段的學生有能力理解內容複雜，觀點多元的文章，並能對內容分析，形成初步的批判。由 Chall 的閱讀發展階段可知，小學一至三年級的學習重點在發展流暢的字詞辨識能力，四年級之後的學習重點則在閱讀理解能力的發展，尤其是能夠透過閱讀學習新知。因此，藉由合適的工具瞭解四年級以上學生閱讀理解能力發展情形，以及對於閱讀理解能力落後的學生即時補救，顯得格外重要。

目前國內正式出版或常被縣市特殊教育中心學校所採用的標準化閱讀理解能力測驗包括：柯華葳（1999）的「閱讀理解困難篩選測驗」，林寶貴與錡寶香（2000）的「中文閱讀理解測驗」，柯華葳與詹益綾（2006）「國民小學（二至六年級）閱讀理解篩選測驗」，柯華葳與詹益綾（2006）「國民中學閱讀推理測驗」，及王木榮與董宜俐（2006）的「國小學童中文閱讀理解測驗」。其中，除了柯華葳與詹益綾（2006）「國民小學（二至六年級）閱讀理解篩選測驗」，為每個年級提供 A、B 兩個複本外，其他均為單一版本的測驗。Kameenui 等人（2006）特別強調，能力成長的監控需要有複本（Alternative parallel forms）的測驗工具，而且對於不同需求、不同教學目的的學生，成長監控的頻率也應有所不同。國內現有的複本測驗最多只有兩個複本，僅能提供前後測比較，難以進行能力成長的持續追蹤。由於閱讀理解複本測驗的不足，所以發展多複本的閱讀理解測驗，在國內有實質上的需要。因此，本研究乃為小學進入「閱讀新知階段」的四至六年級學生，編製多複本的閱讀理解測驗。

就閱讀理解測驗的內容而言，前述國內近年正式出版或常被國小採用的標準化閱讀理解能力測驗之內容與題型如表 1 所示。其中，林寶貴與錡寶香（2000）的「中文閱讀理解測驗」及王木榮與董宜俐（2006）的「國小學童中文閱讀理解測驗」，均以測量文章閱讀理解為主，也就是受試者需先讀完文章，然後再回答文章後面所附的問題。而柯華葳（1999）的「閱讀理解困難篩選測驗」及柯華葳與詹益綾（2006）「國民小學（二至六年級）閱讀理解篩選測驗」，由於編製的主要目的是為篩選閱讀困難的學童，為了避免難度高產生地板效應，或是使落後的學童抗拒作答，所以二分之一到三分之二的題目是屬於「部份處理」的題目，也就是測量對一個句子裡的某個字詞，或是對一個句子整體理解的題目。即使柯華葳（1999）「閱讀理解困難篩選測驗」的閱讀（一）、閱讀（二），及柯華葳與詹益綾（2006）「國民小學（二至六年級）閱讀理解篩選測驗」的「本文處理」，所採用也是 100 字左右的短文。根據 Keenan、Betjemann 與 Olson（2008）的研究，以一、兩個句子的短題測量閱讀理解，所測得分的總變異能被字詞辨識能力解釋的程度，往往高於能被理解能力解釋的比重。原因可能是只有一、兩個句子的短式閱讀理解題，因為缺乏前後文的線索，因此學生容易因為對於題目中的某一兩個字詞不甚理解，而影響其答題表現。由於本研究所編製的測驗期能適用於一般的小學四、五、六年級學生，而這階段的學生需發展閱讀文章的能力，以便透過閱讀吸取新知，所以本研究所發展的測驗以文章閱讀理解的型式為主。為了顧及學生能力上的個別差異，所以每個複本都有短文的題組及長文的題組，文體則記敘文與說明文兼備。

表 1 國內近年出版的國小標準化閱讀理解能力測驗之內容與題型

編製者	測驗名稱	適用年齡	測驗內容／題型
柯華葳 (1999)	閱讀理解困難篩選測驗	小學 2~6 年級	字意題、命題組合、句理解題、閱讀(一)(二)
林寶貴與錡寶香 (2000)	中文閱讀理解測驗	小學 2~6 年級	6篇故事類記敘文與6篇說明文，在每篇文章之下含有：音韻處理、語法、語意、理解文章基本事實、摘要重點大意、推論、比較分析等能力之題目。
柯華葳與詹益綾 (2006)	國民小學(二至六年級)閱讀理解篩選測驗	小學 2~6 年級	2年級：命題組合、句子理解、短文理解 3~6年級： 部份處理：多義字、命題組合 本文處理：句子理解、短文理解
王木榮與董宜俐 (2006)	國民小學閱讀理解測驗	國小 6 年級	4 篇文章，每篇文章之下含有：字義理解、文本理解、推論理解、摘要能力、布題能力 5 類題目。

關於閱讀理解的次成分，最簡單的分法就是「字面理解」(literal comprehension)與「推論理解」(inferential comprehension) (Taylor, Harris, Pearson, & Garcia, 1995, pp. 267-268)。Taylor 等人 (1995) 對「字面理解」題目的界定是：若題目與答案是在文章裡的同一個句子，就是字面理解的題目。Gagn'e (1985) 則認為字面理解包含兩個部分，一個是「字彙觸接」(lexical access)，另一個是「剖析」(parsing)。所謂「字彙觸接」是指認出文字的字面意義，例如：能夠了解「小明每天都穿制服上學」句中的「服」是衣裳的意思，而「小明這學期擔任班上的服務股長」句中的「服」是做事的意思。至於「剖析」是指將幾個文字的字義、詞義，依照它們之間的關係(如：字序、文法)組合在一起，行成一個命題，也就是能夠瞭解句子所要表達的意思。例如：能夠瞭解「小明叫小華」和「小華叫小明」這兩句是不同的，前一句裡叫人的是小明，但後一句裡叫人的是小華。由於中文有不少多義字、多義詞，讀者需透過前後文的線索，才能對多義字詞在句子裡的意義正確解讀，一旦讀者對多義字詞的解讀錯誤，就容易對文義產生誤解。因此，多義字詞的理解在中文閱讀裡格外重要。所以本研究的命題架構裡，把「字彙觸接」獨立為一個命題成份，主要是測量多義字詞的理解，以及文章裡重要語詞的理解。「字彙觸接」的題幹型式如：「上文裡，『秀才寫完後，呈給縣官看』，句中『呈』可以替換為下列哪一個詞，意思最接近？」，又如：「上文裡，『你這個兩腳書櫥』，句中『兩腳書櫥』一詞在下面哪一句的使用最恰當？」。至於「字面理解」的題目，則兼採 Taylor 等人 (1995) 對字面理解的界定及 Gagn'e (1985) 對「剖析」的定義來設計題目，也就是學生可以在文章裡直接找到答案的問題。「字面理解」的題幹型式如：「根據這篇文章的敘述，以下哪一項是秀才開始找工作的原因？」，又如：「根據這篇文章，吉姆·羅傑斯和女兒分享了什麼事？」，「字面理解」的題目都可以直接在文章的字裡行間找到答案。

關於「推論理解」，Taylor 等人 (1995) 的界定是：讀者需統整文章裡兩個以上的句子才能產生答案，或是需把文章裡的訊息跟讀者自己的生活經驗、背景知識加以整合，才能產出答案的題目。Gagn'e (1985) 則認為推論理解包含統整、摘要、精緻化 (elaboration) 三個歷程。「統整」是指能發覺句子之間的隱含關係。例如：學生讀了「黑熊向小明走過來」，「他趕快跑開」這兩個句子，能夠推論小明可能為了躲避黑熊的攻擊而跑開。所謂「摘要」是指讀完一段文字後，能夠歸納出該段文字的主要大意。所謂「精緻化」是指能運用自己的先備知識，賦予文句或段落更豐富的意義。例如：學生讀到「一場颱風把還沒成熟的柚子吹落了一地，今年和爸爸上山採柚子的希望又落空了」，能夠推論這個事件發生在夏秋之間，而且主角的心情是失望的。

此外，張雅如與蘇宜芬 (2003) 也曾比較 Magliano 與 Graesser (1991)、Kintsch (1993)、Trabasso 與 Suh (1993)、van den Broek、Fletcher、與 Risden (1993)、Pressley 與 Afflerbach (1995) 的推論分類架構，嘗試找出這些理論之間推論類別的主要交集。茲將這五個推論分類架構先分述如下。

## 一、Graesser 以知識為本 ( knowledge-based ) 的推論分類架構

這個分類架構強調讀者閱讀時，融入文章訊息之外的知識結構，以擴充文章表徵所提供的有限訊息。提倡的學者以 Graesser 為主，他也陸續與其他學者合作修訂此分類架構。其中，由 Magliano 與 Graesser (1991) 所提出的推論類別包括以下十一項：

(一) 指稱推論 (anaphoric reference)：指讀者閱讀「焦點陳述句」(focal statement，也就是讀者當下正在讀的句子)時，必須連結前文出現的元素，以找出焦點陳述句指稱的內容或對象是前文所提過的哪個抽象或具體名詞、代名詞、片語、或事件，其目的在達成局部文章意義的連貫。

(二) 前因推論 (causal antecedent inference)：是指連結焦點陳述句與先前事件之間的因果關係，主要在推論目前行動、事件、狀態發生的原因。

(三) 後果推論 (causal consequence inference)：是指根據目前的文章訊息，預測後續的因果關係，也就是對文章內容的預期。

(四) 工具推論 (instrument inference)：是指讀者對於角色在進行行動時，所使用的物體、身體部位、資源等所做的推論。由於這與動作具有高度的相關，因此讀者可憑其背景知識或經驗達成推論。

(五) 名詞類別的示例 (instantiation of noun category)：是指推論名詞所包含的次級類型或特殊例子。例如：舉出蘭花、茶花做為花朵的例子，或說出籃球賽為球賽的一種等等。

(六) 上位目標與行動 (superordinate goal)：是指推論角色行動的原因。

(七) 下位目標與行動 (subordinate goal/actions)：是指推論角色如何完成行動。

(八) 狀態推論 (state)：是指與文章事件沒有因果關連，但為持續進行的狀態，例如：角色特質、知識信念、空間位置、物品特性等。

(九) 主旨推論 (theme)：是指推論文章的主旨、寓意。

(十) 角色情緒反應 (emotion)：是指推論文中角色對於事件的情緒反應。

(十一) 作者的意圖 (author's intent)：是指推論作者寫作的態度或動機。

## 二、Kintsch (1993) 的三向度分類架構 (triple classification scheme)

Kintsch(1993)以三個向度對推論進行分類，第一個向度是「自動化歷程」(automatic processes)或「控制歷程」(controlled processes)，前者是指在閱讀過程中，讀者不太耗費認知資源、自動化地產生推論，後者則指讀者需運用到認知資源，需要思索、統整地產生推論。第二個向度是「訊息增加」(accretion of information)或「訊息減少」(reduction of information)，如前面「精緻化推論」所提過的例子「讀者讀到『一場颱風把還沒成熟的柚子吹落了一地，今年和爸爸上山採柚子的希望又落空了』，讀者能夠推論這個事件發生在夏秋之間，而且主角的心情是失望的」，這就屬於「訊息增加」的推論；至於主旨、大意的推論，就屬於「訊息減少」的推論。第三個向度是「提取 (retrieval) / 產出 (generation)」或「刪除 (deletion) / 產出 (generation)」，其中，「提取」是指提取背景知識或經驗，「產出」是指遞移推論、邏輯推論、或歸納原則；「刪除」則指刪除不重要的細節或萃取重點。依這三個向度，推論可分為八個類型，如表 2 所示。

表 2 Kintsch ( 1993 ) 的三向度分類架構

	訊息增加		訊息減少	
	提取	產出	刪除	產出
自動化歷程 (具備豐富先備知識且處理迅速)	文章線索充足，讀者可自背景知識中，提取精緻化文章表徵所需的訊息。	訊息產出的來源來自文章內容，而非提取自背景知識，例如：根據文章內容所產出的遞移推理。	刪除文章裡不重要的細節	在熟悉的主題領域，建構、歸納訊息。
控制歷程 (先備知識較缺乏且處理時間長)	文章線索不足，讀者需以控制的歷程，自背景知識中提取所需的訊息。	根據文章內容，經過複雜邏輯推理過程而產生的新訊息。	在萃取重點的過程裡，把經過判斷，不具重要性的內容刪除。	在不熟悉的主題領域，建構、歸納訊息，其歷程需時較長。

### 三、Trabasso 與 Suh ( 1993 ) 推論的心理運作模式

此模式是以分析放聲思考所得的資料發展而成，共分為九種推論與心理運作的類型：

(一) 複述文本 (Repeat Text)：是指對焦點陳述句的主要命題進行部份或全部的複述或以自己的話重述。

(二) 複述思考 (Repeat Thought)：是指對焦點陳述句閱讀時所產生的思考，進行部份或全部的複述或以自己的話重述。

(三) 提取文本 (Retrieve Text)：是指提取焦點陳述句之前的文本訊息。

(四) 提取思考 (Retrieve Thoughts)：是指提取焦點陳述句之前進行的思考訊息。

(五) 精緻化文本後果 (Elaborate Consequence of Text)：是指將焦點陳述句與從背景知識中增加的訊息連結，進行因果推論。

(六) 精緻化連結 (Elaborate Associate of Text)：是指將焦點陳述句與從背景知識中增加之訊息連結，進行非因果的推論。

(七) 解釋 (Explanation)：是指提取前文或思考產生的因果推論。

(八) 預測 (Prediction)：是指對文章將發生的事件進行推論。

(九) 後設評論 (Meta-Comments)：是指對文中角色之所為或焦點句是否容易理解等進行評鑑性的推論。

### 四、van den Broek、Flecher 與 Risdien ( 1993 ) 的推論類型

van den Broek 等人 (1993) 是以讀者目前閱讀中的「焦點陳述句」(focal statement) 為推論產生的中心，根據推論的功能與所運用的訊息來源，把推論分成四種類型：

(一) 回溯 (橋樑) 推論 (backward/bridging inference)：是指連結焦點陳述句與先前之訊息以達成連貫的文章表徵。這又可細分為：1. 與短期記憶中的文章訊息相連的連結推論 (connecting

inference)；2. 從長期記憶提取文章訊息的再生激發 (reinstatement)；3. 與讀者經驗、先備知識相連的回溯精緻化 (backward elaboration)。

(二) 向前精緻化 (forward elaboration)：預測文章尚未閱讀部分的内容。

(三) 直交精緻化 (orthogonal elaboration)：產生與焦點陳述句並存、且增加對其理解的訊息。這有許多不同的形式，目前研究較多的形式包括視覺、空間的推論，與讀者情感投入的推論。

(四) 聯想推論 (associative inference)：透過擴散激發 (spreading activation) 機制而自動產生與焦點陳述句相關的訊息，為其他推論產生的基礎。

## 五、Pressley 與 Afflerbach (1995) 的推論分類架構

這是以「資料紮根分析法」(grounded analysis of the data) 分析放聲思考的研究文獻，企圖達到理論飽和，歸納出所有閱讀理解處理歷程。其所包含的推論類別包括：

(一) 指稱推論

(二) 填補被刪掉的訊息，也就是推論者在文章裡沒有直接提及的訊息。

(三) 根據前後文脈絡推論字義

(四) 推論字與句的言外之意 (connotation)

(五) 連結先前知識與文章的訊息

(六) 有關文章作者的推論，例如：作者的目的、觀點、信念、寫作手法、專長領域等。

(七) 透過文章後面的内容，檢驗讀者自己所產生的推論是否正確。

(八) 提出隱含的結論。

在這五個推論分類架構裡，交集最多的推論理解成份為：指稱推論 (referential inference 或 anaphoric reference)、因果推論、摘取大意主旨、及精緻化。其中，「摘取大意」是閱讀理解很重要的目標，因此本研究的命題架構中，也把「摘取大意」獨立為一個命題成份，其題幹型式如：「這篇文章主要的焦點是什麼？」，「以下哪一個題目最能表達文章的主要內容？」。至於「推論理解」，則以張雅如與蘇宜芬 (2003) 從各個推論分類架構所整理出交集最多的核心成份，即指稱、因果推論、及精緻化進行命題。「指稱推論」的題幹型式如：「在這篇文章裡，『叮囑秀才幫他寫邀請函』句中的『他』指的是誰？」。「因果推論」的題幹型式如：「根據這篇文章的敘述，以下哪一項是縣官最後氣得臉紅脖子粗的原因？」。「精緻化推論」的題幹型式如：「當秀才聽了縣官對他說：『你這個兩腳書櫥，我看你還是另謀高就吧！』，以下哪一項最能符合秀才的心情？」。「摘取大意」及「推論理解」的題目都無法直接在文章的字裡行間直接找到答案，讀者必須根據文章訊息進行推論，才能得到答案，這是「摘取大意」與「推論理解」不同於「字面理解」之處。所以，綜合言之，本研究據以編製題目的閱讀理解成份主要為：字彙觸接、字面理解、摘取大意、及推論理解。

綜上所述，本研究主要目的為編製適用於小學四、五、六年級學生的多複本閱讀理解測驗，提供一套瞭解學生閱讀理解能力發展情形的工具，也可用於閱讀低成就學生的補救教學成效評估，作為介入反應之決定。題目型式為測量文章理解的題組。每套複本都有四個題組，這四個題組的文章分別是：短記敘文、短說明文、長記敘文、長說明文。每個題組中的題目所測量的閱讀理解成份包括：字彙觸接、字面理解、摘取大意、及推論理解。關於本測驗的編製過程，則進一步說明如下。

## 方法

### 一、研究對象

**(一) 第一次預試**

第一次預試主要為評估每篇文章各道題目的難度、鑑別度，做為選題依據。第一次預試，在台灣北、中、南區依不同的都會程度共選 6 所學校施測，每篇文章至少有 180 位學生的作答資料，作為項目分析之用。施測人數及區域分佈如表 3 所示。

表 3 第一次預試各年級取樣人數摘要表

區域	台北市	新北市	新竹縣	台南市	總計
四年級	100	126	124	133	483
五年級	101	137	148	132	518
六年級	109	134	136	135	514
總計	310	397	408	400	1515

**(二) 第二次預試**

第二次預試主要是為瞭解同一年級的複本之間一致性。施測人數及區域分佈如表 4 所示。每份複本均至少有 160 位學生受測。

表 4 第二次預試各年級取樣人數摘要表

區域	新北市	宜蘭縣	基隆市	台南市	總計
四年級	70	65	69	137	341
五年級	66	70	70	135	341
六年級	68	63	69	137	337
總計	204	198	208	409	1019

**(三) 第三次預試**

第三次預試為針對第二次預試中，難以透過刪題調整難易度和鑑別度的五、六年級各兩個複本，重新調整文章，並從五、六年級各挑一份不需調整題組的複本加入第三次預試做為對照，以瞭解調整後的五、六年級各兩個複本之難度是否接近其他複本。施測人數及區域分佈如表 5 所示，每位學童均需在三個複本作答。

表 5 第三次預試各年級取樣人數摘要表

區域	台北市	新北市	宜蘭縣	台南市	高雄市	總計
五年級	28	35	30	16	33	142
六年級	34	31	34	31	35	165
總計	62	66	64	47	68	307

**(四) 信度研究及常模取樣**

本研究經過三次預試後才確定正式題本，由於施測經費上的限制，因此，以常模樣本的資料進行信度分析。本研究於 100 年 9、11、1 月，及 101 年 2、4、6 月，在全國北、中、南各區共選取 3600 位學童進行施測。每個複本在每一個施測月份都至少有 100 位學童的施測資料。為增加學校配合施測意願，以及提升學生施測動機，本研究與學校採合作方式，於施測一年後提供學生和學校第一次測驗結果做為參考。

本研究取樣採用叢集取樣，先依據全國人口比率，分北一、北二、中、南、東，和離島等六區，然後在北、中、南三區再依據城鄉選學校，各校再依據人口比率選班級，原則上各年級抽 1~2 班。東區由於人數太少，不再細分區域。各區各年級計畫抽樣人數如表 6，施測有效人數如表 7，有效人數樣本與計畫抽樣比率之差距如表 8。

表 6 各區各年級計畫抽樣人數

區域	G4			G5			G6		
	人口	人口 比率	預計 抽樣人數	人口	人口 比率	預計 抽樣人數	人口	人口 比率	預計 抽樣人數
北一	75,562	0.2748	330	79,024	0.2768	332	75,219	0.2770	332
北二	53,004	0.1927	231	54,089	0.1894	227	51,341	0.1890	227
中部	66,229	0.2408	289	68,264	0.2391	287	65,728	0.2420	290
南部	71,978	0.2617	314	75,427	0.2642	317	70,988	0.2614	314
東區	6,434	0.0234	28	6,901	0.0242	29	6,626	0.0244	29
離島	1,808	0.0066	8	1,816	0.0064	8	1,682	0.0062	7
合計	275,015	1.0000	1200	285,521	1.0000	1200	271,584	1.0000	1199

註：預計抽樣人數為 1200 人乘以各區各年級人口比率。

資料來源：教育部統計處 99 學年度縣市統計指標「各縣市國小概況統計」

([http://www.edu.tw/files/site\\_content/B0013/99citye.xls](http://www.edu.tw/files/site_content/B0013/99citye.xls))。

表 7 實際施測和有效人數

地區	人數	四年級	五年級	六年級
北一	施測人數	330	371	323
	有效人數	109	155	134
北二	施測人數	186	229	216
	有效人數	106	157	113
中部	施測人數	317	318	342
	有效人數	179	164	203
南部	施測人數	272	278	304
	有效人數	156	192	175
東部	施測人數	21	17	24
	有效人數	16	10	7
總計	施測人數	1126	1213	1209
	有效人數 (%)	566 (50.27%)	678 (55.89%)	632 (52.27%)

註：有效人數乃指六次受測均有效作答的學生，若學生資料有下列情況，則不納入分析：

1. 六次當中至少有一次以上未接受測驗，例如：在某次施測日請假，在學年中轉入或轉出，該班（或該校）無法於某次施測月份安排施測...等。



2. 檢驗題未通過，或因身心障礙因素經導師反映資料不宜納入計分者。

表 8 實際樣本比率

區域	G4			G5			G6		
	有效 人數	人口 比率	與抽樣 比率相差	有效 人數	人口 比率	與抽樣 比率相差	有效 人數	人口 比率	與抽樣 比率相差
北一	109	0.193	-0.0818	155	0.229	-0.0478	134	0.212	-0.065
北二	106	0.187	-0.0057	157	0.232	0.0426	113	0.179	-0.01
中部	179	0.316	0.0752	164	0.242	0.0029	203	0.321	0.079
南部	156	0.276	0.0143	192	0.283	0.0188	175	0.277	0.0156
東區	16	0.028	0.0046	10	0.015	-0.0092	7	0.011	-0.0134
合計	566	1.000	0.0066	678	1.000	0.0073	632	1.0000	0.0062

(五) 效度研究

本研究以「國民小學閱讀理解篩選測驗版本 A」(柯華葳、詹益綾, 2006) 為效標, 對「閱讀理解成長測驗」四、五、六年級的複本 1 進行效標關連效度的分析。研究對象取自新北市, 共有 149 位四年級學生, 144 位五年級學生, 及 146 位六年級學生。

再者, 由於詞彙能力是閱讀理解的基礎, 所以本研究也於常模樣本在 100 年 9 月接受本測驗時, 同時蒐集學生在「詞彙成長測驗」(洪儷瑜、陳心怡、陳柏熹、陳秀芬, 2014) 的得分資料, 進行相關分析。常模樣本的取樣方式及人數如前段所述。

二、測驗編製過程

(一) 文章選擇與題目編寫

本測驗以瞭解學生文章理解能力為主, 每套複本都有四個題組, 每個題組都有一篇文章及根據這篇文章所命題的閱讀理解題目。四個題組的文章分別是短記敘文、短說明文、長記敘文、長說明文。關於短篇、長篇、記敘文、說明文的界定及字數, 請見表 9。

本研究在命題前, 先為每個年級挑選短篇記敘文、長篇記敘文、短篇說明文、長篇說明文各 8 篇, 其中 4 篇為學生所熟悉的主題, 另 4 篇為學生較不熟悉的主題, 所以每個年級有 32 篇文章。

關於命題架構, 預試前, 每篇文章均編寫字彙觸接題 5 題, 字面理解題 3 題, 摘要主旨題 3 題, 推論理解題 3 題, 故每篇文章有 14 道題目。預試後, 每篇文章保留的題數為: 字彙觸接題 4 題, 字面理解題 2 題, 摘要主旨題 2 題, 推論理解題 2 題, 故每篇文章有 10 道題目。

表 9 文章長度與文體之界定

	記敘文	說明文
定義	1. 純描述式：描述景物、描述情境等。 2. 故事體：有主角、有故事情節。	1. 連續性或時間順序：說明事件的始末、程序，例如：法印戰爭、細胞分裂等。 2. 條列式：用於解釋物品或事件的特徵，例如：生物課本條列爬蟲類的特徵、體溫、再生能力、飲食習慣等等。 3. 比較和對比：比較相似點或相異點。 4. 因果關係：描述事件的原因，解釋引起的原因或影響。 5. 問題解決：討論一個問題，想出可能的解決方法等。
短文	200~300字	200~300字（4年級為200~250字）
長文	500字以上	450字以上（4年級為400字以上）

### （二）第一次預試

第一次預試主要為評估每篇文章各道題目的難度、鑑別度，做為選題依據。本研究在預試部分的項目分析，主要是依古典測驗理論分析每道題目的難度、鑑別度。第一次預試，在台灣北、中、南區依不同的都會程度共選 6 所學校施測，每篇文章至少有 180 位學生的作答資料，作為項目分析之用。

根據預試分析結果，選題標準如下：文章的題目整體通過率在 .40~.60 之間，一致性在 .60 以上的題組才予保留。根據此標準，於各年級的短篇記敘文、長篇記敘文、短篇說明文、長篇說明文各挑出較佳的 6 個題組，且每個題組保留字彙觸接題 4 題，字面理解題 2 題，摘要主旨題 2 題，推論理解題 2 題，故每篇文章有 10 道題目。

然後，把每個年級所選出的正式題組，進一步組成 6 個複本，每份複本均有短篇記敘文、長篇記敘文、短篇說明文、長篇說明文各一篇，每篇文章有 10 道題目，因此每份複本有 40 道題。

### （三）第二次預試

第二次預試主要是為瞭解各複本內試題的內部一致性，以及複本之間是否等值。每份複本均至少有 160 位學生受測。第二次預試各年級各複本的題目難度、鑑別度、Cronbach  $\alpha$  係數、折半信度係數，受試者得分的平均數、標準差，如表 10 所示。根據受試者在各複本得分的平均數，除了五、六年級各有一個複本（R51、R65）與其他複本的難度差距較大，四年級的六個複本難易度均相當，Cronbach  $\alpha$  係數均在 .80 以上。

表 10 第二次預試各年級各複本的題目難度、鑑別度和信度係數

版本	人數 (N)	信度		平均數與標準差		各題 難度	各題 鑑別度
		ALPHA	折半信度	平均數	標準差		
四年級	R41	170	.801	.697	26.58		.37~.98 .068~.456
	R42	168	.830	.737	26.43	6.484	.26~.93 .084~.542
	R43	162	.807	.777	25.60	6.173	.13~.91 -.107~.526
	R44	161	.861	.846	25.09	7.254	.20~.89 -.134~.561
	R45	160	.862	.807	24.29	7.429	.32~.84 -.018~.625
	R46	167	.854	.841	24.54	7.315	.29~.80 -.024~.589
五年級	R51	160	.849	.672	22.35	7.100	.26~.88 -.014~.575
	R52	164	.861	.789	26.38	7.098	.12~.88 -.092~.593
	R53	164	.870	.837	24.72	7.377	.20~.90 -.078~.623
	R54	165	.841	.771	23.78	6.973	.23~.85 .076~.591
	R55	161	.873	.821	23.45	7.661	.19~.85 .150~.594
	R56	160	.848	.750	24.25	7.160	.24~.88 .022~.535
六	R61	168	.835	.740	25.81	6.541	.29~.92 .050~.566
	R62	167	.841	.792	26.26	6.496	.10~.91 -.247~.494

年	R63	163	.806	.796	25.93	5.980	.19~.95	-.181~.539
級	R64	164	.814	.787	25.18	6.465	.37~.91	.027~.457
	R65	164	.837	.783	23.74	6.960	.27~.82	-.104~.544
	R66	168	.831	.762	24.38	6.593	.07~.92	-.283~.613

#### (四) 第三次預試

第三次預試是針對第二次預試中，難度差距較大，且難以透過修題調整難易度和鑑別度的複本（即 R51、R52 及 R62、R65）重新調整文章題組，並從五、六年級複本中，再挑一份不需調整題組的複本（即 R56、R66），加入第三次預試做為對照，以瞭解調整後的 R51、R52 及 R62、R65 之難度是否接近其他四個複本。

根據第三次預試資料，五、六年級各三個複本的題目難度、鑑別度和信度分析結果如表 11 所示。其中五年級調整過的複本比第二次預試時不理想，所以決定以第二次預試時的題本為正式題本。至於六年級，調整過的複本比第二次預試時理想，因此六年級以第三次預試調整過的題本為正式題本。但是，R62 第 2 題及 R65 第 33 題的鑑別度不佳，經分析後，發現其錯誤選項誘答力太高，所以將這兩題修改為難度很低的題目，做為檢驗題，以瞭解學生是否認真作答。由於這個因素，所以每個年級每份複本中，也都挑出鑑別度相對較低的兩題，也就是鑑別度為負值或接近 0 的題目，修改為難度很低的題目，做為檢驗題。因此，最後正式定題的題本，每個複本依然是 40 題，但有兩題檢驗題不計分，需計分的題目為 38 題。兩題檢驗題的功能主要是做為判斷受試者是否認真作答之用，兩道檢驗題都需答對，該份作答資料才被視為有效資料。原始分數的計分方式為：答對一題得一分，滿分為 38 分。

表 11 第三次預試五、六年級三個複本的題目難度、鑑別度和信度係數

版本	人數 (N)	信度		平均數與標準差		各題 難度	各題 鑑別度	
		ALPHA	折半信度	平均數	標準差			
五年級	R51	137	.849	.841	24.36	6.803	.19~.87	.026~.561
	R52	137	.923	.933	20.88	9.352	.28~.74	-.080~.850
	R56	138	.919	.923	19.18	9.249	.25~.70	.032~.750
六年級	R62	163	.833	.820	24.72	6.428	.31~.91	-.135~.658
	R65	161	.841	.846	25.93	6.365	.33~.93	.068~.657
	R66	161	.845	.851	24.83	6.721	.35~.91	.078~.520

此外，由於考慮到本測驗也可能做為閱讀理解能力落後一個年級以上的低成就學生補救教學效果評估工具，為顧及學生的感受，所以四、五、六年級的題本封面分別以 A、B、C 做為代碼。因每個年級都有 6 個複本，所以四年級第一個複本就以 A1 為代碼，四年級第二個複本就以 A2 為代碼，依此類推。

## 結果與討論

茲將信度研究、效度研究、及常模樣本的分析結果分述如下，並進行討論。

### 一、信度研究

如前所述，本研究經過了三次預試，才確定正式題本。由於施測經費所限，因此以常模樣本的資料進行信度分析。為了建立各年級的成長曲線常模，本研究於 100 學年度上、下學期取六個時間點，共施測六次。上學期第一次為 9 月底至 10 月初，第二次為 11 月中旬至下旬，第三次於 1 月中旬之前完成施測。下學期的施測時間，第四次是 2 月下旬至 3 月中，第五次是 4 月下旬至 5 月初，第六次為 6 月初至 6 月底。

表 12 至表 17 說明第一期次至第六期次，各年級各複本的 Cronbach's  $\alpha$  信度係數、折半信度係數，試題難度範圍、及得分平均數等資料。每個複本的 Cronbach's  $\alpha$  信度係數及折半信度係數多在 .8 以上，顯示每個複本的試題具有良好的內部一致性。另外，每個年級內各複本平均數的差距約為 0~3 分之間，各複本之間的難度也頗為接近。

表 12 第一期次各年級各複本信度分析資料、試題難度範圍及得分平均數

版本	人數 (N)	信度		平均數與標準差		試題 難度	
		ALPHA	折半信度	平均數	標準差		
四年級	A1	191	.827	.843	23.50	6.399	.35~.97
	A2	180	.854	.844	21.34	7.111	.24~.92
	A3	157	.833	.821	23.25	6.552	.13~.83
	A4	150	.805	.844	22.06	6.272	.31~.83
	A5	178	.807	.788	20.38	6.361	.19~.78
	A6	179	.847	.876	20.72	7.163	.27~.77
五年級	B1	195	.806	.835	22.45	6.193	.23~.86
	B2	181	.866	.854	24.81	7.117	.33~.90
	B3	184	.800	.783	22.05	5.992	.23~.91
	B4	178	.804	.796	21.20	6.396	.38~.77
	B5	182	.836	.822	21.09	6.800	.22~.81
	B6	195	.817	.841	21.50	6.384	.20~.93
六年級	C1	198	.828	.831	23.40	6.499	.29~.89
	C2	190	.771	.758	23.65	5.569	.23~.94
	C3	181	.771	.777	24.71	5.487	.20~.94
	C4	181	.843	.847	22.25	6.906	.31~.87
	C5	202	.784	.767	23.35	5.774	.13~.95
	C6	169	.840	.841	22.33	6.873	.33~.88

表 13 第二期次各年級各複本信度分析資料、試題難度範圍及得分平均數

版本	人數 (N)	信度		平均數與標準差		試題 難度	
		ALPHA	折半信度	平均數	標準差		
四年級	A1	182	.836	.852	21.25	6.671	.269~.901
	A2	151	.850	.840	23.32	6.915	.278~.927
	A3	142	.858	.882	22.36	7.152	.127~.810
	A4	117	.817	.880	21.32	6.511	.299~.786
	A5	160	.824	.842	20.68	6.674	.231~.781
	A6	168	.845	.839	20.92	7.086	.244~.762
五年級	B1	190	.795	.797	21.51	6.066	.195~.821
	B2	167	.866	.885	24.25	7.205	.353~.862

年級	B3	179	.848	.868	23.34	6.628	.201~.905
	B4	145	.842	.825	21.06	7.024	.297~.772
	B5	162	.863	.841	20.12	7.398	.247~.815
	B6	172	.806	.858	21.15	6.298	.203~.895
六年級	C1	186	.827	.864	22.63	6.551	.269~.849
	C2	166	.821	.826	22.28	6.389	.265~.916
	C3	162	.744	.788	24.42	5.176	.136~.951
	C4	165	.838	.838	22.68	6.810	.345~.891
	C5	178	.821	.837	23.44	6.290	.135~.927
	C6	185	.833	.857	23.22	6.667	.346~.903

表 14 第三期次各年級各複本信度分析資料、試題難度範圍及得分平均數

版本	人數 (N)	信度		平均數與標準差		試題 難度	
		ALPHA	折半信度	平均數	標準差		
四年級	A1	144	.864	.876	21.48	7.306	.264~.896
	A2	142	.888	.908	22.13	7.962	.324~.845
	A3	173	.841	.852	25.47	6.415	.150~.884
	A4	173	.849	.843	21.81	7.052	.283~.809
	A5	164	.855	.859	19.76	7.271	.201~.732
	A6	150	.848	.865	20.98	7.175	.260~.753
五年級	B1	145	.864	.885	21.26	7.309	.172~.793
	B2	146	.857	.843	22.64	7.230	.397~.815
	B3	185	.849	.857	23.03	6.702	.222~.908
	B4	204	.859	.885	23.20	7.228	.397~.824
	B5	164	.856	.857	19.88	7.313	.274~.756
	B6	160	.867	.892	18.93	7.540	.256~.794
六年級	C1	153	.847	.878	23.73	6.821	.288~.876
	C2	158	.815	.853	20.53	6.472	.259~.854
	C3	149	.859	.853	23.66	6.943	.181~.872
	C4	210	.819	.841	22.67	6.463	.381~.895
	C5	160	.839	.848	23.55	6.653	.100~.906
	C6	161	.840	.863	23.04	6.790	.366~.919

表 15 第四期次各年級各複本信度分析資料、試題難度範圍及得分平均數

版本	人數 (N)	信度		平均數與標準差		試題 難度	
		ALPHA	折半信度	平均數	標準差		
四年級	A1	192	.837	.858	23.16	6.764	.307~.984
	A2	176	.838	.840	23.79	6.857	.233~.972
	A3	131	.877	.897	24.44	7.745	.153~.916
	A4	174	.844	.864	25.85	6.832	.316~.937
	A5	186	.854	.858	22.04	7.294	.263~.984
	A6	150	.860	.902	21.89	7.600	.300~.887
五	B1	189	.869	.890	22.16	7.636	.222~.937

年級	B2	172	.875	.872	24.73	7.704	.355~.872
	B3	156	.822	.871	23.97	6.403	.263~.981
	B4	189	.876	.897	24.17	7.776	.328~.963
	B5	199	.878	.879	24.29	7.710	.251~.965
	B6	170	.868	.893	21.16	7.714	.229~.959
六年級	C1	191	.832	.841	25.63	6.521	.272~.927
	C2	195	.822	.819	21.96	6.409	.200~.923
	C3	95	.872	.879	23.93	7.659	.189~.853
	C4	206	.844	.852	23.49	7.056	.393~.981
	C5	238	.822	.834	25.50	6.368	.126~.992
	C6	145	.839	.858	25.48	6.792	.352~.938

表 16 第五期次各年級各複本信度分析資料、試題難度範圍及得分平均數

版本	人數 (N)	信度		平均數與標準差		試題 難度	
		ALPHA	折半信度	平均數	標準差		
四年級	A1	164	.860	.886	21.68	7.202	.305~.890
	A2	187	.849	.834	22.16	7.017	.294~.925
	A3	167	.883	.868	23.16	7.592	.096~.856
	A4	125	.866	.894	22.71	7.370	.304~.808
	A5	178	.849	.853	22.39	7.005	.247~.820
	A6	181	.865	.860	20.31	7.645	.326~.696
五年級	B1	171	.826	.840	21.12	6.659	.211~.801
	B2	182	.896	.893	21.81	8.359	.324~.764
	B3	168	.849	.872	22.12	6.898	.274~.869
	B4	162	.833	.856	21.67	6.821	.346~.778
	B5	191	.879	.897	22.26	7.694	.230~.859
	B6	202	.865	.882	22.77	7.296	.223~.896
六年級	C1	151	.832	.856	23.80	6.497	.272~.907
	C2	190	.817	.797	22.44	6.311	.226~.932
	C3	180	.846	.860	23.58	6.719	.222~.894
	C4	158	.860	.873	19.56	7.440	.297~.778
	C5	201	.871	.872	22.70	7.410	.189~.900
	C6	190	.869	.881	23.48	7.434	.421~.921

表 17 第六期次各年級各複本信度分析資料、試題難度範圍及得分平均數

版本	人數 (N)	信度		平均數與標準差		試題 難度	
		ALPHA	折半信度	平均數	標準差		
四年級	A1	189	.876	.903	23.40	7.761	.312~.979
	A2	161	.894	.906	23.19	8.345	.280~.981
	A3	141	.849	.859	25.38	6.894	.113~.894
	A4	131	.876	.890	24.94	7.707	.305~.893
	A5	116	.851	.868	23.16	7.185	.233~.974
	A6	154	.869	.864	23.63	7.740	.325~.981
五	B1	203	.821	.804	25.46	6.412	.217~.956

年 級	B2	167	.909	.912	22.51	9.055	.317~ .832
	B3	180	.865	.870	23.15	7.425	.294~ .911
	B4	170	.873	.867	23.63	7.790	.347~ .918
	B5	155	.867	.902	22.91	7.598	.219~ .916
	B6	196	.880	.877	23.88	7.895	.219~ .964
六 年 級	C1	174	.855	.849	26.03	6.980	.264~ .885
	C2	142	.831	.808	22.42	6.626	.218~ .901
	C3	191	.851	.854	24.74	7.003	.157~ .906
	C4	189	.838	.852	22.39	6.936	.323~ .899
	C5	119	.875	.902	20.91	7.835	.101~ .941
	C6	171	.892	.914	23.45	8.431	.363~ .854

## 二、效度研究

本研究以「國民小學（二至六年級）閱讀理解篩選測驗」（柯華葳、詹益綾，2006）為效標，分析四、五、六年級學生在「國民小學閱讀理解篩選測驗版本 A」的得分與本測驗複本 A1（四年級複本 1）、B1（五年級複本 1）、C1（六年級複本 1）得分的相關，結果兩個測驗在四年級的相關係數為 .67，五年級的相關係數為 .59，六年級的相關係數為 .75，均達 .001 顯著水準。本測驗與「國民小學（二至六年級）閱讀理解篩選測驗」為中度相關，一方面可能是因本測驗採用的是文章理解的題組，但「國民小學（二至六年級）閱讀理解篩選測驗」，有二分之一到三分之二的題目是屬於「部份處理」的題目，也就是測量對一個句子裡的某個字詞，或是對一個句子整體理解的題目。即使「本文處理」題的閱讀（一）、閱讀（二）題組，也是 100 字左右的短文。根據 Keenan 等人（2008）的研究，以一、兩個句子的短題測閱讀理解，得分的總變異被字詞辨識能力解釋的程度，往往高於能被理解能力解釋的比重；但以較長文章設計的題組測閱讀理解，得分的總變異能被理解能力解釋的比重，則高於被字詞辨識能力解釋的比重。另一個可能原因是：「國民小學（二至六年級）閱讀理解篩選測驗」的主要編製目的是為篩選閱讀困難的學童，為了避免難度高產生地板效應，可能選題時會避免難度高的題目；但本測驗主要的適用對象是一般的四、五、六年級學生，所以選題的難度範圍較廣，也就是每個複本都有較簡單的題目，但也有挑戰性高的題目。

此外，由於詞彙能力是閱讀理解的基礎，所以本研究也在第一期次施測時，蒐集學生在本測驗及「詞彙成長測驗」（洪儷瑜、陳心怡、陳柏熹、陳秀芬，2014）的得分資料，進行相關分析。結果四、五、六年級的相關係數分別為 .68、.65、.64，均達 .01 顯著水準。雖然，詞彙能力是閱讀理解的基礎，但閱讀理解的涵蓋更廣，由於「詞彙成長測驗」與本測驗所測量的構念雖然彼此相關，但並不同，因此兩個測驗之間具中度相關是合理的。

## 三、試題及能力參數的分析

為了使本測驗能夠分析學生的閱讀理解能力成長係數，因此本測驗需提供原始分數與每一個施測時間點之能力值的轉換常模。茲將分析過程說明如下。

### （一）試題參數

本研究採用 Embretson（1991）的成長模式對試題參數進行估計，採用 Embretson 成長模式的理由是：Embretson 的學習成長 IRT 模式（multidimensional Rasch model for learning and change, MRMLC）主要是針對階層能力逐漸累積的學習模式所提出，例如：Bloom（1956）將認知領域學習目標分成記憶、理解、應用、分析...等六項，其中，後面階段的學習需建立在前面階段的基礎之上，如下列表 18 的模式所示：

表 18 階層能力學習成長模式示例

	能力 1	能力 2	能力 3
階段 1	1	0	0
階段 2	1	1	0
階段 3	1	1	1

表中數值 1 代表具備該能力，0 代表尚未具備該能力。

由於本測驗在不同階段進行閱讀能力評量時，後面階段的能力都是奠基在前一階段的基礎上，因此能力 2 可視為階段 1 到階段 2 的能力成長，而能力 3 可視為階段 2 到階段 3 的能力成長，與 Embretson (1991) 的學習成長 IRT 模式概念較符合，因此本研究採用其模式來進行閱讀能力成長之資料分析。也就是，Embretson 成長模式將多向度 Rasch 模式延伸至縱貫的情境當中，假定第一期次的作答受到  $\theta_1$  的影響， $\theta_1$  是第一期次的初始能力，第二期次的作答則同時受到  $\theta_1$  及  $\Delta_1$  的影響，其中  $\Delta_1$  代表第二期次成長的能力。第三期次則同時受到  $\theta_1$ 、 $\Delta_1$  及  $\Delta_2$  的影響，其中  $\Delta_2$  代表第三期次成長的能力。依此類推，最後一期次（第六期次）的作答則同時受到  $\theta_1$ 、 $\Delta_1$ 、 $\Delta_2$ 、 $\Delta_3$ 、 $\Delta_4$  及  $\Delta_5$  的影響。本研究採用 ConQuest 軟體，以邊緣最大概似法進行同時估計。表 19 為各年級各版本試題參數的平均數、最大值與最小值等相關資料。

表 19 各年級各版本之試題參數平均數與最大值最小值

年級	版本	最大值	最小值	平均數	標準差
四年級	A1	1.12	-3.33	-0.27	0.95
	A2	1.39	-2.45	-0.19	0.83
	A3	2.99	-1.25	0.10	0.83
	A4	1.52	-1.08	0.22	0.66
	A5	1.46	-1.07	0.13	0.66
	A6	1.07	-0.91	0.01	0.53
五年級	B1	1.73	-1.58	0.12	0.85
	B2	0.87	-1.98	-0.49	0.71
	B3	1.69	-1.83	-0.07	0.92
	B4	1.00	-1.04	0.07	0.62
	B5	2.17	-0.93	0.39	0.73
	B6	2.00	-1.96	-0.01	0.78
六年級	C1	1.62	-1.65	-0.01	0.79
	C2	1.17	-1.89	-0.18	0.74
	C3	2.21	-2.32	-0.20	0.99
	C4	1.19	-1.72	0.02	0.72
	C5	1.26	-1.13	0.22	0.60
	C6	1.74	-1.68	0.16	0.68

試題反應理論以訊息量 (information) 來代表試題對於各種能力的受試者所能提供的測量精準度，若試題難度與受試者的能力愈相近，則能提供愈大的訊息量，亦即提供愈高的測量精準度。每一道題目的訊息量都是能力值的函數，而不同題目之間的訊息量可以相加，因此本研究將各複本所有題目的訊息函數相加，形成各複本的訊息量函數圖（橫軸為能力值，縱軸為訊息量），如圖 1 至圖 3 所示。



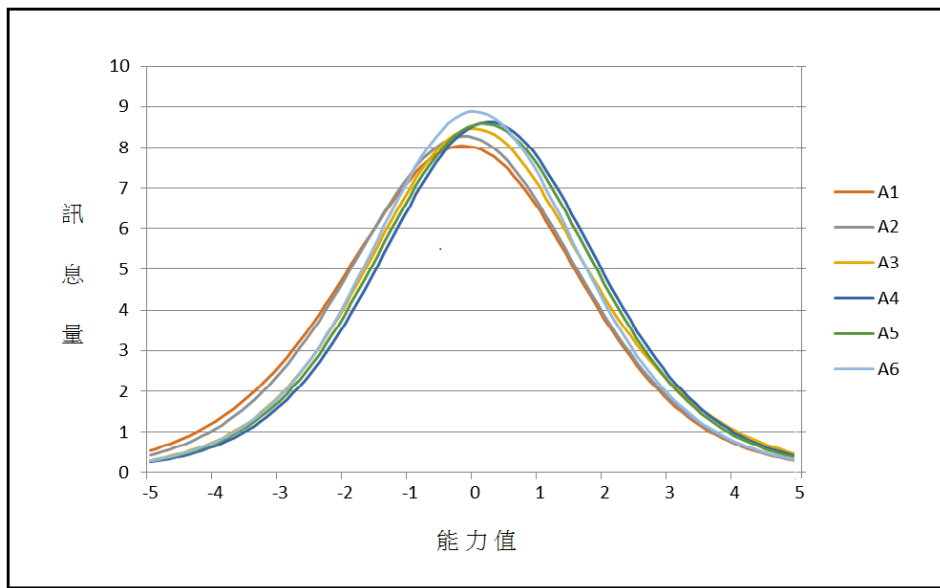


圖 1 四年級各複本訊息量

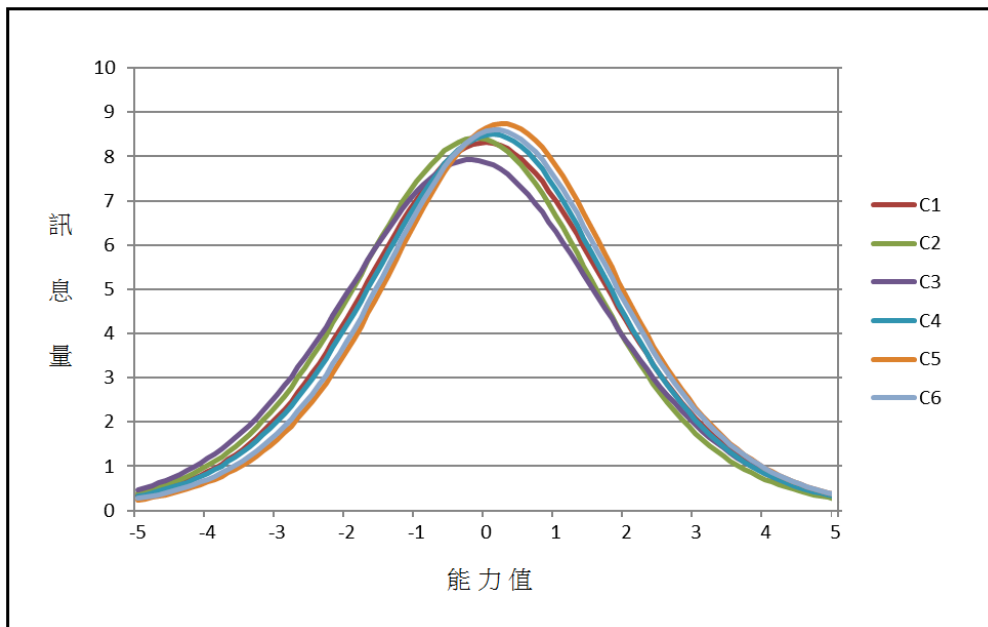


圖 2 五年級各複本訊息量

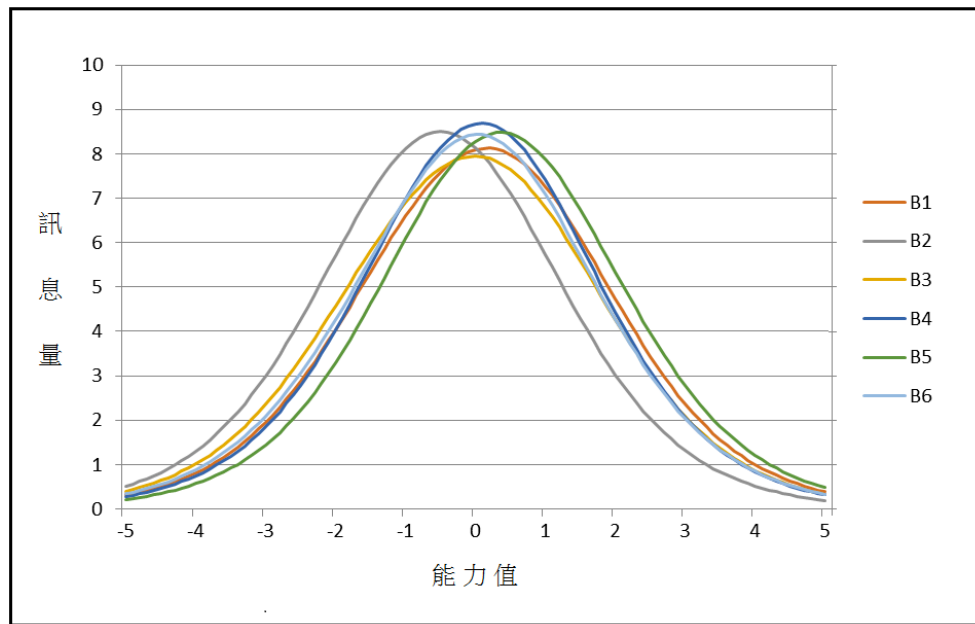


圖 3 六年級各複本訊息量

### (二) 能力參數

本研究採用 Embretson 的成長模式，以最大概似法 (Maximum Likelihood) 進行個人能力參數的估計。依據 Embretson 的成長模式，每一個學生分別能估計出第一個期次的初始能力 ( $\theta_1$ )、第二個期次成長的能力 ( $\Delta_1$ )、第三個期次的成長能力 ( $\Delta_2$ )、第四個期次的成長能力 ( $\Delta_3$ )、第五個期次的成長能力 ( $\Delta_4$ ) 及第六個期次的成長能力 ( $\Delta_5$ )。因此，第二個期次的能力 ( $\theta_2$ ) 便等於  $\theta_1 + \Delta_1$ ，第三個期次的能力 ( $\theta_3$ ) 便等於  $\theta_1 + \Delta_1 + \Delta_2$ ，依此類推，表 20 為各年級各期次的初始及成長能力對照表。由於呈現數值皆採四捨五入，每一橫列數值的總和與最右欄的能力值可能稍有差異。

表 20 各年級各期次初始 ( $\theta_1$ ) 及成長能力 ( $\Delta$ ) 對照表

四年級							
期次	1st ( $\theta_1$ )	2nd ( $\Delta_1$ )	3rd ( $\Delta_2$ )	4th ( $\Delta_3$ )	5th ( $\Delta_4$ )	6th ( $\Delta_5$ )	能力值
01	0.53						0.53
02	0.53	-0.01					0.52
03	0.53	-0.01	0.03				0.55
04	0.53	-0.01	0.03	-0.01			0.54
05	0.53	-0.01	0.03	-0.01	0.10		0.65
06	0.53	-0.01	0.03	-0.01	0.10	0.04	0.69
五年級							
期次	1st ( $\theta_1$ )	2nd ( $\Delta_1$ )	3rd ( $\Delta_2$ )	4th ( $\Delta_3$ )	5th ( $\Delta_4$ )	6th ( $\Delta_5$ )	能力值
01	0.57						0.57
02	0.57	-0.11					0.46
03	0.57	-0.11	0.10				0.56
04	0.57	-0.11	0.10	0.04			0.60
05	0.57	-0.11	0.10	0.04	0.06		0.66

06	0.57	-0.11	0.10	0.04	0.06	0.02	0.68
六年級							
期次	1st ( $\theta_1$ )	2nd ( $\Delta_1$ )	3rd ( $\Delta_2$ )	4th ( $\Delta_3$ )	5th ( $\Delta_4$ )	6th ( $\Delta_5$ )	能力值
01	0.71						0.71
02	0.71	0.01					0.72
03	0.71	0.01	0.08				0.79
04	0.71	0.01	0.08	0.00			0.79
05	0.71	0.01	0.08	0.00	0.05		0.84
06	0.71	0.01	0.08	0.00	0.05	-0.05	0.79

### (三) 成長曲線

本研究採用潛在成長曲線模式，對學生各期次估計出來的能力值進行潛在成長曲線的估計。本研究將所有學生各期次的能力值估計出來之後，第二階段再將這些能力值套入潛在成長模式當中，使用 Mplus 6.0 進行估計，以具強韌標準誤的最大概似法 (MLR) 進行分析，以獲得成長曲線的資料。各年級閱讀理解能力的成長曲線估計結果如表 21 至表 23 所示。各年級的閱讀理解成長理論曲線與實際資料的結果如圖 4 至圖 6 所示。

表 21 四年級閱讀理解成長曲線估計

成長因子	估計值	標準誤
成長截距	0.505	0.031
成長斜率	0.058	0.013
截距與斜率共變數	0.024	0.011
適配指標		
RMSEA	0.039	
CFI	0.993	
TLI	0.994	
SRMR	0.024	

表 22 五年級閱讀理解成長曲線估計

成長因子	估計值	標準誤
成長截距	0.519	0.027
成長斜率	0.061	0.012
截距與斜率共變數	0.069	0.009
適配指標		
RMSEA	0.093	
CFI	0.944	
TLI	0.947	
SRMR	0.110	

表 23 六年級閱讀理解成長曲線估計

成長因子	估計值	標準誤
成長截距	0.717	0.027
成長斜率	0.045	0.010
截距與斜率共變數	0.027	0.008
適配指標		
RMSEA	0.011	
CFI	0.997	

TLI	0.998
SRMR	0.021

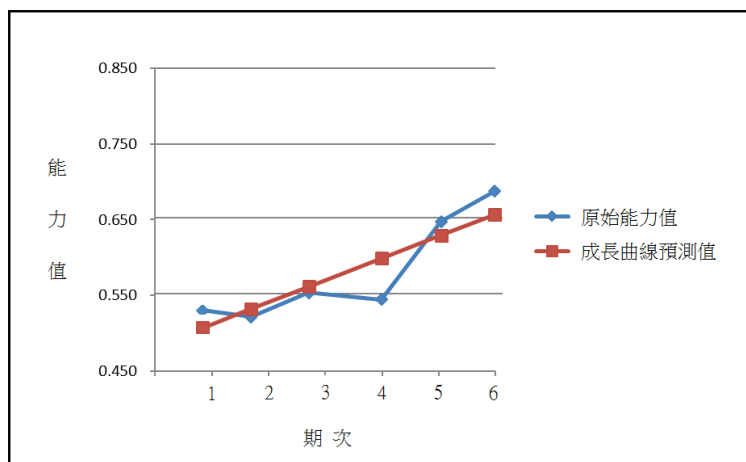


圖 4 四年級閱讀理解原始能力值及成長曲線預測值

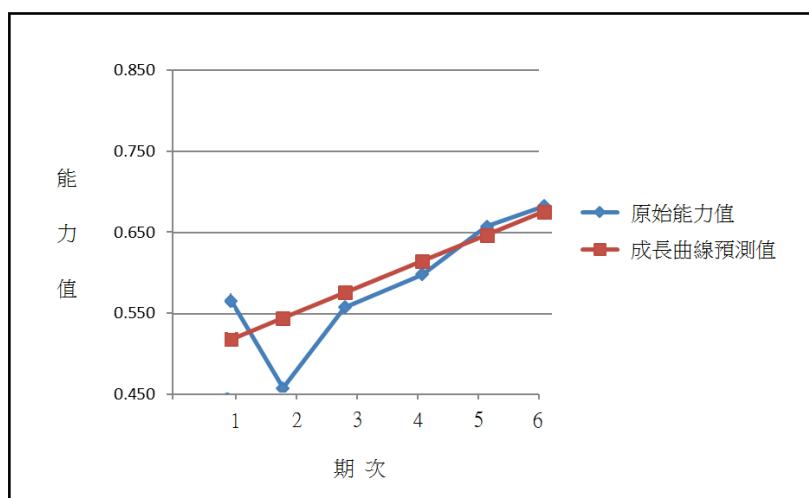


圖 5 五年級閱讀理解原始能力值及成長曲線預測值

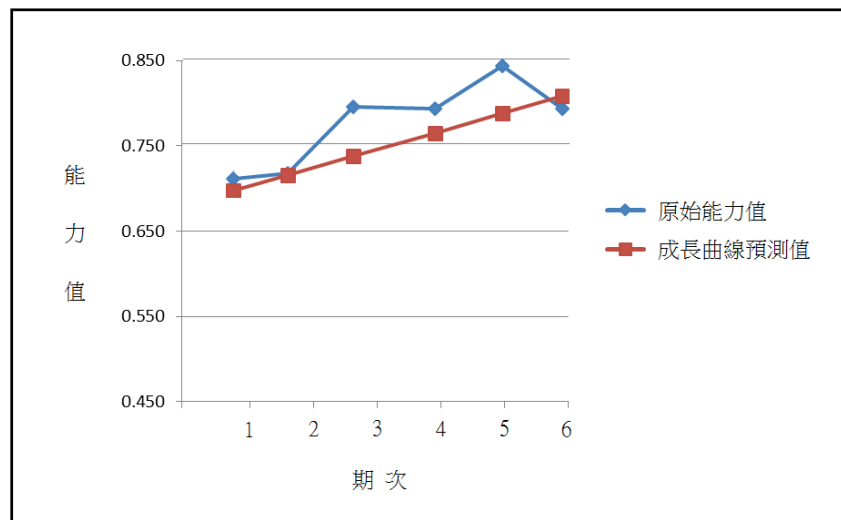


圖 6 六年級閱讀理解原始能力值及成長曲線預測值

依據 Hu 與 Bentler (1999) 的建議標準, CFI 及 TLI 應大於 0.95, RMSEA 應小於 .06, SRMR 則應小於 .08, 因此本研究中四年級與六年級閱讀理解能力的潛在成長模型達到適配的標準。五年級的模型則是接近但未完全達到 Hu 與 Bentler (1999) 所建議的適配標準。從表 19 及圖 5 可知, 五年級常模樣本的能力值雖在第二期次下降, 但三至六期次仍大致呈現成長的趨勢。

綜合以上分析結果, 本測驗具有良好的信度與效度, 而且難度適中, 是一套適合用來瞭解一般四、五、六年級學生在學年內閱讀理解能力成長情形的測驗。

## 結論與建議

綜合前述信度、效度及成長曲線模型分析結果, 本測驗具有良好的信、效度, 且難度適中, 適合用來做為追蹤瞭解四、五、六年級學生在學年內閱讀理解能力成長情形的工具。關於本測驗在教育應用上的建議與限制, 以下將做進一步的說明。

### 一、教育應用上的建議

(一) 本測驗適用於國小四至六年級學生, 或實足年齡高於小四至小六這段範圍, 但閱讀理解能力相當於小四至小六水準的學生。

(二) 本測驗非計時測驗, 但一般而言, 施測時間約在 35~45 分鐘內即可完成, 如果學生作答時間過長或過短, 應進一步了解情形, 以免影響結果。

(三)本測驗可以利用複本進行多次的評量，而不會有題目練習效果的情形。如果需持續在不同的時間點，做閱讀理解成長監控 (progress monitor)，不論監控時間是一學期或一學年，都建議至少要有三個施測點，而且監控期間不宜低於一個學期。

(四)為了能監控學生閱讀理解能力的成長情形，便於繪出學生的閱讀理解能力成長曲線，本測驗提供能力值轉換常模。另外，根據學生在不同期次所測得的能力值，還可計算出成長係數，亦即學生在兩個時間點之間，能力的成長幅度。學校或研究人員除了可藉能力值瞭解學生在不同時間點能力成長的情形外，也可透過成長係數瞭解學生能力成長的幅度是高於或低於全國常模。所以，能力值與成長係數兩個資料所提供的成長訊息不一樣，建議學校或研究人員在進行成長監控時可運用這兩個轉換分數，兩項分數均為標準分數，可以直接做統計分析。此外，本測驗也提供百分等級常模，以便於學校瞭解學生在同年齡團體中，閱讀理解能力的相對位置。

(五)評估學生是否為低成就與低成長的雙重缺陷時，可以參考成長監控的作法，把學生的成長係數與該生所屬年級常模樣本之成長係數做比較，依此判斷該生的閱讀理解能力是否為低成長。或者，也可利用學校的年級樣本平均數，換算出學校本位之成長係數，瞭解該生相對於學校同年級學生的成長係數是否偏低。

(六)本測驗每份複本裡都有 14~16 個推論理解題 (包含摘取大意題在內)，這些題目可做為推論理解教學效果評估的工具。

## 二、應用上的限制

本測驗主要適用對象是已進入「透過閱讀學習」(read to learn)的四、五、六年級學童，對於閱讀能力尚處於「學習閱讀」(learn to read)的一至三年級學童，本測驗之結果可能會受到其識字能力的影響，難以正確評估其閱讀理解能力，所以在運用上需留意。此外，本測驗在編製時，沒有設計跨年級題本的共同題進行跨年級題本的等化，因此，不同年級版本的分數不能進行跨年級的比較，這個部分值得未來研究繼續發展。

## 參考文獻

- 王木榮、董宜俐 (2006): 國民小學閱讀理解測驗。台北: 心理。[Wang, M. R., & Dong, Y. L. (2006). *Reading comprehension test for elementary school students*. Taipei, Taiwan: Psychological Publishing.]
- 林寶貴、錡寶香 (2000): 中文閱讀理解測驗之編製。特殊教育研究學刊, 19, 79-104。[Lin, B. G., & Chi, P. H. (2000). The development of reading comprehension test. *Bulletin of Special Education, 19*, 79-104.]
- 柯華葳 (1999): 閱讀理解困難篩選測驗。測驗年刊, 42 (2), 1-11。[Ko, H. W. (1999). Reading Comprehension Screening Test. *Psychological Testing, 42*(2), 1-11.]
- 柯華葳、詹益綾 (2006): 國民小學 (二至六年級) 閱讀理解篩選測驗。台北: 教育部特殊教育小組。[Ko, H. W., & Chan, Y. L. (2006). *Reading comprehension screening test for second to*

- sixth graders*. Taipei, Taiwan: Department of Student Affairs and Special Education, Ministry of Education.]
- 柯華葳、詹益綾 (2006): **國民中學閱讀推理測驗**。台北: 教育部特殊教育小組。[Ko, H. W., & Chan, Y. L. (2006). *Inferential comprehension test for junior high school students*. Taipei, Taiwan: Department of Student Affairs and Special Education, Ministry of Education.]
- 洪麗瑜、陳心怡、陳柏熹、陳秀芬 (2014): **詞彙成長測驗**。台北: 中國行為科學社。[Hung, L. Y., Chen, H. Y., Chen, P. H., & Chen, S. F. (2014). *Progress monitoring test of vocabulary*. Taipei: Chinese Behavioral Science Corporation.]
- 張雅如、蘇宜芬 (2003): **國小學童推論理解測驗之編製與研究**。中國心理學會第四十二屆年會宣讀之論文。[Chang, Y. R., & Su, Y. F. (2003). *The development of inferential comprehension test for elementary school students*. Paper presented at the 42nd annual meeting of Taiwanese Psychological Association, Taipei, Taiwan.]
- Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.
- Burns, M. S., Griffin, P., & Snow, E. C. (1999). *Starting out right: A guide to promoting children's reading success*. Washington, DC: National Research Council.
- Chall, J. (1996). *Stages of reading development* (2nd ed.). Orlando, FL: Harcourt Brace & Co.
- Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, 56(3), 495-515. DOI: 10.1007/BF02294487.
- Gagne, E. D. (1985). *The cognitive psychology of school learning*. Boston, MA: Little, Brown, and Company.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55. DOI: 10.1080/10705519909540118.
- Kameenui, E., Fuchs, L., Francis, D., Good III, R., O'Connor, R., Simmons, D., Tindal, G., & Torgesen, J. (2006). The adequacy of tool for assessing reading competence: A framework and review. *Educational Researcher*, 35(4), 4-10. DOI: 10.3102/0013189X035004003.
- Keenan, J., Betjemann, R., & Olson, R. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281-300. DOI: 10.1080/10888430802132279.
- Kintsch, W. (1993). Information accretion and reduction in text processes: inference. *Discourse Processes*, 16, 193-202. DOI: 10.1080/01638539309544837.

- Magliano, J. P., & Graesser, A. C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics*, *20*, 193-232. DOI: 10.1016/0304-422X(91)90007-C.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Associates.
- Taylor, B., Harris, L. A., Pearson, P. D., & Garcia, G. (1995). *Reading difficulties: Instruction and assessment*. New York, NY: McGraw-Hill.
- Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through on-line inference and mental operations in working memory. *Discourse Processes*, *16*, 3-34. DOI: 10.1080/01638539309544827.
- van den Broek, P., Fletcher, C. T., & Risdien, K. (1993). Investigations of inference processes in reading: A theoretical and methodological integration. *Discourse Processes*, *16*, 169-180. DOI: 10.1080/01638539309544835.

收稿日期：2015年08月24日

一稿修訂日期：2016年09月05日

二稿修訂日期：2017年01月02日

接受刊登日期：2017年01月03日

Bulletin of Educational Psychology, 2018, 49(4), 557-580

National Taiwan Normal University, Taipei, Taiwan, R.O.C.

## **The Development of Progress Monitoring Test of Reading Comprehension**



**Yi-Fen Su**

Department of Educational Psychology and  
Counseling

National Taiwan Normal University

**Li-Yu Hung**

Department of Special Education

National Taiwan Normal University

**Po-Hsi Chen**

Department of Educational Psychology and  
Counseling

National Taiwan Normal University

**Hsin-Yi Chen**

Department of Special Education

National Taiwan Normal University

The purpose of this study was to develop a set of equivalent tests for reading comprehension progress monitoring for grades four to six. The reading comprehension equivalent tests can also be used as tools for evaluating instructional effects of remedial intervention. In this test, there were six alternate-forms for each grade. Four testlets were included in each alternate-form. Each testlet had one passage and ten test items. The four passages in one alternate-form included one short narrative, one short expository, one long narrative, and one long expository texts. There were 200-300 characters in short passages, and 400-500 characters in long passages. The test items were designed to measure lexical access, literal comprehension, summarization, and inferential comprehension. There were forty items in total in one alternate-form. Regarding reliability, the Cronbach's  $\alpha$  coefficients and split-half coefficients of all alternate-forms were mostly above .80. Using "Reading Comprehension Screening Test for Elementary School Students-Version A" as criterion, the criterion validity coefficient were .67 ( $p < .001$ ), .59 ( $p < .001$ ), .75 ( $p < .001$ ) for grades 4, 5, and 6 respectively. In addition, based on the data collected from six time-points in 2011 school year, the growth model analysis revealed that the reading comprehension abilities of the fourth, fifth, and sixth graders increased from the beginning to the end in one school year. In order for teachers to monitor students' progress, this test provided not only the norm of percentile rank, but also the norm of ability value (theta score). Using the ability values between two time-points, a growth coefficient could be calculated. Several suggestions and one limitation for educational application were provided.

**KEY WORDS: Ability value, Elementary school students, Equivalent test, Growth curve, Reading comprehension**