

教育研究集刊

2025年6月，71（2），頁1-57

[https://doi.org/10.6910/BER.202506/SP_71\(2\).0001](https://doi.org/10.6910/BER.202506/SP_71(2).0001)



拾級而上：第四學習階段 國語文閱讀表現水準描述芻議

謝佩蓉

摘要

研究目的

本研究目的在於建立第四學習階段國語文閱讀表現量尺及表現水準描述，並提出範例試題使表現水準描述具象化。

主要理論或概念架構

標準設定的方法各有所長，大多數研究採用「以內容為基礎」的方法，但設定程序可能受到人為因素影響。本研究比照國際大型教育調查所採用的「量尺定錨法」，包括TIMSS和PISA均採用此方法，讓資料使用者理解量尺分數所代表的意義。

研究設計／方法／對象

採用調查研究法蒐集標準設定所需實證資料，以縱貫性研究設計測量學生國語文閱讀學習表現，再以國際大型調查採用的量尺定錨法設定表現標準並產出表

謝佩蓉，國家教育研究院測驗及評量研究中心副研究員

電子郵件：pjh@mail.naer.edu.tw

投稿日期：2024年10月24日；修改日期：2025年02月17日；接受日期：2025年06月13日

現水準描述。研究對象以107學年度和108學年度的七年級學生為母群體，107學年度為第一組、108學年度為第二組追蹤樣本。採二階段分層叢集抽樣設計，第一階段採「依比例的機率」抽取學校、第二階段針對抽中的學校進行校內班級隨機抽樣，所抽中的班級，全班學生均為樣本。「國語文閱讀」於第一組追蹤樣本七年級時分派2,803人，八年級分派2,807人；第二組追蹤樣本於七年級分派2,565人，八年級分派2,780人。評量工具為國語文閱讀素養導向電腦化線上測驗，總計344題，採部分平衡不完全區塊組成題本，信、效度證據良好。以試題反應理論之部分給分模式估計試題難度。

研究發現或結論

表現水準的四個切截點分數為400、475、550、625，並命名為M1、M2、M3及M4，撰寫各表現水準的描述，具體描寫學生的能力發展。依閱讀題組的閱讀特性，可概分為評量學生的一般閱讀和數位閱讀兩種素養。以找出訊息、理解及評價與省思三向度的閱讀認知歷程，描述達到每一水準學生所展現的兩種類型閱讀素養。整體而言，當認知複雜性逐漸提升時，試題的難度通常也會隨之增加。從本研究的實證資料可知，即便是認知複雜度較低的任務，也可能會出現令學生感到困難的題目；而認知複雜度較高的試題，可能反而較為容易。

理論或實務創見／貢獻／建議

本研究以嚴謹大規模調查之實證證據支持標準設定的有效性，並根據實證數據選取合適的範例題。所完成的表現水準描述揭示各閱讀歷程於不同水準可能涉及的能力差異，有助於教師精準掌握學生閱讀理解歷程中的困難點。建議未來研究能建構第三和第五學習階段國語文閱讀素養表現水準描述，使學習階段能力描述具備縱貫脈絡，並進一步研發涵蓋不同難度層級的數位閱讀試題，補充評價與省思M1至M3的數位閱讀表現描述，確立完整的數位閱讀素養能力階層。此外，可以表現水準描述為基礎，針對不同能力層級學生設計教學方案，發揮表現水準描述作為「學習地圖」在實務現場的應用潛力。

關鍵詞：大型評量、素養導向評量、量尺定錨法、標準設定、閱讀理解

Bulletin of Educational Research

June, 2025, 71(2), pp. 1-57

[https://doi.org/10.6910/BER.202506/SP_71\(2\).0001](https://doi.org/10.6910/BER.202506/SP_71(2).0001)

Step-by-Step Progression: A Preliminary Discussion on Performance Level Descriptors for Mandarin Reading in the Fourth Learning Stage

Pei-Jung Hsieh

Abstract

Purpose

The study aims to develop a performance scale and performance level descriptors (PLDs) for Mandarin reading in the fourth learning stage, and to provide sample items to illustrate the PLDs.

Main Theories or Conceptual Frameworks

Each standard-setting method has its own strengths. Most studies use a “content-based standard-setting method” approach, but the standard-setting process can be influenced by subjective factors. This study, therefore, adopts the “scale anchoring” method, which is commonly employed in international large-scale educational

Pei-Jung Hsieh, Associate Research Fellow, Research Center for Testing and Assessment,
National Academy for Educational Research

Email: pjh@mail.naer.edu.tw

Manuscript received: Oct. 24, 2024; Revised: Feb. 17, 2025; Accepted: Jun. 13, 2025.

assessments such as TIMSS and PISA, to help data users interpret the meaning represented by scale scores.

Research Design/Methods/Participants

A survey research method was employed to collect empirical data required for standard setting. A longitudinal study design was used to measure students' performance in Mandarin reading, and the scale anchoring method was adopted to set performance standards and generate performance descriptors. The study population consisted of 7th-grade students from the 107th and 108th academic years, with the 107th academic year as Panel 1 and the 108th academic year as Panel 2 for follow-up samples. A two-stage stratified cluster sampling design was implemented: in the first stage, schools were selected using "probabilities proportional to size" sampling; in the second stage, classes within the selected schools were randomly sampled. All students in the sampled classes were included in the study. In Panel 1, 2,803 students were assigned for the Mandarin reading assessment in 7th grade and 2,807 in 8th grade. In Panel 2, 2,565 students were assigned in 7th grade and 2,780 in 8th grade. The assessment tool was a competency-based computerized online test of Mandarin reading, consisting of 344 items. Test booklets were assembled using a partially balanced incomplete block design. The assessment provided strong evidence of reliability and validity. Item difficulty was estimated using the partial credit model of item response theory.

Research Findings or Conclusions

The four cut scores for performance levels were set at 400, 475, 550, and 625, corresponding to levels M1, M2, M3, and M4. Descriptions for each performance level were developed to portray students' developmental abilities. Reading tasks were categorized into two literacy types: general reading and digital reading. The reading cognitive processes were classified into three dimensions: locating information, understanding, and evaluating and reflecting. Performance level descriptors (PLDs) detailed the literacy skills demonstrated by students at each level for both types of reading. In general, item difficulty increased with cognitive complexity. However, the empirical data revealed that tasks with lower cognitive complexity could still challenge

students, while tasks with higher cognitive complexity might be relatively easier.

Theoretical or Practical Insights/Contributions/Recommendations

This study validates the effectiveness of standard setting with large-scale empirical data and selects appropriate sample items based on empirical evidence. The resulting performance level descriptors (PLDs) reveal the differences in reading processes across performance levels, enabling teachers to accurately identify students' difficulties in reading comprehension. Future research is recommended to develop PLDs for Mandarin reading literacy in the third and fifth learning stages, providing a longitudinal framework for PLDs across learning stages. Additionally, the development of digital reading items covering various difficulty levels is suggested to enrich the PLDs for evaluating and reflecting processes at M1 to M3 levels. Furthermore, the PLDs can serve as a “learning map” for designing instructional plans tailored to students at different performance levels, enhancing their practical application in teaching and learning contexts.

Keywords: large-scale assessments, competency-based assessment, scale anchoring, standard setting, reading comprehension

壹、緒論

一、前言

全球各國在面對21世紀多元挑戰與教育轉型需求下，紛紛啟動課程改革，普遍將「核心素養」（core competency）作為課程設計與學習願景的核心，包括美國、英國、德國、法國、澳洲、南韓與紐西蘭等國，皆已在課程改革中導入素養導向理念，強調學生應具備能夠跨情境運用知識、解決問題、與人溝通合作、進行反思與創造的能力（Gouëdard et al., 2020; New Zealand Ministry of Education, 2015）。我國教育部自2014年起實施《十二年國民基本教育課程綱要》（以下簡稱十二年國教課綱），亦明定核心素養為課程發展之主軸，強調學習不應侷限於學科知識與技能，而應重視學習與生活的結合，藉由實踐力行促進全人發展（教育部，2014，2021）。其中，「溝通互動」作為核心素養三大面向之一，強調學生在語文表達與理解上的能力建構；而閱讀素養則為其關鍵內涵，涵蓋學生在面對多元文本與複雜情境時，能否有效解讀、評鑑並運用資訊的能力。

為檢視各國教育政策與課程實施的成效，國際大型評量應運而生，藉由跨國學生學習表現資料，評估教育體系的整體效能與課程落實狀況（Wagner & Hastedt, 2022）。以經濟合作暨發展組織（Organisation for Economic Co-operation and Development, OECD）主導的「國際學生能力評量計畫」（Programme for International Student Assessment, PISA）為例，其針對15歲學生每三年進行一次跨領域素養評量，涵蓋閱讀、數學及科學三大領域，成為全球教育政策制定的重要依據。PISA除以量尺分數呈現學生表現外，更以「表現水準描述」（Performance Level Descriptors, PLDs）詮釋學生能力表現水準。PLDs用以描述學生在特定水準下所應具備的知識、理解與策略應用能力，揭示不同等級學生在表現上的具體差異（Perie, 2008）。例如，在閱讀素養領域，PISA自水準1c至水準6設定八個等級（OECD, 2020），每一水準均說明該層級學生面對特定任務類型時的反應特徵與理解深度，使得評量結果不僅止於分數比較，更能轉化為課程設計與教學調整的實證依據。

國內針對國語文領域進行表現標準設定的研究甚少，曾建銘與王暄博（2012）利用臺灣學生學習成就評量資料庫的資料，以書籤標定法（bookmark method）針對國小六年級國語文進行表現標準設定。除此之外，幾乎缺乏其他研究，顯示國內在不同學習階段的國語文閱讀表現水準描述仍有極大發展空間，亟待深入探討。十二年國教課綱第四學習階段國語文閱讀素養學習表現僅有六項，僅能整體性、概括性描寫學習表現，再加上並未對準十二年國教課綱學習表現建立量尺，難以具體描述不同能力等級的閱讀素養表現作為學習地圖，無法像生長曲線圖般提供閱讀能力成長的實證參照，亦缺乏不同能力等級的閱讀素養表現描述，作為課綱修訂參考。

二、研究目的

鑑於國內對於第四學習階段的閱讀表現水準描述極少，缺乏細緻且具實證基礎的研究。本研究目的在於建立第四學習階段國語文閱讀表現量尺及表現水準描述，並提出範例試題使表現水準描述具象化。創新性與貢獻在於：（一）透過實證數據建構第四學習階段國語文閱讀表現水準描述，並根據實證數據選取合適的範例題。（二）藉由量尺定錨法（scale anchoring）建立不同能力等級的閱讀表現標準，精細化閱讀學習表現的內容，為課綱修訂與教學評量提供基礎。本研究不僅能為現行課綱提供更精細的閱讀表現水準描述，也能協助教師針對不同能力等級的學生進行精準教學，進而提升學生的閱讀素養。

三、研究倫理審查

本研究已通過國立臺灣師範大學研究倫理中心微小風險審查，案件編號201803ES009。

貳、文獻探討

國家課程描寫的是國家對於學生學習的目標和願景，因而撰寫各學習領域的學習重點時，隨著學習進程，應蘊含著一份架構，縱向銜接該學習領域從低能力到高能力學生的學習表現。

一、閱讀領域課程綱要

美國州共同核心課程標準（Common Core State Standards, CCSS）將標準視為學習地圖，選用須具備必要性、嚴謹性、清晰性、具體性及連貫性，並符合國際基準（Common Core State Standards Initiative, 2009）。在CCSS架構下，語文素養建立K至12年級的標準，著重核心內涵為：熟悉複雜文本與學術語言，基於文本證據的閱讀、寫作、口說，透過非文學類文本建構知識（Common Core State Standards Initiative, 2022）。由此可知，語文素養不僅是技能的訓練，而是要在高中之前培養學生具備大學和職場所需的閱讀能力。CCSS論證時不僅根據學生的先備知識和經驗，亦要求學生能根據文本提出證據，隨著年級增加，訊息類文本比重大幅提升，以期學生能獨立學習，透過閱讀習得新知。

為了形塑國家課程的結構性，紐西蘭課綱將各學習領域13年的課程，切分為八個成就目標水準。以語文學習為例，紐西蘭的官方語言包括英語、毛利語及手語，日常生活以英語為主。假如學生未曾學習過某項語言，無論其年級為何，均由水準1開始學習（New Zealand Ministry of Education, 2015）。為了讓教師在教學時清楚知道學生的學習進展，紐西蘭設計閱讀、寫作及數學的學習進展架構（Learning Progression Framework, LPFs）並開發相對應的評量（The Progress and Consistency Tool, PaCT），使教師能快速掌握學生的狀態，提供最適合的差異化教學。閱讀分為「理解文本」（making sense of text）和「為不同課程目的閱讀」（reading for different curriculum-related purposes）兩種向度，提示教師「閱讀」具有不同的面向與功能，教學時除了使學生理解文本，也要考慮閱讀的功能性目的，也就是透過閱讀學習不同領域的專業知識，更重要的是，提醒教師在進行閱讀教學時是否為學生設計各面向的學習（New Zealand Ministry of Education, 2025b）。

首先，理解文本包含四項技能：使用處理系統（using a processing system）、利用文本結構和特徵知識（using knowledge of text structure and features）、詞彙知識（vocabulary knowledge）、批判性閱讀（reading critically），亦即「學習閱讀」時不僅要讀懂字詞和讀出字裡行間的隱含訊息，還要能理解作者如何架構文章，甚至要能評論文章。其次，為不同課程目的閱讀，則包括透過閱讀組織學習的思想和資訊（reading to organise ideas and

information for learning）、在資訊文本中獲取和使用資訊和思想（acquiring and using information and ideas in informational texts）及透過閱讀體驗文學（reading for literary experience），不論是英語、科學或社會科學的文本，學生都要能從中讀到作者所表達的核心概念（New Zealand Ministry of Education, 2025b）。紐西蘭教育部亦提出各項技能在各年級的學習範例，將「學習進展」具象化，作為良好的示範（New Zealand Ministry of Education, 2025a）。

反觀十二年國教課綱第三至第五學習階段的閱讀，其敘寫架構並不一致（詳見表1），第三學習階段學習表現12條、第四和第五學習階段各6條，每條涵蓋的範圍不一。以階段流暢朗讀和標點符號為例，第三學習階段以5-III-1和5-III-2兩項學習表現加以描寫，第四學習階段則合併為5-IV-1。

表1

十二年國教課綱第三至第五學習階段的閱讀學習表現

第三學習階段 （國小五、六年級）	第四學習階段 （國中七至九年級）	第五學習階段 普通型高級中等學校 （高中10至12年級）
5-III-1流暢朗讀各類文本，並表現抑揚頓挫的變化。	5-IV-1比較不同標點符號的表達效果，流暢朗讀各類文本，並表現情感的起伏變化。	5-V-1辨析文本的寫作主旨、風格、結構及寫作手法。
5-III-2理解各種標點符號的用法與表達效果。	5-IV-2理解各類文本的句子、段落與主要概念，指出寫作的目的與觀點。	5-V-2歸納文本中不同論點，形成個人的觀點，發展系統性思考，以建立論述體系。
5-III-3讀懂與學習階段相符的文本。	5-IV-3理解各類文本內容、形式和寫作特色。	5-V-3大量閱讀多元文本，探討文本如何反應文化與社會現象中的議題，以拓展閱讀視野與生命意境。
5-III-4區分文本中的客觀事實與主判斷之間差別。	5-IV-4應用閱讀策略增進學習效能，整合跨領域知識轉化為解決問題的能力。	5-V-4廣泛運用工具書及資訊網絡蒐集、分析資料，提高綜合學習和應用能力。
5-III-5認識議論文文本的特徵。	5-IV-5大量閱讀多元文本，理解重大議題內涵及其與個人生活、社會	5-V-5主動思考與探索文本的
5-III-6熟習適合學階段的摘要策略，擷取大意。		
5-III-7連結相關的知識和經驗，提出自己觀點評述文本內容。		

（續下頁）

第三學習階段 (國小五、六年級)	第四學習階段 (國中七至九年級)	第五學習階段 普通型高級中等學校 (高中10至12年級)
5-III-8運用自我提問、推論等策略，文本隱含的因果訊息或觀點。	結構的關聯性。 5-IV-6能運用圖書館(室)、科技工具，蒐集資訊、組織材料，擴充閱讀視野。	意涵，建立終身學習能力。 5-V-6在閱讀過程中認識多元價值、尊重多元文化，思考生活品質、人類發展及環境永續經營的意義與關係。
5-III-9因應不同目的，運用閱讀策略。		
5-III-10結合自己的特長和興趣，主動尋找閱讀材料。		
5-III-11大量閱讀多元文本，辨識中議題的訊息或觀點。		
5-III-12運用圖書館(室)、科技與網路，進行資料蒐集解讀與判斷、提升多元文本的閱讀和應用能力。		

註：摘錄自十二年國民基本教育課程綱要國民中小學暨普通型高級中等學校：語文領域—國語文，教育部，2018。

閱讀相關理論一致指出，隨著年齡增長，讀者透過閱讀學習新知，因而學習重點從字詞、句段，轉移至篇章(Chall, 1983, 1996)。然十二年國教課綱第三學習階段並未提及句意理解，第四學習階段卻出現相關描寫(5-IV-2)，顯示階段間的學習表現描寫出現斷層；若此為十二年國教課綱有意鋪排，則和閱讀能力發展背道而馳。第三學習階段和第五學習階段都有提到評述文本內容或建立論述體系(5-III-7和5-V-2)，然第四學習階段卻付之闕如。

十二年國教課綱意欲擘劃一個完整的12年的學習藍圖，從閱讀學習表現隱約可看出，自第三學習階段至第五學習階段的認知思考深度提升、學習廣度更多元，然國小至國中學習表現的文字描寫方式不一致，階段間的縱向銜接亦出現斷裂，缺乏整體性的結構。

二、本研究評量架構理論基礎及命題重點

本研究使用「國語文閱讀素養測驗」作為評量工具，為電腦化線上測驗。以謝佩蓉（2018）提出的評量架構作為命題架構，由研究團隊邀請現場教師共同研發試題。該架構考量十二年國教課綱國語文閱讀的學習內容與學習表現，以及第四學習階段學生適合閱讀的材料特性，以Kintsch（1988, 2019）建構統整模式（construction-integration model, C-I model）和OECD（2019）閱讀認知歷程為基礎。評量架構分為找出訊息（locate information）、理解（understand）及評價與省思（evaluate and reflect）三個向度，界定出16項評量指標，且每一項評量指標均分別對應國語文第二、第三、第四學習階段的學習表現（謝佩蓉、林明佳，2022），以之處理課綱學習表現縱向連貫性不足的問題。

對照分析謝佩蓉（2018）之評量架構和紐西蘭閱讀學習進展架構，發現兩者高度相似，雖然命名與結構有所不同，但兩者均為建構理論派典，強調閱讀核心能力均由字面理解至高層次省思。例如，紐西蘭LPFs中提到的「利用文本結構和特徵知識」乃利用文本特徵的知識進行閱讀與理解文本內容，這些知識包括辨識文本的結構、理解其使用的視覺元素（如標題與插圖）、標點符號、所使用的語言風格，以及語氣與語體，學生會逐漸熟練於辨識不同文本類型的目的與特徵（New Zealand Ministry of Education, 2025a）。這在謝佩蓉架構中，屬「R-RE-1分析寫作風格與寫作手法」。又如，紐西蘭LPFs中提到的「詞彙知識」係指讀者對文本中多數詞語意義的掌握，能夠辨認這些詞語的書寫形式，並理解其含義（New Zealand Ministry of Education, 2025a）。這在謝佩蓉架構中，屬「R-U-1在語境中解釋字詞意涵」。

評量架構為試題研發的依歸。找出訊息，意指從閱讀素材找到特定的訊息，或於網頁環境中搜尋並選擇相關資料（OECD, 2019）。本研究命題時，強調找出訊息的提問是為了引導更深的理解，而非文章中零碎的文字訊息。提問時，若題幹缺乏主要概念，學生難以理解題目究竟想問什麼，直到將選項逐一讀完（Haladyna & Rodriguez, 2013）。因而本研究強調找出訊息的提問須為文章中的關鍵訊息，且將解題概念寫在題幹而非選項，較貼近真實世界的提問方式，也就是符合素養導向試題的要素「在真實情境下所提出的真實問題」（任宗浩，2018）。

理解意謂從句子到段落，解析並整合閱讀素材，以瞭解其意涵、產生推論，辨識隱含其中的主要概念，為閱讀素材總結摘要或確立標題（OECD, 2019）。文本的理解是個複雜的歷程，涉及許多不同的元素，仰賴許多不同種類的訊息，產生複雜的心理表徵（Kintsch & Rawson, 2005）。

評價與省思乃超越閱讀素材的字面意涵和推論，針對素材內容或形式進行反思，評估訊息的品質與有效性（OECD, 2019）。學者指出，撰寫高層次思考試題有三項原則：（一）創造一個情境或允許受試者可使用素材；（二）使用新的情境避免受試者直接觸摸其記憶；（三）認知複雜性和試題難度不一定對等，記憶有難題，高層次思考也有易題（Brookhart, 2010）。

三、國際大型調查表現標準設定方法

內容標準（content standards）由一系列陳述所構成，說明特定的學習成果或學習目標；表現標準（performance standards）界定學習者在某項評量上需達到的具體表現水準（performance level），以判定其表現屬於何種等級（如達標、未達標等）（Cizek, 2012）。表現標準設定的方法各有所長，大多數研究採用「以內容為基礎」（content-based standard setting method）的方法，像是書籤標定法或Yes/No Angoff。學者指出，此類標準設定方法有兩項隱憂：（一）訂定標準的過程中，標準設定委員可能受限於課堂教學經驗或個人生活經驗，難以有足夠高度審視標準的適切性。（二）標準設定會議開始時，不會提供實證資料，致使標準設定委員進行判斷時，所得訊息不夠周延（Phillips, 2012）。

國際大型教育調查則採用量尺定錨法（Beaton & Allen, 1992）作為表現標準設定方法。除了國際數學與科學教育成就趨勢調查（Trend in International Mathematics and Science Study, TIMSS）、PISA使用之外，英國針對童年貧困所進行的跨國調查研究亦採用（James & Rossiter, 2018）。

（一）TIMSS採用的表現標準設定程序

為了讓資料使用者理解TIMSS量尺分數所代表的意義，TIMSS 1995和TIMSS 1999兩波調查採用百分位數切出四個切截點：90百分位數、75百分位數、50百分位數及25百分位數。到了2004年，切截點調整為「盡可能靠近既有百分位數」的四個量尺分數，形成等距切截點：625、550、475、400，稱作進階、高

級、中級及初級國際標竿，並界定每一階學生答對該階選擇題的機率為.60至.65（Gonzalez et al., 2004）；若為建構題，則界定每一水準學生答對該水準建構題的機率為.50。決定表現標準切截點之後，將所有的試題依照難度，由下而上、由易至難排列，再將學生能力由下而上、由低至高對應排列，形成試題圖（item map），再針對該水準試題，逐題描寫學生解答該題所需運用的特定內容知識與認知歷程，統合完整的陳述之後，由專家團隊建立表現水準描述共識。程序七步驟如下（Olsen & Nilsen, 2017）：

1. 專家撰寫評量架構，闡述所欲測量構念，包括量尺分數低至高學生的表現特徵的概略性假設。
2. 根據評量架構研發試題。
3. 命題者與專家依照評量架構分類試題，並描述解答每一試題所需的特定知識與技能。
4. 以現代測驗理論分析資料，獲得學生能力值與試題參數，繪製試題圖。
5. 依據實務和實證資料，決定切截點數及切截點分數。
6. 辨識切截點附近的試題作為表現水準描述的對象。
7. 撰寫表現水準描述，根據步驟3所描寫解答試題的共同性，以及步驟1評量架構所涵蓋構念的一般性描述。以TIMSS八年級數學的表現水準描述為例，描寫達到400、475、550、625分數的數學能力，並未描寫「未達400」學生的表現（von Davier et al., 2024）。

（二）PISA採用的表現標準設定程序

PISA界定學生答對和自己能力相當試題的機率（Response Probability, RP）為.62、並界定表現水準的區間寬度為.8 logits，來建立標準設定的量尺。閱讀表現分為八水準，命名為1c、1b、1a、2、3、4、5、6，最低水準的切截分數189.33，最高水準的切截分數698.32分。描述學生答對每一水準試題所需的技能和知識範圍，以及需要展現的能力，便完成PISA表現水準描述。和TIMSS描寫方式相似處在於，描寫達到1c、1b、1a、2、3、4、5、6的閱讀素養，未描寫「未達1c」的學生能力（OECD, 2020）。由於每個表現水準的區間寬度為.8 logits，假設某個表現水準下界（bottom of level）試題難度為0、表現水準上界（top of level）試題難度為.8。此外，由於界定學生答對機率為.62，即表示能力

位於表現水準下界的學生答對表現水準下界試題的機率為.62、能力位於表現水準上界的學生答對表現水準上界試題的機率為.62（OECD, 2017）。

綜整TIMSS與PISA的表現標準設定程序重點，作為本研究研發程序的核心概念：1. 基於實證數據決定切截點，採用量尺定錨法來設定標準。2. 繪製試題難度分布圖，確保表現水準描述能夠精確對應學生能力等級。3. 針對不同能力等級撰寫表現水準描述，展現閱讀素養的遞進關係。4. 僅描寫達到最低水準以上的能力特徵。

（三）撰寫表現水準描述的參考指引

評分指引（rubrics）是一種評量工具，內含評價學生作品的規準，以及描述作品在不同品質層次上的表現水準描述（Brookhart & Chen, 2015）。換句話說，評分指引能清楚呈現學生作業表現「達標」會是什麼樣貌、「待加強」會有哪些特徵，和標準設定撰寫表現水準描述的概念相同。本研究為盡可能撰寫各水準客觀、具體的描述，參採Stanley（2019）提出的撰寫評分指引注意事項：

1. 用詞具體、避免主觀

使用「相當」、「有點」、「好」和「完全」等詞語，將使讀者覺得很模糊。如果形容某人「有點聰明」可能產生兩種完全不同的解讀，可能意味著評論者對此人的智力評價不高，也可能表述此人被認為是聰明的。

2. 用詞過於具體

若標準描述為「學生在所有時間（即95%-100%）內清晰且明確地講話」，反而使得表現難以衡量，讀者如何評估何謂94%的時間？

3. 平衡描述的詳細度

描述愈簡單，使用起來就愈方便。描述的詳細程度既要提供足夠的說明使讀者關注關鍵能力，又不能讓描述過於繁瑣，讓讀者感到迷惑或混亂。

4. 提供具體的範例

提供具體的範例，使讀者易於理解和使用。例如，為「學生完全準備好」提供一個實例：「學生帶齊了所有材料，並且能夠順暢且清晰地轉換，使聽眾更好地理解主題。」，可以讓讀者明確地知道該關注什麼。

參、方法

本研究採用調查研究法蒐集標準設定所需實證資料，以縱貫性研究設計（longitudinal design）測量學生國語文閱讀學習表現，再以量尺定錨法設定表現標準，並產出表現水準描述。

一、標準設定程序

以量尺定錨法執行學生表現標準設定實際操作如下：

1. 召開說明會，招募命題團隊，說明評量目的與架構。為提升文本素材多樣性，也為了讓教學現場瞭解素養導向試題的樣貌，廣邀現場教師共同參與撰寫閱讀理解題組。

2. 命題與審題：依循文本的挑選或撰寫原則，選定並改寫文本、分析文本，再依據評量架構和命題原則撰寫試題初稿。將文本和題組初稿提交討論會，和外聘講座專家共同修審，使每篇文本均歷經二至三位講座調整文本及其題組。

3. 於命題與審題階段，透過討論會逐題確認所測量的構念及解題歷程。

4. 採試題反應理論估計試題難度和學生能力可能值。

5. 基於下列考量決定切截點及其分數（cut score）：

- （1）歐洲語言學習、教學、評量共同參考架構（Common European Framework of Reference for Languages: Learning, Teaching, Assessment, CEFR）為國際間採用的通用性語言能力標準，針對語言能力撰寫而非對應特定年級。CEFR初版將語言能力分為A1、A2、B1、B2、C1、C2（Council of Europe, 2001），A2能力標準為「日常對話」，而B1為「談論興趣、描述事件，並在工作與學校場合進行互動」，跳躍性的任務形成「不成比例的難度躍升」，不利學習（Alderson, 2007），故新版的CEFR將A2分為兩級，B1也分為兩級（Council of Europe, 2020），使各水準之間的難度落差較為相近，提供語言學習者更好的指引。為使各水準難度間距相似、發揮最佳的引導學習功能，本研究比照TIMSS和PISA採「等距」切截點。

(2) 本研究第1組追蹤樣本 (panel 1) 就讀七年級PV 1之「25百分位數、50百分位數、75百分位數及90百分位數」所對應的量尺分數分別為396、478、544、612，和TIMSS的切截點分數十分接近。

(3) 本研究界定第2組追蹤樣本 (panel 2) 閱讀素養表現平均數500、標準差100。樣本能力值呈常態分配的情況下，低於量尺分數400 (1個標準差) 人數比例約為16%，意即最低切截點分數定在400，則表現水準描述約可涵蓋84%學生表現。參照TIMSS和PISA的撰寫方式，表現水準描述乃將最低切截點視為基礎能力，故針對「最低切截點」以上的學生能力進行撰寫。

(4) 比照TIMSS採用400、475、550、625作為表現水準描述的切截點分數，並依照研究團隊共識，將四表現水準命名為M1、M2、M3及M4。

6. 將所有的試題依照量尺分數排列，由易至難，找出四個切截點分數附近的試題。

7. 於2020年4月至10月以及2021年12月召開表現標準討論會議，邀集相關領域專家學生共同完成表現水準描述。除由計畫主持人擔任評量領域專家外，出席的學者專長包括中文系教授二人、認知領域教授二人、閱讀理解試題專家以及本研究團隊成員，歷經20場表現標準討論會議凝聚共識，共譜表現水準描述。基於量尺定錨法乃針對試題，逐題描寫學生解答該題所需運用的特定內容知識與認知歷程再行綜整，和評分指引針對評量任務撰寫表現水準描述的歷程十分相近，故參酌Stevens與Levi (2013) 提出的建構評分指引四階段：反思 (reflecting)、列舉 (listing)、分組與標記 (grouping and labeling) 以及應用 (application)，本研究執行過程如下：

(1) 釐清評量目的：針對切截點附近的試題，思考當初設計這些題目想要評量的目標為何？

(2) 列出具體細節：題目希望學生能展現的語文能力表現有哪些？由研究團隊逐題寫出答題的認知歷程（要如何得出這個答案）與必備的關鍵能力或先備知識。

(3) 建構上位能力：將相似的能力表現歸類在一起，統整撰寫為較通用的句子。這個過程需要來回反覆的調整、確認句子能適切的涵蓋、闡述試題所反映的能力。

(4) 建立水準描述：將前三個階段完成的內容，填入表現水準描述表格之中，並注意表現水準M1、M2、M3及M4能力的縱向連貫性。

二、研究對象

研究對象母群體為107和108兩個學年度的七年級學生。

(一) 母群體

第一組母群體為九年一貫課綱最後一屆的學生，乃107學年度臺灣地區七年級學生。根據教育部統計處(2020)資料，排除特殊教育學校(包含戲曲學校和未其他未使用教育部課綱教學的學校)，學校總計949校，學生總計201,188人。第二組母群體為十二年國教課綱第一屆的學生，乃108學年度臺灣地區七年級學生，學校總計946校，學生總計197,858人。

(二) 樣本

採二階段分層叢集抽樣設計(two-stage stratified cluster sample design)，第一階段抽取學校，第二階段針對抽中的學校進行校內班級隨機抽樣，所抽中的班級，全班學生均為樣本。

第一階段以「偏遠地區」為顯性分層變項，再以「校層級平均分數」和「都市化等級」為隱性分層變項，於各層內依照第一組母群體各校七年級人數排序，再採「依比例的機率」(Probabilities Proportional to Size, PPS)抽樣，抽出256校。第二階段進行班級抽樣，於抽中的256校的七年級抽出269班，全班均為樣本，總計7,008人為本研究第一組追蹤樣本(panel 1)。選擇第二組追蹤樣本(panel 2)時，第一階段校層級的抽樣，優先詢問panel 1學校是否願意再參加新一屆的調查，若願意，則直接成為panel 2學校，若不願意，則以該分層中的替代學校(replacement school)作為樣本學校，並在其中抽出一班，以全班學生為樣本。歷經此歷程，panel 1學校同意成為panel 2學校共102校，不同意為132校。各分層詳細的樣本分布，請詳見吳正新(2022)。

於調查第一年，所有研究樣本隨機分派作答「國語文閱讀」、「英語文」、「數學」、「社會」及「自然科學」五科中的兩科，之後第二年、第三年的調查均作答第一年被分派的相同科目。由於總樣本因轉出、轉入略有變動，panel 1中，本研究對象「國語文閱讀」於七年級分派到2,803人，到考數2,709人，到考

率為96.65%；於八年級分派2,807人，到考數2,711人，到考率為96.58%。panel 2於七年級分派2,565人，到考數2,511人，到考率為97.89%；於八年級分派2,780人，到考數2,440人，到考率為87.77%。

三、研究工具

良好的評量工具為標準設定有效性的重要前提。為確保評量工具具備足夠的信、效度，除採用具備信、效度證據的評量架構，且設定嚴謹的研發歷程，文體類型、文本議題及題本組成均為研發過程審慎評估的項目。

（一）評量工具效度證據

1. 文體類型分布

PISA 2018以閱讀素養為主要評量標的，規劃七種文體類型，包括描述（description）、記敘（narration）、說明（exposition）、議論（argumentation）、說明書（instruction）、互動式（interaction）、事務（transaction）（OECD, 2019）。電腦施測題數共244題，文體類型題數及其比例分別為：描述30題（12%）、記敘45題（18%）、說明72題（30%），議論39題（16%）、說明書12題（5%）、事務25題（10%）以及混合型21題（9%）（OECD, 2020）。美國教育進展評量（National Assessment of Educational Progress, NAEP）僅將文體分為兩類，規劃八年級記敘文（literary text）和說明文（informational text），比例分別為45%和55%，2024年實際施測的比例則為48%和52%（U.S. Department of Education, 2025）。十二年國教課綱強調五種文體並重，然而，真實世界中的文本甚難明確區分（OECD, 2020），故本研究僅將其區分為「記敘文與抒情文」和「論說文」（說明文和議論文）兩大類，參採其他大型評量的設計，規劃「記敘文與抒情文」33%、「論說文」67%，四波調查的文體類型分布大致均符合擬訂目標（詳如表2）。

2. 文本對應議題

為了理解文本內容，讀者必須將從文本中獲取的訊息與自身已有的知識相整合，以在心智中建構對文本意義的理解模型（Dijk & Kintsch, 1983）。若讀者對特定主題的知識較豐富，他們在閱讀該主題的文本時，其理解能力會顯著優於背景知識較弱的讀者，這一現象對高能力讀者與低能力讀者皆適用（Smith et al.,

表2
評量工具文體類型分布百分比

	Panel 1		Panel 2	
	2019 wave 1	2020 wave 2	2020 wave 1	2021 wave 2
	Grade 7	Grade 8	Grade 7	Grade 8
	<i>n</i> = 24	<i>n</i> = 24	<i>n</i> = 24	<i>n</i> = 24
記敘文與抒情文	29.17	25.00	29.17	29.17
論說文	70.83	66.67	70.83	66.67
混合	0	8.33	0	4.17

2021)。為降低背景知識對於母群體閱讀能力估計的影響，本研究所研發的文本涵蓋15項議題（教育部，2020），包括人權教育、環境教育、科技教育、能源教育、家庭教育、品德教育、生命教育、法治教育、資訊教育、安全教育、防災教育、生涯規劃教育、多元文化教育、國際教育、閱讀素養教育。鑑於本研究評量構念為學生的閱讀素養，每一篇都和閱讀素養有關，因而對應十二年國教課綱議題時，以其他18項議題為優先歸屬類別，而文學性作品通常難以對應其他議題而歸類至閱讀素養議題。

3. 題本組成

限於疲勞與測驗時間等因素，一位學生不可能做完當年所有的試題，故將數個試題（item）或題組（testlet）組成區塊（block），再將數個區塊透過部分平衡不完全區塊（Partially Balanced Incomplete Block, PBIB）組成測驗題本（booklet），以平衡試題的位置效應（positional effects），避免影響受試者的能力估計結果（Chen & Jen, 2012）。

panel 1就讀七年級時210題，就讀八年級時則為238題，兩波調查的共同題為14題組、138題，占七年級施測題數60.53%、八年級施測題數57.98%。panel 2就讀七年級時210題，就讀八年級時則為234題，兩波調查的共同題為13題組、130題，占七年級施測題數61.90%、八年級施測題數55.56%。七年級與八年級作答的相同試題，作為垂直等化校準量尺（Kolen & Brennan, 2004）。

（二）評量工具信度證據

自1910年以來，信度係數即為測量工具品質的重要指標。大型調查設計關注的是母群體潛在特質的估計，以期望後驗法（Expected A Posteriori, EAP）估計母群體能力可能值，因而採EAP reliability表徵測驗工具的誤差（Adams, 2005）。本研究panel 1於調查七年級時，測驗工具EAP reliability為.81，八年級則為.82。panel 2於調查七年級時，測驗工具EAP reliability為.75，八年級則為.81。信度證據良好。

四、資料分析方法

試題難度估計採試題反應理論（Item Response Theory, IRT）之部分給分模式（Partial Credit Model, PCM）（Masters, 1982），共同估計的資料包括panel 1就讀七年級和八年級，以及panel 2就讀七年級和八年級，四波次作答國語文閱讀的作答反應。為能進行不同調查母群及不同年級之量尺等化，主要界定如下（陳冠銘，2022）：

1. 前提：假設同一試題出現在同屆調查中的不同題本或是不同屆調查，試題難度均相同。
2. 答對試題的機率：試題難度參數採試題通過率.62之閾值。
3. 試題難度和學生能力值位於相同量尺：估計試題難度採logit量尺，之後利用線性轉換將同一學科的試題難度以及後續估計的學生素養表現量尺分數，轉換為「panel 2就讀七年級的量尺分數平均數為500、標準差為100」，亦即定錨於十二年國教課綱第一屆的七年級。

肆、結果與討論

經由嚴謹的抽樣、組卷、派題、施測、估計、量尺化程序，據之進行表現標準設定。

一、試題難度分布與認知複雜度

圖1縱軸為量尺分數，學生能力分布繪製於圖1左側，試題難度分布和切點繪製於右側。由圖1可知，本研究工具試題難度分布甚廣，亦即所能測得的學生能力值甚廣，低、中、高能力學生均有相對應試題，可作為未來向下對接國小量尺、向上對接高中量尺的準備。表現水準M1、M2、M3及M4切截點上、下具有足量的試題，可提升切截點附近學生能力的估計值精確度，表現水準描述的撰寫素材豐富。

整體而言，三個向度的試題（綠色為找出訊息，藍色為理解，紅色為評價與省思）由圖的左下分布至右上，隨著認知複雜性增加，整體試題難度亦增加。本研究正式調查題數共344題，表現水準M4以上的試題占107題（31.10%），M3水準試題占76題（22.09%），M2水準試題占74題（21.51%），M1水準試題占44題（12.79%），未達M1水準試題占43題（12.50%）。

進一步分析各向度題數占該表現水準的百分比，理解於M1至M4所占比例為47.37%至59.09%（詳見表3）。參照NAEP閱讀評量架構分為找訊息與再憶（locate/recall）、統整與詮釋（integrate/interpret）以及批判與評鑑（critique/evaluate）三個向度，和本研究所採用的架構相似。NAEP規劃八年級試題在三個向度的比例為20%、50%、30%，2024年實際施測的比例則為17%、62%、21%（U.S. Department of Education, 2025），本研究「理解」於各表現水準所占比例和NAEP「統整與詮釋」所占比例十分相近。

找出訊息於M1至M4所占比例由M1之31.82%遞減至M4之14.02%；另一方面，評價與省思於M1至M4所占比例由M1之9.09%遞增至M4之30.84%，除呈現良好的建構效度證據，亦彰顯高層次思考的試題對於學生相當具有挑戰（詳見表3）。

認知複雜性和試題難度不一定對等（Brookhart, 2010），實證研究結果指出，認知複雜度和試題難度為中度相關（Schneider et al., 2013）。以高層次思考技能並不需要「完全掌握低層次思考技能」作為前提，所有層次的思考在思維過程中均相互關聯且非線性（Shalihin, 2024）。由圖1可知，找出訊息、理解及評價與省思各自的試題難度，自「未達M1」至「M4」均有分布，顯示試題難度的

圖1
試題難度分布

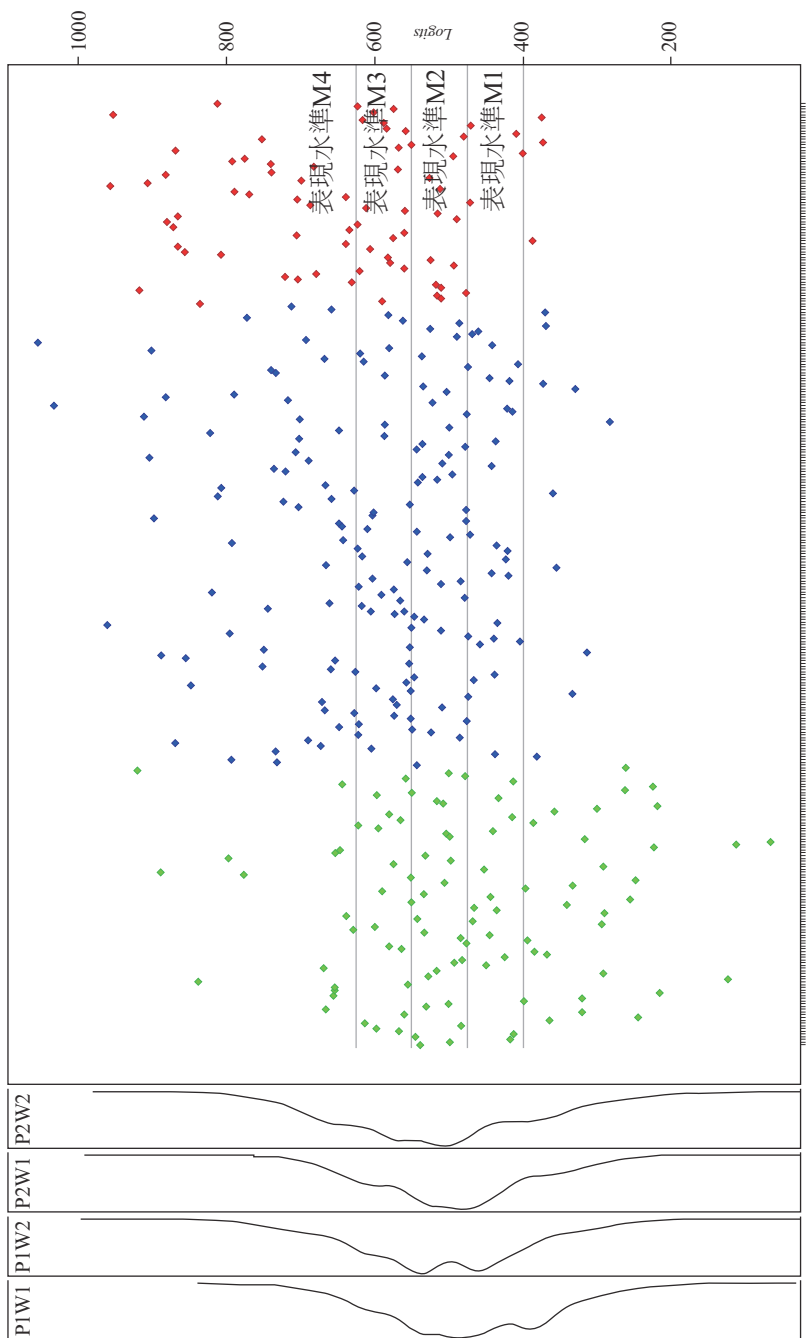


表3

各表現水準題數及向度題數比例分布

表現水準	不分向度 <i>n</i> = 344	區分向度					
		找出訊息 <i>n</i> = 102		理解 <i>n</i> = 168		評價與省思 <i>n</i> = 74	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
M4	107	15	14.02	59	55.14	33	30.84
M3	76	19	25.00	36	47.37	21	27.63
M2	74	24	32.43	37	50.00	13	17.57
M1	44	14	31.82	26	59.09	4	9.09
未達M1	43	30	69.77	10	23.26	3	6.98

註：表格中的百分比為各向度題數占該表現水準的百分比。

影響因素不只認知複雜性。以本研究範例題〈微物跡證〉題組問題1「哪兩個人爭奪羊皮？」為例，只要找出第二段第一句「人有負鹽負薪者」就是答案，然該段為學生較不熟悉文言文，即便解題的認知歷程為「找出訊息」，難度仍達到M4；而〈妮妮與橘橘的對話〉題組問題6，請學生使用關鍵字查資料，難度則只有M2。就省思與評價而言，以〈微物跡證〉題組問題8為例，請學生評論「以現今科學辦案的角度，李惠辦案過程，證據充足嗎？」為難度M4的高難度試題，而問題6「『鹽屑』在這則故事裡扮演的作用，相當於現今科學辦案中的哪一種證據？」，乃根據文本情境對應生活中的熟悉例子，難度則僅為M2（請參見附錄）。由此可知，「低認知複雜度但難度高」或「高認知複雜度但難度低」均能透過本研究的實證資料得到印證。

二、各水準的表現水準描述

依閱讀題組的閱讀特性，可概分為評量學生的一般閱讀和數位閱讀兩種類型素養。若閱讀素材形式為虛擬網站，或題目要求學生自行至某一特定網站查詢資料，學生的閱讀歷程為「非線性」，則歸類為數位閱讀素養；若閱讀任務為靜態頁面，即使形式為部落格、粉絲專頁等，其閱讀歷程仍為「線性」，則歸類為一般閱讀素養（本研究試題均為電腦化數位試題，不宜稱為紙本閱讀）。

閱讀是建構性的歷程，有助於學生獲取新知並發展重要能力。它不僅僅是語言活動，更是認知歷程，包含低層次與高層次能力的多面向技能（DJEFAFLIA, 2018）。本研究採用的評量架構乃區分為找出訊息、理解及評價與省思三個向度的閱讀認知歷程（謝佩蓉，2018），與國語文領綱第四學習階段的閱讀學習表現在邏輯與內涵上高度一致，可從國語文課程綱要研修紀錄得到印證（國家教育研究院，2015）。「找出訊息」的核心是辨識文本中的關鍵資訊，並在多元文本或數位環境中篩選並選擇相關資料，與第四學習階段學習表現所述的「理解與解釋各類文本的句子、段落的主要概念」以及「應用閱讀策略，整合跨領域的知識，增進學習的效能」相呼應，著重學生應具備準確擷取與統整資訊的能力。其次，「理解」涵蓋從句子到段落的意義建構，包括訊息整合、詮釋及推論，與「指出文本寫作的目的與觀點」、「比較不同的標點符號在文本中所達到的效果」學習表現相符，皆強調學生應能超越個別句子的表層意義，整合資訊並建構意義。「評價與省思」強調超越文本字面意涵，針對內容或形式進行批判性反思與評估，與「理解各類文本的寫作特徵」、「評鑑文本內容和形式的適切性」、「閱讀多元文本，分析文本中的重大議題內涵，探索其與個人生活、社會結構的關聯」及「統整閱讀的內容，並轉化為日常生活解決問題的能力」學習表現密切相關，期望培養學生跨領域整合、批判思維及多角度分析能力。可見領綱研修時，確實採用認知歷程的派典撰寫學習表現，最終精煉、定稿為第四學習階段閱讀學習表現（如表1）。

為能更細緻闡述各表現水準，以評量架構的閱讀認知歷程描述達到每一水準學生所展現的兩類型閱讀素養，據以強調學生應具備的訊息辨識、理解到批判省思的能力，並透過多元文本的閱讀與分析，發展系統性思考與素養導向的閱讀能力。在表4中，能力位於M1學生所面對的閱讀任務最簡單，接續為表現水準M2、表現水準M3，難度遞增，而表現水準M4為難度最高的閱讀任務。

以範例題流水號[7]為例，該題位於評價與省思向度M3水準，為〈微物跡證〉的問題7。〈微物跡證〉為一篇說明文，旨在說明微物跡證於案件鑑定中的關鍵作用，並輔以文言文故事「李惠審判羊皮」作為具體例證。鑑於「理解各類文本形式和寫作特色」為國語文課綱學習表現之一（如表1），本研究採用的評量架構中，以「R-RE-1分析寫作風格與寫作手法」評量指標呼應課綱這項學習

表4

第四學習階段國語文閱讀素養表現水準描述

表現 水準	分數	表現描述		
		找出訊息	理解	評價與省思
M4	625 以上	- 一般閱讀素養：能在較不熟悉的文本主題或寫作文體[1]、形式中，找出與問題設定條件相符的訊息[2]，或需跨段落才能找到訊息。 - 數位閱讀素養：能在網站情境中，從不同段落找到與問題設定條件相符的訊息，或從網路搜尋到的複雜資訊中，找到所需的訊息。	- 一般閱讀素養：能在較不熟悉的文本主題[3]或寫作文體、形式中，掌握具有多元觀點的文本主旨，運用多個段落的訊息判斷立場、因果關係、事件順序等，或推論文句深層涵義、故事角色情緒、作者意圖等。能根據對文言文語句的理解，判斷關鍵字詞的意義，且該關鍵字詞具有高度複雜的多義性。能以精簡且連貫性佳的文字，完整寫出單一段落或全文的大意。 - 數位閱讀素養：能整合文本和外部網頁的訊息[4]，完成任務。	- 一般閱讀素養：能精確辨識多樣性的寫作手法，運用文本以外個人經驗和知識，評估較不熟悉的圖、文訊息在文本中所產生的閱讀效果，評論文章內容[5]及相關訊息的可信度。能根據文本啟發，結合生活經驗，提出與文本議題相似的例子，並解釋該例子的適切性。能根據文本主旨與問題設定條件，選擇適切的標題，並說明其他標題不適切的理由。 - 數位閱讀素養：能在網站情境中，根據從文本得到的啟發，結合生活經驗，提出與文本議題相似的例子，並解釋該例子的適切性。
M3	624 至 550	- 一般閱讀素養：能從不易觸及的解題線索中找到訊息，或從多個可能的訊息中找到恰當的訊息。 - 數位閱讀素養：能在問題設定條件中爬梳正確的關鍵詞，或在	一般閱讀素養：能掌握主題明確或單一觀點的文本主旨，運用鄰近段落間的訊息判斷關係、立場、事件順序等。能根據前後文判斷文言文的詞義[6]。	一般閱讀素養：能辨析圖文訊息、文本結構及寫作手法呈現的閱讀效果[7]。當閱讀不熟悉主題的文本，能結合文本以外的個人經驗及知識，合理回應真實情境並提出

(續下頁)

表現 水準	分數	表現描述		
		找出訊息	理解	評價與省思
		網站情境中，利用網站地圖的知識，找到與問題設定條件相同的訊息。	• 數位閱讀素養：能將網站情境中的文字訊息，轉化成具象的圖片。	看法。 • 數位閱讀素養：NA
M2	549 至 475	- 一般閱讀素養：能從較明顯的解題線索中找到訊息，或從文本中找出意義相近的訊息。 - 數位閱讀素養：能在網站情境中，從單一分頁找到正確的訊息，或從問題設定條件中選擇關鍵詞，並運用該關鍵詞在搜尋引擎中，查詢所需的資訊[8]。	- 一般閱讀素養：能整合段落內的訊息形成主要概念，針對訊息進行分析、比較、類推[9][10]。能根據對文言文語句的理解，判斷關鍵字詞的意義，且該關鍵字詞的意義仍為現今所慣用[11]。 - 數位閱讀素養：能在網站情境中，整合單一分頁中的訊息，形成主要概念。	- 一般閱讀素養：能辨析常見的文本結構及熟悉的寫作手法所呈現的閱讀效果[12]。當文本屬於學生不熟悉的主題，能運用文本中所獲得的啟發，辨別生活中相似的例子[13]。 - 數位閱讀素養：NA
M1	474 至 400	- 一般閱讀素養：能從文本開端找到訊息，或從文本段落內找到單一明確的訊息。 - 數位閱讀素養：能利用問題設定中給定的訊息作為關鍵詞，並運用該關鍵詞在搜尋引擎中，查詢所需的資訊。	- 一般閱讀素養：能根據段落內容辨認出段落的主要概念[14]，或形成簡單的推論，或判斷較明確的指稱詞指涉對象。 - 數位閱讀素養：能在網站情境中，整合單一分頁中的訊息。	- 一般閱讀素養：能根據文本情境，從生活中舉一個熟悉的例子。 - 數位閱讀素養：NA

註：

1. 數位閱讀素養的評價與省思歷程，主要以表現水準4為試題研發目標，現階段無法提供表現水準1、2、3的描述。
2. [1]至[14]為表現水準描述所對應的範例題流水號。

表現，故設計此題評量學生於「一般閱讀素養」中「評價與省思」向度的表現，聚焦於其分析寫作手法的能力，辨識作者所運用的寫作技巧及其對閱讀效果的影響。

[7]問題7.「李惠審判羊皮」的故事，在本文的功用是什麼？

- ①指出問題來源
- ②提供例證資料
- ③製造反諷效果
- ④比較古今異同

本研究完成的第四學習階段國語文閱讀素養表現水準描述具有下列四項特色：

（一）M1到M4的能力遞進關係以精緻化課綱學習表現

本研究的表現水準描述強調M1至M4能力的遞進，使閱讀素養的學習發展脈絡更加清晰，幫助教師與學生明確掌握不同能力層級的表現，並設定適切的學習目標。以第四學習階段國語文課綱學習表現「5-IV-2理解各類文本的句子、段落與主要概念」為例，該學習表現描述的是「理解」認知歷程，本研究將之細緻化，描寫學生在不同表現水準（M1至M4）展現的能力：

1. 表現水準M4的學生能「掌握具有多元觀點的文本主旨」，亦即能夠理解並統整來自不同立場或角度的論述，綜合文本的主要論點與結論，展現高度的推理與整合能力。

2. 表現水準M3的學生能「掌握主題明確或單一觀點的文本主旨」，表示能夠準確辨識文章的主要論點，但尚未發展出比較、對照與多視角整合的能力。

3. 表現水準M2的學生能「整合段落內的訊息形成主要概念」，即能辨識並組織文本中具關聯性的訊息，初步建立對文本內容的理解，但仍侷限於單一段落內的訊息統整。

4. 表現水準M1的學生能「辨認出段落的主要概念」，即能從段落內容中找到關鍵訊息，但尚未能進一步整合或推論更深層的意涵。

能力遞進架構使教師與學生能夠根據學習者目前的能力水準，逐步設定學習目標，以高一個水準的表現作為努力方向。例如，當學生已能夠在M2層級掌握

段落內的主要概念時，教師可引導學生嘗試提升至M3層級，進一步掌握完整文本的主旨；若學生已達到M3層級，則可培養其批判思考能力，使其能整合多元觀點，朝向M4層級發展。

（二）閱讀歷程為軸線使表現水準描述更具系統性

如同紐西蘭以七項核心能力描述不同年級的閱讀學習進展（New Zealand Ministry of Education, 2025b），本研究以「找出訊息」、「理解」、「評價與省思」三項閱讀歷程分別闡述第四學習階段不同表現水準的學生能力。教師進行閱讀教學時，可針對不同認知歷程設計不同的提問，以此協助教師規劃完整且均衡的閱讀教學，避免只聚焦於單一能力，確保學生在不同閱讀歷程中均獲得適當的訓練。

國中教育會考國文科評量九年級學生的紙本閱讀和寫作學力表現，表現標準分為精熟、基礎及待加強三個等級（國立臺灣師範大學心理與教育測驗研究發展中心，無日期）。由表5可知，國中教育會考以概念性語句描寫學生的國文能力，採用的規準包括語文知識、文本內容與形式及處理問題，以「副詞」區分三個等級的表現。本研究所研發國語文閱讀素養表現水準描述（如表4）提出各水準於各認知歷程的能力描述，精緻化第四學習階段的閱讀能力表現。

表5
國中教育會考國文等級描述

等級	等級描述
精熟	能具備與教材相關的語文知識，能深入地理解、評鑑各類文本的內容與形式，並能適切統整應用以處理問題。
基礎	大致能具備與教材相關的語文知識，大致能理解、評鑑各類文本的內容與形式，並大致能統整應用以處理問題。
待加強	僅能具備部分與教材相關的語文知識，僅能有限地理解、評鑑各類文本的內容與形式，並有限地統整應用以處理問題。

註：引自「各科等級描述」，國立臺灣師範大學心理與教育測驗研究發展中心，無日期，2025年5月4日，取自<https://cap.rcpet.edu.tw/score1.html>。

PISA的閱讀表現水準描述，自水準1c至水準6，位於每一個表現水準的讀者面對各種不同的任務特性（如表6）：解題線索由明確而漸隱晦、理解整合的

範圍由相鄰而遠距、詮釋由字面至深層、評價與省思由比較對照進階到結合多項外在規準、主題由熟悉至不熟悉，以及文本由短而長、由單至多、相互干擾的訊息由少而多。然而，PISA對於15歲的閱讀表現水準描述敘寫方式為整體性（holistic），若作為課綱學習表現參考指引仍不夠明確。本研究所研發之國語文閱讀素養表現水準描述（如表4）細緻分析並呈現第四學習階段學生能達成的閱讀能力，可作為未來國語文領綱一至十二年級縱貫銜接、調整的關鍵基礎。

表6

PISA閱讀素養表現水準描述

水準	切截分數	任務特性
6	698.32	表現水準6的讀者能理解長而抽象的文本，解題的線索隱晦不明，和提問之間的連結亦不明確。讀者能比較、對照及整合多元且潛在衝突的觀點，運用多項規準並整合遠距離段落之間的訊息，產生推論，以決定該如何應用資訊。 表現水準6的讀者能使用文本未述及的規準，深度省思文本出處和內容之間的關係。藉由推論訊息的來源（不論它是明確的或潛藏的），以及其他可供判斷訊息有效性的線索，作為比較、對照跨文本間的訊息，辨識和解決文本間的差異與衝突。 表現水準6的任務主要是要求讀者根據任務要求和文本內容制定縝密的計畫，該計畫需要結合多項規準並產生推論。這個層級的閱讀材料包括單一或數個複雜且抽象的文本，涉及多個且可能不一致的觀點。解決任務所需的關鍵訊息可能是嵌入文本內或跨文本的細節，可能被其他相互干擾的訊息所掩蓋。
（水準1b至5的任務特性，略）		
1c	189.33	表現水準1c的讀者能理解並確認簡短且文法簡單句子的字面意涵，在一小段有限的時間內，為明確且簡單的目的而閱讀。 這個層級的任務包括簡單的詞彙和文法結構。

註：譯自PISA 2018 Technical Report, by Organisation for Economic Co-operation and Development, 2020。

（三）納入數位閱讀素養反映當代閱讀環境的轉變

數位科技與人工智慧的快速發展，改變了人們在家庭、學校與職場中的閱讀與資訊交換方式。面對大量且多樣化的數位內容，讀者需具備搜尋、篩選、比較、評估與統整資訊的能力，才能有效應對數位時代的閱讀挑戰。

OECD（2021）指出，閱讀不再只是訊息的提取與理解，而是建構與驗證知識的能力。數位閱讀素養較高的國家，其學生在面對複雜的數位文本時，能夠更準確地區分事實與觀點，並在處理多文本資訊時展現更高層次的理解與評價能力。例如，能夠辨識資訊的來源是否可信、判斷文章是否具有偏見，並整合來自不同來源的資訊，以形成完整的知識框架。這顯示數位閱讀能力已成為學習與生活不可或缺的核心素養。

然而，當前第四學習階段國語文課綱對於數位閱讀的描述相對有限，僅在「5-IV-6能運用圖書館（室）、科技工具，蒐集資訊、組織材料，擴充閱讀視野」中提及數位工具的使用，未明確闡述學生在數位環境中應具備的閱讀理解與資訊評估能力。為了彌補這一不足，本研究將數位閱讀素養納入表現水準描述，使其涵蓋當代閱讀環境的關鍵能力，引導學生在科技社會中成為具備高層次閱讀理解與批判性思考能力的智慧型讀者。

（四）信實數據作為表現標準及試題範例的支持證據

本研究採量尺定錨法，依據學生實際作答結果，分析學生在不同難度試題上的表現特徵，歸納「找出訊息」、「理解」、「評價與省思」三項閱讀歷程之M1至M4水準的表現水準描述，使各表現水準具備實證數據支持，而非憑專家主觀判斷。表現水準描述不僅提出概括性能力說明，更搭配具代表性的試題範例，以具體呈現學生在該水準可能展現的閱讀行為。

本研究完成的表現描述（如表4）提供更細膩的能力描述，界定每一水準能力的情境脈絡，再撰寫該脈絡下，學生於特定認知歷程所能達成的任務。除了適合作為課綱學習表現的參考指引，亦回應全球教育革新趨勢——評量即學習（assessment as learning）的理念。評量即學習強調學生扮演主動性角色，「評量」產出學習機會，使學生主動投入於尋找、關聯和使用證據。不只是回憶和使用先備知識，而是促進後設認知和自我調整學習，使學生自我監控表現，顧及自我學習需求，使得評量一方面促進學習，另方面判斷學生的表現（Yan & Yang, 2022）。也就是說，一份好的評量標準可以引導學習發生。本研究所建構的表現水準描述，使測驗分數不只是分數，而能讓教與學雙方理解，得到這個分數的學生能解答的學習任務特性。藉由描寫不同表現水準學生於不同向度的表現，使教與學雙方均能得知目前的學習狀態，以高一階學習表現為目標，充分作為引導學

習發生的地圖。

三、表現水準描述範例題

為具體呈現表現水準描述的意涵，本研究提出〈妮妮與橘橘的對話〉和〈微物跡證〉兩組範例試題。前者融入環境教育與閱讀素養教育兩項議題，為對話形式的說明文本，後者融入法治教育與閱讀素養教育兩項議題，形式上為包含文言文故事的說明文本，據以呈現閱讀素材的多元樣貌。

茲將兩組試題依照表現水準及其難度整理於表7，題目由上而下的排序乃依照試題難度編排，排在上方的難度較高，排在下方的難度較低。為擷節版面，文本請詳見表7列出的連結，範例試題則詳述於附錄。

表7

第四學習階段國語文閱讀素養各表現水準範例題一覽

表現水準	分數	題組名稱與題號
M4	625以上	〈妮妮與橘橘的對話〉：問題5[4] 〈微物跡證〉微物跡證：問題8[5] 〈妮妮與橘橘的對話〉：問題3[3] 〈微物跡證〉：問題5[2] 〈微物跡證〉：問題1[1]
M3	624至550	〈微物跡證〉：問題7[7] 〈微物跡證〉：問題3[6]
M2	549至475	〈妮妮與橘橘的對話〉：問題1[9] 〈妮妮與橘橘的對話〉：問題4[12] 〈微物跡證〉：問題2[11] 〈微物跡證〉：問題6[13] 〈妮妮與橘橘的對話〉：問題6[8] 〈妮妮與橘橘的對話〉：問題2[10]
M1	474至400	〈微物跡證〉：問題4[14]

註：

1. 〈妮妮與橘橘的對話〉文本詳見https://tasal2020.naer.edu.tw/home/simulation/sim_itemset.php?encitemsetid=wTNFAV93vxl9qykX3YFt97_XL_vgBluxFO6gvPI6UTU=&ver=NE。
2. 〈微物跡證〉文本詳見https://tasal2020.naer.edu.tw/home/simulation/sim_itemset.php?encitemsetid=-ZrQdU9p6R59JPeT6vgXZfVRXahNcutvKfnHaFsPVDQ=&ver=NE。

四、研究限制

本研究比照TIMSS和PISA國際大型教育調查所採用的量尺定錨法設定標準，界定之最低切截點量尺分數為400，未達最低標準的人數比例約為16%，意即表現水準描述約可涵蓋84%學生表現。對照2024年國中教育會考國文科「待加強」比例為13.02%、2025年比例為12.18%（國立臺灣師範大學心理與教育測驗研究發展中心，2024，2025），本研究未達最低切截點的比例略高。

本研究係以第四學習階段學生為對象進行標準設定，未來若能向下延伸至小學階段，則原屬「未達M1」的16%學生群體有機會依其能力表現進一步細分，發展為連結國小至國中的表現等級。透過縱貫性建構表現標準，設計信、效度良好的試題，更細緻描繪學生能力表現，提升教學與學習的回饋。因此，本文所呈現的表現標準與學生分布結果，可視為階段性成果，仍有進一步精緻化與延伸應用的空間。

伍、結論與建議

茲將本研究之結論與建議分述如下。

一、結論

本研究針對十二年國教課綱第四學習階段國語文閱讀學習表現，建立表現水準描述，並提出具體範例試題，精緻化現行課綱學習表現。綜整研究成果，歸納以下五點結論：

（一）以大型調查實證數據支持第四學習階段閱讀學習表現水準描述

本研究採用國際大型調查使用的量尺定錨法作為閱讀表現標準設定方法，透過四波次的隨機調查樣本10,955人次和344道試題，為標準設定提供堅實的數據支持，確保研究結果的可靠性。參酌TIMSS、PISA及CEFR能力進階概念，設定400、475、550、625四個量尺切截點，比照國際以核心能力撰寫學習進展的方式，以「找出訊息」、「理解」、「評價與省思」三項閱讀認知歷程為軸線，分

別撰寫M1至M4四個水準之一般閱讀和數位閱讀的能力描述，並根據實證數據選取合適的範例題，彰顯十二年國教國語文領域課綱閱讀學習表現的豐富內涵。

（二）表現水準描述精緻化課綱學習表現並引導能力遞進

本研究完成的表現水準描述，揭示第四學習階段國語文閱讀能力由M1至M4的遞進層次，展現十二年國教課綱培養學生深入理解文本、綜合統整訊息以及批判反思能力的學習期待。第四學習階段涵蓋七、八、九三個年級，學生的認知發展與閱讀能力均發生變化，無論是文本理解的深度，或是對於不同文體的處理能力。藉由表現水準描述具體呈現學生閱讀素養隨能力提升而展現的層次變化，有助於教師與學生明確掌握不同能力水準的表現，據以調整教學內容與策略，並設定適切的學習目標。

（三）表現水準描述展現數位閱讀素養的能力特徵，契合當代閱讀環境發展趨勢

學生在數位環境下的閱讀理解歷程，需運用搜尋、篩選與整合跨頁面或跨文本訊息的能力。表現水準描述將數位閱讀素養的能力特徵具體呈現於各能力水準之中，展現學生在數位情境中辨識訊息、理解文本與評價內容的能力歷程，與當代閱讀環境發展趨勢相呼應。此描述可作為教師設計教學活動及評量時的參考指引，促使學生在數位資訊環境中，培養精準擷取訊息、批判省思與綜合運用的閱讀素養，進一步回應課綱所強調的多元文本閱讀與審慎使用各類資訊的能力。

（四）認知歷程與試題難度並非線性關係，表現水準描述揭示找出訊息亦可能為高難度任務

實證數據分析結果顯示，「找出訊息」認知歷程並非必然對應低難度試題，自「未達M1水準」至「M4水準」皆有試題，部分涉及陌生文本或跨段落定位的任務，對學生而言具高度挑戰性。「評價與省思」認知歷程亦有相似的狀況，評論文章內容對學生而言有難度，根據文本情境對應生活中的熟悉例子則容易得多。本研究透過表現水準描述，揭示各閱讀歷程於不同水準可能涉及的能力差異，有助於教師精準掌握學生閱讀理解歷程中的困難點。

（五）表現水準描述搭配範例試題為教師挹注素養導向評量後盾

表現水準描述不僅界定不同能力水準學生的閱讀表現，更搭配實證試題範例，使抽象的能力敘述具象化，方便教師參照應用。教師得以據此設計素養導向

課室提問或段考試題，評量學生閱讀理解能力，引導學生朝向更高水準的表現邁進。表現水準描述與範例試題相輔相成，能成為連結課堂教學與學生學習表現之橋梁，使教學、學習與評量三者相互扣合，共同促進學生閱讀素養的提升。

二、建議

（一）建構國語文閱讀素養縱向銜接的表現水準描述

本研究聚焦於第四學習階段國語文閱讀素養表現水準描述之建置，惟十二年國教課綱各學習階段之學習歷程應具縱向連貫性，國語文閱讀素養學習表現亦應展現能力漸進發展的軌跡。然而，現行第三、四、五學習階段的閱讀學習表現敘寫方式不一，能力遞進關係模糊，顯示課綱縱向銜接仍待細緻化。建議未來研究擴展本研究成果，建構第三和第五學習階段國語文閱讀素養表現水準描述，使學習階段能力描述具備縱貫脈絡，提供教學現場一套完整、連貫的學生能力發展參考架構。

（二）擴充「評價與省思」之數位閱讀素養表現水準描述

隨著數位科技融入學習情境，數位閱讀已成為國語文閱讀素養不可忽視的重要面向。然而，本研究建構之數位閱讀素養表現水準描述，「評價與省思」以表現水準M4為主，M1至M3之表現描述尚不完整，顯示目前評價與省思之數位閱讀素養評量試題多集中於高能力學生，低至中能力學生之數位閱讀表現樣態尚待探索。建議未來研究進一步研發涵蓋不同難度層級的數位閱讀試題，補充評價與省思M1至M3的數位閱讀表現描述，確立完整的數位閱讀素養能力階層，使表現水準描述更臻完備，以符合學生閱讀行為與科技發展趨勢。

（三）探討表現水準描述於教學現場的實踐

表現水準描述不僅能詮釋學生能力水準，亦能作為教師進行差異化教學與學習診斷之工具，促進評量與學習的融合。然而，表現水準描述如何轉化為課堂教學實務，發揮其診斷與引導學習之功能，尚待實證研究支持。建議後續研究以表現水準描述為基礎，針對不同能力層級學生設計教學方案，探討教師應用表現水準描述於課堂教學、學生學習診斷及學習成效提升的可行性，發揮表現水準描述作為「學習地圖」在實務現場的應用潛力。

致謝：本研究為國家教育研究院「中小學生成就長期追蹤評量」整合型計畫之子計畫六「第四學習階段中文閱讀素養長期追蹤」（NAER-2019-041-A-1-1-E1-06）部分成果。研究歷程中，謹此向整合型總計畫主持人任宗浩副研究員致以誠摯謝意，感謝為計畫總體推動提出前瞻性發想與協助跨子計畫資源之協調。本研究亦蒙子計畫二陳冠銘助理研究員提供數據支援、子計畫三吳正新副研究員負責整合型計畫之抽樣設計，為本研究奠定重要基礎，謹此併謝。

此外，特此感謝本子計畫共同主持人蔡曉楓副研究員參與研究討論。感謝陳毓欣博士後研究員於表現水準描述設計與規劃階段提供專業建議，對本研究之概念發展貢獻良多。莊淳斌命題專員協助本研究相關評量工具之研發，其貢獻對本研究至關重要。林佳曄專任助理執掌研究庶務，讓研究工作如期推進。

最後，誠摯感謝為本研究提供專業指導的教授、學者，有您們的提攜、指點，研究才能順利完成。

參考文獻

- 任宗浩（2018）。素養導向評量的界定與實踐。載於蔡清華（主編），*課程協作與實踐第二輯*（頁75-82）。教育部。
- [Jen, T.-H. (2018). The definition and practice of competency-based assessment. In C.-H. Tsai (Ed.), *Curriculum collaboration and practice II* (pp. 75-82). Ministry of Education.]
- 行政院（2018）。推動循環經濟——創造經濟與環保雙贏。<https://www.ey.gov.tw/Page/5A8A0CB5B41DA11E/12c0a2b8-485d-49d7-ba9e-a9a10b82828e>
- [Executive Yuan. (2018). *Promoting a circular economy: Creating a win-win situation for the economy and the environment*. <https://www.ey.gov.tw/Page/5A8A0CB5B41DA11E/12c0a2b8-485d-49d7-ba9e-a9a10b82828e>]
- 吳正新（2022）。長期追蹤調查抽樣技術與權重校正（NAER-2019-113-A-1-1-E1-03）。國家教育研究院。
- [Wu, J.-S. (2022). *Sampling techniques and weight adjustment in longitudinal surveys* (NAER-2019-113-A-1-1-E1-03). National Academy for Educational Research.]

財團法人資源循環臺灣基金會（2019）。邁向循環臺灣：循環經濟實踐案例手冊。

[Circular Taiwan Network. (2019). *Toward a circular Taiwan: Handbook of circular economy case practices.*]

國立臺灣師範大學心理與教育測驗研究發展中心（2024）。113年國中教育會考各科等級加標示人數百分比統計表。 <https://cap.rcpet.edu.tw/exam/113/113Table2.pdf>

[Research Center for Psychological and Educational Testing, National Taiwan Normal University. (2024). *Percentage statistics of students reaching each proficiency level in each subject in the 2024 comprehensive assessment program for junior high school students.* <https://cap.rcpet.edu.tw/exam/113/113Table2.pdf>]

國立臺灣師範大學心理與教育測驗研究發展中心（2025）。114年國中教育會考各科等級加標示人數百分比統計表。 <https://cap.rcpet.edu.tw/exam/114/114%E5%B9%B4%E5%9C%8B%E4%B8%AD%E6%95%99%E8%82%B2%E6%9C%83%E8%80%83%E5%90%84%E7%A7%91%E8%A8%88%E5%88%86%E8%88%87%E9%96%B1%E5%8D%B7%E7%B5%90%E6%9E%9C%E8%AA%AA%E6%98%8E.pdf>

[Research Center for Psychological and Educational Testing, National Taiwan Normal University. (2025). *Percentage statistics of students reaching each proficiency level in each subject in the 2025 comprehensive assessment program for junior high school students.* <https://cap.rcpet.edu.tw/exam/114/114%E5%B9%B4%E5%9C%8B%E4%B8%AD%E6%95%99%E8%82%B2%E6%9C%83%E8%80%83%E5%90%84%E7%A7%91%E8%A8%88%E5%88%86%E8%88%87%E9%96%B1%E5%8D%B7%E7%B5%90%E6%9E%9C%E8%AA%AA%E6%98%8E.pdf>]

國立臺灣師範大學心理與教育測驗研究發展中心（無日期）。各科等級描述。2025年5月4日，取自<https://cap.rcpet.edu.tw/score1.html>

[Research Center for Psychological and Educational Testing, National Taiwan Normal University. (n.d.). *Proficiency level descriptions by subject.* Retrieved May 4, 2025, from <https://cap.rcpet.edu.tw/score1.html>]

國家教育研究院（2015）。十二年國教語文領域一國語文課程綱要研修第7次全體委員會會議紀錄。 [https://www.naer.edu.tw/upload/1/9/doc/1051/1040614%E5%8D%81%E4%BA%8C%E5%B9%B4%E5%9C%8B%E6%95%99%E8%AA%9E%E6%96%87%E9%A0%98%E5%9F%9F\(%E5%9C%8B%E8%AA%9E%E6%96%87\)%E8%AA%B2%E7%A8%8B%E7%B6%B1%E8%A6%81%E7%A0%94%E4%BF%AE%E7%AC%AC7%E6%AC%A1%E5%85%A8%E9%AB%94%E5%A7%94%E5%93%A1%E5%A4%A7%E6](https://www.naer.edu.tw/upload/1/9/doc/1051/1040614%E5%8D%81%E4%BA%8C%E5%B9%B4%E5%9C%8B%E6%95%99%E8%AA%9E%E6%96%87%E9%A0%98%E5%9F%9F(%E5%9C%8B%E8%AA%9E%E6%96%87)%E8%AA%B2%E7%A8%8B%E7%B6%B1%E8%A6%81%E7%A0%94%E4%BF%AE%E7%AC%AC7%E6%AC%A1%E5%85%A8%E9%AB%94%E5%A7%94%E5%93%A1%E5%A4%A7%E6)

%9C%83%E6%9C%83%E8%AD%B0%E7%B4%80%E9%8C%84.pdf

[National Academy for Educational Research. (2015). *Meeting minutes of the 7th plenary session on the curriculum guidelines for the 12-year basic education elementary school, junior high school, and upper secondary school: The domain of language arts—Mandarin*. [https://www.naer.edu.tw/upload/1/9/doc/1051/1040614%E5%8D%81%E4%BA%8C%E5%B9%B4%E5%9C%8B%E6%95%99%E8%AA%9E%E6%96%87%E9%A0%98%E5%9F%9F\(%E5%9C%8B%E8%AA%9E%E6%96%87\)%E8%AA%B2%E7%A8%8B%E7%B6%B1%E8%A6%81%E7%A0%94%E4%BF%AE%E7%AC%AC7%E6%AC%A1%E5%85%A8%E9%AB%94%E5%A7%94%E5%93%A1%E5%A4%A7%E6%9C%83%E6%9C%83%E8%AD%B0%E7%B4%80%E9%8C%84.pdf](https://www.naer.edu.tw/upload/1/9/doc/1051/1040614%E5%8D%81%E4%BA%8C%E5%B9%B4%E5%9C%8B%E6%95%99%E8%AA%9E%E6%96%87%E9%A0%98%E5%9F%9F(%E5%9C%8B%E8%AA%9E%E6%96%87)%E8%AA%B2%E7%A8%8B%E7%B6%B1%E8%A6%81%E7%A0%94%E4%BF%AE%E7%AC%AC7%E6%AC%A1%E5%85%A8%E9%AB%94%E5%A7%94%E5%93%A1%E5%A4%A7%E6%9C%83%E6%9C%83%E8%AD%B0%E7%B4%80%E9%8C%84.pdf)]

教育部（2014）。十二年國民基本教育課程綱要總綱。

[Ministry of Education. (2014). *Curriculum guidelines of 12-year basic education: General guidelines*.]

教育部（2018）。十二年國民基本教育課程綱要國民中小學暨普通型高級中等學校：語文領域—國語文。

[Ministry of Education. (2018). *Curriculum guidelines for the 12-year basic education: Elementary school, junior high school, and upper secondary school—The domain of language arts—Mandarin*.]

教育部（2020）。十二年國民基本教育課程綱要國民中小學暨普通型高級中等學校：議題融入說明手冊。

[Ministry of Education. (2020). *Manual for integrating curriculum issues into the 12-year basic education: Elementary school, junior high school, and upper secondary school*.]

教育部（2021）。十二年國民基本教育課程綱要總綱修正。

[Ministry of Education. (2021). *Revised curriculum guidelines of 12-year basic education: General guidelines*.]

教育部統計處（2020）。108（2019-2020）學年度高級中等學校科別資料。https://depart.moe.edu.tw/ed4500/News_Content.aspx?n=5A930C32CC6C3818&sms=91B3AAE8C6388B96&s=596D9D77281BE257

[Department of Statistics, Ministry of Education. (2020). *Data on high school departments for the 2019-2020 academic year*. https://depart.moe.edu.tw/ed4500/News_Content.aspx?n=5A930C32CC6C3818&sms=91B3AAE8C6388B96&s=596D9D77281BE257]

陳冠銘（2022）。長期追蹤調查似真值估算與量尺等化研究（NAER-2019-113-A-1-

1-E1-02)。國家教育研究院。

[Chen, K.-M. (2022). *Plausible value estimation and scale equating in longitudinal surveys* (NAER-2019-113-A-1-1-E1-02). National Academy for Educational Research.]

曾建銘、王暄博（2012）。標準設定之效度評估：以TASA國語科為例。《教育學刊》，39，77-118。

[Cheng, C.-M., & Wang, H.-P. (2012). Assessing the standards set by TASA and its standard-setting procedures. *Educational Review*, 39, 77-118.]

鄒倫、張祖恩（主編）（2018）。循環經濟系列叢書第三冊：資源及產品循環應用技術。財團法人中技社。

[Tsou, L., & Chang, T.-E. (Eds.). (2018). *Circular economy series volume 3: Technologies for resource and product recycling*. CTCI Foundation.]

劉文海（2007）。中國大陸再生鋁市場現況與展望。https://www.mirdc.org.tw/FileDownload/IndustryNews/2007431016121252.doc

[Liu, W.-H. (2007). *Current status and prospects of the recycled aluminum market in Mainland China*. https://www.mirdc.org.tw/FileDownload/IndustryNews/2007431016121252.doc]

環境部（2018）。107至109年資源回收再利用推動計畫。

[Ministry of Environment. (2018). *Resource recycling and reuse promotion plan (2018-2020)*.]

謝佩蓉（2018）。108課綱第四學習階段國語文閱讀素養線上評量之建構。《教育科學研究期刊》，63（4），193-228。https://doi.org/10.6209/JORIES.201812_63(4).0007

[Hsieh, P.-J. (2018). Development and validation of an online assessment of Mandarin reading literacy for middle school students of the 12-year basic education program. *Journal of Research in Education Sciences*, 63(4), 193-228. https://doi.org/10.6209/JORIES.201812_63(4).0007]

謝佩蓉、林明佳（2022）。國小中文閱讀素養長期追蹤（NAER-2019-077-A-1-1-E1-13）。國家教育研究院。

[Hsieh, P.-J., & Lin, M.-C. (2022). *A longitudinal study of elementary students' Mandarin reading literacy* (NAER-2019-077-A-1-1-E1-13). National Academy for Educational Research.]

Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31(2-3), 162-172. https://doi.org/10.1016/j.stueduc.2005.05.008

Alderson, J. C. (2007). The CEFR and the need for more research. *Modern Language Journal*, 91(4), 659-663. https://doi.org/10.1111/j.1540-4781.2007.00627_4.x

Beaton, A. E., & Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of*

- Educational Statistics*, 17(2), 191-204. <https://doi.org/10.2307/1165169>
- Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.
- Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, 67(3), 343-368. <https://doi.org/10.1080/00131911.2014.929565>
- Chall, J. S. (1983). *Stages of reading development*. McGraw-Hill.
- Chall, J. S. (1996). *Stages of reading development* (2nd ed.). Harcourt Brace College Publishers.
- Chen, K.-M., & Jen, T.-H. (2012, March 5-7). *Positional effect of item blocks in international large-scale assessment caused the item difficulty different between Taiwan and the USA* [Paper presentation]. The 6th International Technology, Education and Development Conference, Valencia, Spain.
- Cizek, G. J. (2012). An introduction to contemporary standard setting: Concepts, characteristics, and contexts. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 3-14). Taylor & Francis.
- Common Core State Standards Initiative. (2009). *Common core state standards initiative standards-setting criteria*. <http://www.corestandards.org/about-the-standards/development-process/>
- Common Core State Standards Initiative. (2022). *Key shifts in English language arts*. <http://www.corestandards.org/other-resources/key-shifts-in-english-language-arts/>
- Council of Europe. (2001). *Common European framework of reference for languages*. Cambridge University Press.
- Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment— Companion volume*. Council of Europe Publishing.
- Dijk, T. A. v., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.
- DJEFAFLIA, K. (2018). *Developing critical thinking skill through reading: Both teachers and first year master students' perspectives*. Guelma University.
- Gonzalez, E. J., Joseph Galia, A. A., Erberber, E., & Diaconu, D. (2004). Reporting student achievement in mathematics and science. In M. O. Martin, I. V. S. Mullis, & S. J. Chrostowski (Eds.), *TIMSS 2003 technical report* (pp. 275-307). TIMSS & PIRLS International Study Center.
- Gouëdard, P., Pont, B., Hyttinen, S., & Huang, P. (2020). *Curriculum reform: A literature review to support effective implementation* (OECD Education Working Paper No. 239). OECD. <https://doi.org/10.1787/efe8a48c-en>

- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
- James, Z., & Rossiter, J. (2018). *Using scale-anchoring to interpret the young lives 2016-17 achievement scale*. Young Lives.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychotocical Review*, 95(2), 163-182. [https://doi.org/10.1016/S0166-4115\(08\)61551-4](https://doi.org/10.1016/S0166-4115(08)61551-4)
- Kintsch, W. (2019). Revisiting the construction-integration model of text comprehension and its implications for instruction. In D. E. Alvermann, N. J. Unrau, M. Sailors, & R. B. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th ed., pp. 178-203). Routledge.
- Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209-226). Blackwell. <https://doi.org/10.1002/9780470757642.ch12>
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174. <https://doi.org/10.1007/BF02296272>
- New Zealand Ministry of Education. (2015). *The New Zealand curriculum*. Crown.
- New Zealand Ministry of Education. (2025a). *The learning progression frameworks*. <https://curriculumprogresstools.education.govt.nz/lpf-tool/>
- New Zealand Ministry of Education. (2025b). *Understanding the reading framework*. <https://curriculumprogresstools.education.govt.nz/lpfs/understanding-the-reading-framework/>
- Organisation for Economic Co-operation and Development. (2017). Proficiency scale construction. In OECD (Ed.), *PISA 2015 technical report* (pp. 275-287).
- Organisation for Economic Co-operation and Development. (2019). *PISA 2018 assessment and analytical framework*.
- Organisation for Economic Co-operation and Development. (2020). *PISA 2018 technical report*.
- Organisation for Economic Co-operation and Development. (2021). *21st-century readers: Developing literacy skills in a digital world*.
- Olsen, R. V., & Nilsen, T. (2017). Standard setting in PISA and TIMSS and how these procedures can be used nationally. In S. Blömeke & J.-E. Gustafsson (Eds.), *Standard setting in education: The Nordic countries in an international perspective* (pp. 69-84). Springer.
- Perie, M. (2008). A guide to understanding and developing performance-level descriptors.

- Educational Measurement: Issues and Practice*, 27(4), 15-29. <https://doi.org/10.1111/j.1745-3992.2008.00135.x>
- Phillips, G. W. (2012). The benchmark method of standard setting. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd ed., pp. 323-345). Routledge.
- Schneider, M. C., Huff, K. L., Egan, K. L., Gaines, M. L., & Ferrara, S. (2013). Relationships among item cognitive complexity, contextual demands, and item difficulty: Implications for achievement-level descriptors. *Educational Assessment*, 18(2), 99-121. <https://doi.org/10.1080/10627197.2013.789296>
- Shalihin, R. R. (2024). *Higher-order versus lower-order thinking skills: How much does the hierarchy matter?* <https://www.bera.ac.uk/blog/higher-order-versus-lower-order-thinking-skills-how-much-does-the-hierarchy-matter>
- Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, 42(3), 214-240. <https://doi.org/10.1080/02702711.2021.1888348>
- Stanley, T. (2019). *Using rubrics for performance-based assessment: A practical guide to evaluating student work*. Routledge.
- Stevens, D. D., & Levi, A. J. (2013). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning* (2nd ed.). Stylus.
- U.S. Department of Education. (2025). *Distribution of reading questions: 2019 and 2024*. <https://nces.ed.gov/nationsreportcard/reading/distributequest.aspx>
- von Davier, M., Kennedy, A., Reynolds, K., Fishbein, B., Khorramdel, L., Aldrich, C., Bookbinder, A., Bezirhan, U., & Yin, L. (2024). *TIMSS 2023 international results in mathematics and science*. Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.timss.rs6460>
- Wagner, J.-P., & Hastedt, D. (2022). Valuing curriculum-based international large-scale assessments: Ensuring alignment with national curricula in IEA studies. *IEA Compass: Briefs in Education*, 16, 1-7.
- Yan, Z., & Yang, L. (2022). Assessment-as-learning in the global assessment reforms. In Z. Yan & L. Yang (Eds.), *Assessment as learning: Maximising opportunities for student learning and achievement* (pp. 1-7). Routledge.

附錄 第四學習階段閱讀表現水準描述範例題

一、妮妮與橘橘的對話

因應地球資源的限制，歐盟於2012年簽訂「循環經濟宣言」、2015年提出「循環經濟推動計畫」，企盼達成零廢棄的目標，更於2020年推出「新循環經濟行動方案」，期望透過具體的行動方案落實循環經濟的政策。本題組以國中學生用餐後的情境，設計妮妮和橘橘兩人一問一答，從中帶出循環經濟的概念與實例。基於評量即學習的理念，透過提問讓學生獲取循環經濟的重要概念，並延伸情境請學生上網查詢相關資料。這個題組除了評量學生的一般閱讀能力與數位閱讀能力，亦能藉由查詢結果解決真實生活中的問題。

[9]問題1. Apple公司2018年最新的拆解機器人黛西，她可以拆解手機零件，以創新的方式處理物料。Apple公司革新回收的歷程，與文章中哪一種廢棄物重新被利用的歷程相似？

- ①廚餘養豬
- ②豬糞發電
- ③咖啡渣製成紗衣
- ④廢輪胎變潛水衣

計分

題型：選擇題

滿分

代碼1：正確選出④得1分

零分

代碼0：選擇其他選項，如①或②或③

說明

本題評量標的為「一般閱讀素養」之理解，檢視學生「推論文本中隱含的訊息」的能力。利用題幹給定的新情境，評量學生是否能統整理解文本中再利用

的歷程，進行比較、類推。「拆解電子垃圾並將原料再利用」已行之有年，即使學生不知道拆解過程中，產品元件可能對人體有害，然題幹已明述Apple公司研發拆解機器人黛西革新歷程來解決這樣的問題。另一方面，對話倒數第四段，橘橘提到廢輪胎製成碳黑並非新技術，然應用於潛水衣卻是新發想。兩者同樣是在舊有的基礎之上創新，學生須能推論字裡行間沒有寫出的訊息。財團法人資源循環臺灣基金會（2019）指出，工業循環（technical cycle）包括材料選擇、產品設計、製程優化、副產品資源化、維修／再使用、翻新／再製造、回收及處理再生八方向，提出「循環再生概念與輪胎業等合作回收廢輪胎，製成環保碳黑，再運用到紡織業，形成跨域合作關係」真實案例。

[10]問題2. 請依循環經濟的分類，判斷下列案例屬於循環經濟的哪種類型？

	生物循環	工業循環
1. 將手機舊電池組合成行動電源	☐	☐
2. 將合金做成不鏽鋼環保餐具	☐	☐
3. 利用甘蔗渣製成吸管	☐	☐
4. 提煉再生鋁合金，製成再生鋁合金汽車外殼	☐	☐
5. 把撞壞、快要過期的蔬果，烹調成料理的剩食餐廳	☐	☐

計分

題型：多重是非題

滿分

代碼1：正確將5項案例分類：

項目1點選工業循環；項目2點選工業循環；項目3點選生物循環；項目4點選工業循環；項目5點選生物循環。

零分

代碼0：任何一項案例點選錯誤或漏答。

說明

本題評量標的為「一般閱讀素養」之理解，檢視學生「推論文本中隱含的訊息」的能力。利用題幹給定的五種新情境，評量學生是否能統整理解文本中生物

循環和工業循環的概念，兩相進行比較、類推。於對話倒數第二段，橘橘說「你看就像這張圖」，下方圖示呈現循環經濟兩大分類：生物循環和工業循環，將資源的效益最大化。學生從圖示理解兩大分類的內涵後，再分別理解五項案例的內涵，對應例子和圖示兩者間的主要概念。在題型設計方面，若只提問一個案例學生就答對，較難確認答題概念是否紮實或只是猜對，故透過五個案例形成多重是非題，確認學生對於概念的理解無誤。

財團法人資源循環臺灣基金會（2019）指出，生物循環（biological cycle）包括多元生產、製程優化、全利用、運銷、惜食及資源化等六個方向，提出「蔗渣作為養菇替代介質，避免原始林濫伐」和「剩食餐廳」等真實案例。題幹中，「甘蔗渣製成吸管」雖然因其添加物和可回收性引發循環利用率（cyclical use rate）方面的爭議，概念上仍符合「回系統中作為養分，開啟下次循環的契機」。

工業循環多指人造之化合物、合金等生物不可分解的物質產品，延長壽命、循環利用等應用模式（鄒倫、張祖恩，2018），並於生產面強調產品設計的耐久循環性，於消費面強調減少拋棄式用品（環境部，2018）。相關實例分述如下：

1. 南韓「Better Re行動電源盒」讓產品在進入回收程序之前盡量增長生命週期，使舊手機電池成為行動電源、賦予新生命，延長電池的使用時間、降低電池「被廢棄」的可能性。
2. 十多年前，大陸地區已從廢鋁原料和鋁渣中回收再生鋁，將75%左右的再生鋁合金用於汽車工業（劉文海，2007）。
3. 「減少使用一次性用品」為我國推動循環經濟的重要政策（行政院，2018），長期重複使用不鏽鋼環保餐具不僅能減少一次性餐具，也能減少地球資源的消耗。

[3]問題3. 下列何者最符合循環經濟的理念？

- ①搖籃到搖籃
- ②搖籃到墳墓
- ③墳墓到墳墓
- ④墳墓到搖籃

計分

題型：選擇題

滿分

代碼1：正確選出①得1分

零分

代碼0：選擇其他選項，如②或③或④

說明

本題評量標的為「一般閱讀素養」之理解，檢視學生「推論全文主旨」的能力。「循環經濟」對第四學習階段的學生而言是較不熟悉的主題，它的宗旨是從搖籃到搖籃，評量學生是否能從大範圍的文本對話中推論出主旨。對話第六段橘橘解釋廚餘的處理「形成永續的循環」，這個題組的問題1帶著學生認識循環經濟的最終目標為「零廢棄」，問題3再帶著學生閱讀循環經濟概念圖，藉這三項搭起鷹架，便於學生推論全文主要說明資源循環零廢棄的永續發展。本題難度甚高，推測可能原因有三：首先，循環經濟並非學生熟悉的主題。第二，以對話方式呈現，對於學生而言並非熟悉的說明文寫作形式。另一方面，學生須運用語文能力，掌握選項中「搖籃」和「墳墓」乃譬喻「自然資源」和「廢棄」，而循環經濟理念在於揚棄過去從搖籃到墳墓的線性資源運用，追求搖籃到搖籃的生生不息。

[12]問題4. 妮妮和橘橘對話之間的關聯性為何？

- ①闡述不同的觀點，提出一致的見解
- ②站在相同的立場，達到誇張的效果
- ③利用二元的角度，說明正反的意見
- ④針對單一的議題，達成相同的共識

計分

題型：選擇題

滿分

代碼1：正確選出④得1分

零分

代碼0：選擇其他選項，如①或②或③

說明

本題評量標的為「一般閱讀素養」之評價與省思，檢視學生「針對文章內容或形式提出個人見解」的能力。評量學生是否能辨識作者採用的寫作結構。對話中，妮妮發問、橘橘回答，妮妮並沒有特定立場，兩人討論單一「循環經濟」議題，最終達成一致看法。

[4]問題5. 妮妮發現親戚開設的養豬場符合行政院農委會養豬場沼氣發電獎勵及補助法規的條件，她上網查詢到底能夠得到多少補助及獎勵。

請你點農業主管法規查詢系統，並在右上角紅圈處的「整合查詢」輸入關鍵字（如下圖所示），找出妮妮親戚的養豬場最多可以得多少金額的補助？



計分

題型：建構反應題

滿分

代碼	評分標準與品質定義	作答實例
1	正確回答385萬（元）	• 385萬 • 385萬元 • 3850000元

零分

代碼	評分規準與品質定義	作答實例
0A	擷取到最高補助金額 600萬元，沒有注意到 妮妮家養多少豬	• 600萬元
0B	回答265萬元	• 265萬元 • 共可得265萬元
0D	其他錯誤類型	• 385 • 1000 • 6000 • 170萬元到600萬元
0E	與題目無關的作答	• 不知道
0X	未作答、空白	(作答區空白)

說明

受限於施測平臺的功能，將學生於數位情境中的閱讀歷程分解為「使用關鍵字檢索特定資料」和「整合文本中的文意訊息」兩題，使得問題5和問題6為系列問題。

本題評量標的為「數位閱讀素養」之理解，檢視學生於「檢索所得網頁內容」和「文本對話」之間，「跨頁整合文意訊息」的能力。評量學生是否能整合文本和外部網頁的訊息。拆解學生的解題歷程如下：使用適當關鍵字查出《行政院農業委員會養豬場沼氣發電獎勵及補助作業要點》第4條所臚列獎勵、補助總額上限，單位為「萬元」，再加上對話第五段，妮妮說「我有個親戚，開設的養豬場有1萬2千頭豬」，兩相整合便能得知最多補助金額為385萬元。

[8]問題6. 你在「農業主管法規查詢系統」輸入什麼關鍵詞，才能找到妮妮親戚的養豬場能夠得到多少補助及獎勵？請將你所輸入的關鍵詞，填寫在下列的方框中。



計分

題型：建構反應題

滿分

代碼	評分規準與品質定義	作答實例
1	在「農業主管法規查詢系統」輸入之關鍵詞，能於五頁的搜尋結果中，查到補助金額。以「關鍵詞是否能於網站中查出能獲得的補助」為給分依據。 為降低查詢偏誤，閱卷者須以「無痕模式」逐筆檢查學生所採用的關鍵詞是否能於該網站中查到題幹的要求。	
1A	關鍵字為「循環經濟」	• 循環經濟
1B	關鍵字包含「沼氣」或「沼氣發電」	• 沼氣 • 沼氣發電 • 沼氣發電獎勵 • 沼氣發電獎勵及補助 • 沼氣發電相關設備 • 養豬場沼氣發電 • 養豬場沼氣發電獎勵 • 豬場沼氣發電 • 行政院農業委員會養豬場沼氣發電獎勵及補助作業要點
1C	關鍵字僅提到「養豬」或「養豬場」	• 養豬 • 養豬場 • 養豬場符合行政院農委會養豬場沼氣發電獎勵 • 我輸入養豬場在整合查詢然後點進去就有文章可以讀找答案

零分

代碼	評分規準與品質定義	作答實例
0	所使用的關鍵字無法於五頁的搜尋結果中，查到補助金額，或沒有作答。 有些零分的關鍵詞（例如：養豬場沼氣發電補助），和滿分的關鍵詞十分相似（例如：沼氣發電獎勵及補助），然而，真實的情況卻無法於該網站中查到題幹的要求，故無法給分。	
0A	以題目的問句搜尋	• 養豬場獎金最多能拿多少 • 養豬場最多可以得多少金額的補助？
0D	其他錯誤類型	• 經濟循環 • 環境經濟 • 養豬補助 • 養豬場補助 • 沼澤發電 • 補助 • 補助金 • 獎金及獎勵 • 養豬場沼氣發電補助 • 養豬場沼氣發電的獎勵及補助 • 有機農業獎勵及補貼辦法 • 廚餘養豬場轉用飼料補助要點 • 行政院農委會養豬場沼氣發電 • 豬糞利用 • 豬大便補助金 • 我們養豬場 • 豬
0E	與題目無關的作答	• 亂猜的 • 無 • 使用咖啡渣製作以衣服
0X	未作答、空白	（作答區空白）

說明

本題評量標的為「數位閱讀素養」之找出訊息，檢視學生「使用關鍵字檢索特定資料」的能力。評量學生是否能選擇適當的關鍵詞，達成題幹所要求的查詢任務。解題線索有三：問題5題幹指出「行政院農委會養豬場沼氣發電獎勵及補助法規」，文本對話第六段橘橘建議妮妮上網查看「養豬場沼氣發電的獎勵及

補助」(須拆解文句)，或者學生也可以使用全文的上位概念「循環經濟」，都有機會達成此題的查詢要求。由於關鍵詞的使用精準度攸關查詢結果的可近性與可行性，故本題將給分標準限定為於五頁的搜尋結果中，查到補助金額。

二、微物跡證

微物跡證常常是刑事案件的破案關鍵。本題組將文言文故事融入微物跡證的說明之中，藉由學生可能喜愛的「穿越」與「偵探」元素，帶領學生一探古今辦案方式。這個題組除了評量學生的文言文閱讀理解，亦引導學生評析寫作布局，並評論真實生活中破案證據的有效性。提問的安排有其順序性，第1題至第4題先確認學生對於文言文的理解是否正確，最後再進行評價與省思。

[1]問題1. 故事中，爭奪羊皮的人是哪兩個人？

- ①李惠、負鹽者
- ②李惠、負薪者
- ③李惠、州綱紀
- ④負鹽者、負薪者

計分

題型：選擇題

滿分

代碼1：正確選出④得1分

零分

代碼0：選擇其他選項，如①或②或③

說明

本題評量標的為「一般閱讀素養」之找出訊息，檢視學生「直接擷取單一段落的字面訊息」的能力。評量學生是否能於不熟悉的文章中，定位題幹要求的訊息。文章第二段開頭寫到「人有負鹽負薪者，同釋重擔息樹陰，二人將行，爭一羊皮」，能得知爭奪羊皮兩個人是負鹽者和負薪者。即使認知歷程僅從文句中找訊息，然故事為文言文，對學生而言為不熟悉文體，推測學生可能在閱讀過程中沒有找到人物角色，或在閱讀過程忽略「惠」指的是判案者，致使難度跳升

至M4表現水準。

[11]問題2. 故事中，「負鹽負薪者」的「負」字意涵，與下列何者相同？

- ①丸尾這次又不「負」眾望，高票當選班長
- ②今晚就是洋基隊和紅襪隊一決勝「負」的時刻
- ③對於昨晚的事他感到很自責，特別登門「負」荊請罪
- ④上次他受鄰人幫忙，現在卻忘恩「負」義，讓人十分不齒

計分

題型：選擇題

滿分

代碼1：正確選出③得1分

零分

代碼0：選擇其他選項，如①或②或④

說明

本題評量標的為「一般閱讀素養」之理解，檢視學生「在語境中解釋字詞意涵」的能力。評量學生是否能透過理解文言文的語句進而判斷詞義。以「詞素」解析「負」字，其意義仍為現今所慣用，因而對大部分的學生而言，本題難度適中。

《教育部重編國語辭典修訂本》中，「負」作為動詞有八種意涵，作為名詞有兩種意涵，作為形容詞有一種意涵，是一個在不同語境之中字義變化、適合評量的多義字。不「負」眾望作為動詞，意思為背棄、違背。一決勝「負」作為名詞，意思為失敗。「負」荊請罪作為動詞，意思為以肩背物，和「負」鹽「負」薪者意義相同。忘恩「負」義作為動詞，意思為背棄、違背。

《教育部重編國語辭典修訂本》中，「詞素」的定義為：構成詞的成分最小，且在意義上不能再分析的單位。例如「日」、「月」、「葡萄」、「蟋蟀」等只包含一個詞素；「電視」、「汽車」、「環保」等則各包含兩個詞素。

[6]問題3. 根據故事，「以杖擊之」指的是杖打何者？

- ①羊皮
- ②負鹽者

③負薪者

④州綱紀

計分

題型：選擇題

滿分

代碼1：正確選出①得1分

零分

代碼0：選擇其他選項，如②或③或④

說明

本題評量標的為「一般閱讀素養」之理解，檢視學生「在語境中詮釋句子意涵」的能力。評量學生是否能根據前後文判斷文言文的詞義。文章第二段寫到「惠令人置羊皮席上，以杖擊之，見少鹽屑」，能得知棒子敲打的是羊皮，「之」字指稱的對象就是羊皮。

[14]問題4. 根據故事，破案的關鍵是什麼？

①被害人的指認

②犯罪者的自首

③發現有效的證據

④訊問裡透露破綻

計分

題型：選擇題

滿分

代碼1：正確選出③得1分

零分

代碼0：選擇其他選項，如①或②或④

說明

本題評量標的為「一般閱讀素養」之理解，檢視學生「整合文本中的文意訊息」的能力。評量學生是否能在段落內辨識出主要概念。故事中，李惠以鹽屑作為辦案證據，讓背柴的人俯首認罪。此外，文章的標題、文章第一段、第三段均

提到「證據」的概念，若學生因而推論全文主旨，便可能成為作答本題的線索，使得本題為簡單的試題。

[2]問題5. 根據本文，微物跡證的特性和功用，分別是什麼？

作答範例：（1）特性是：_____；（2）功用是：_____

計分

題型：建構反應題

滿分

代碼	評分規準與品質定義	作答實例
2	正確回答：（1）特性為它們細小且不易觀察，或是細小的物證（體）；（2）功用是可以成為破案證據。	• 特性是：很不容易察覺；功用是：可當證據做使用 • （1）特性是：細小不易察覺；（2）功用是：協助破案 • 特性是：易被犯罪者遺留的證物；功用是讓鑑識人員近一步的破案。 • 性小且不易觀察容易忽略；功用是能成為破案突破口

部分分數

代碼	評分規準與品質定義	作答實例
1A	僅正確回答（1）特性為它們細小且不易觀察。沒有提及（2）的功用，或是回答錯誤。	• 特性是細小且不易觀察；功用是採集有用的證物 • 特性是用小東西找出關鍵；功用是可以辨別是非
1B	僅正確回答（2）跡證的功用。例如：可以成為破案證據。沒有提及（1）的特性，或是回答錯誤。	• 特性是羊皮；功用是可以當證據 • 有少鹽屑，有用的證據。 • （1）特性是可偵破案件；（2）功用是可偵破案件 • 特性是裨益觀察；功用是協助破案

零分

代碼	評分規準與品質定義	作答實例
0A	沒有確切掌握微物跡證的「特性」與「功用」，回答的是「證據」本身。無法凸顯微物跡證與其他證據的不同	•（1）特性是：纖維、毛髮、玻璃碎片、油漆碎片、火藥殘跡、血跡或花粉等；（2）功用是：因為它們細小且不易觀察，所以容易被犯罪者忽略而遺留現場。 • 洞察案情 採集到有用的證物 • 1指紋2證明
0D	其他錯誤答案	• 1抓嫌疑犯 2有安全姓 • 特性是羊皮功用是羊皮席上
0E	與答案無關的作答	• 很難我不會
0X	未作答、空白	（作答區空白）

說明

本題評量標的為「一般閱讀素養」之找出訊息，檢視學生「直接擷取單一段落的字面訊息」的能力。評量學生是否能依據題幹要求，自文本中找出特定訊息。文章第三段提到「即便是微小證物也可偵破案件，……，因為它們細小且不易觀察，……，證物，協助破案」。題幹要求學生分別寫出微物跡證的「特性」和「功用」兩項概念，而非寫出同一概念下的兩項，故本題設定為2分題，只答出其中一項概念則得到部分分數。

即使微物跡證本身沒有主動性的功用，需仰賴鑑識人員憑藉邏輯推理和縝密的心思，加上豐富的經驗，才能發揮其力量。然而，無論人類選擇或不選擇使用證據，客觀而言，證據的特性和功用並不會改變。

[13]問題6.「鹽屑」在這則故事裡扮演的作用，相當於現今科學辦案中的哪一種證據？

- ①事發時目擊者的證詞
- ②模擬出的嫌疑犯畫像
- ③犯罪者偵訊時的陳述
- ④被害人指縫中的泥土

計分

題型：選擇題

滿分

代碼1：正確選出④得1分

零分

代碼0：選擇其他選項，如①或②或③

說明

本題評量標的為「一般閱讀素養」之評價與省思，檢視學生「應用源自文本之啟發解決真實情境的問題」的能力。評量學生是否能運用文本中所獲得的啟發，辨別生活中相似的例子。本題的解題關鍵為「辦案證據」的背景知識。「鹽屑」在這則故事裡扮演的作用為微物證據，現今科學辦案中，證詞、畫像、陳述、泥土確實都可能是證據，然只有泥土可能具備微物證據的特性。

[7]問題7.「李惠審判羊皮」的故事，在本文的功用是什麼？

⑤指出問題來源

⑥提供例證資料

⑦製造反諷效果

⑧比較古今異同

計分

題型：選擇題

滿分

代碼1：正確選出②得1分

零分

代碼0：選擇其他選項，如①或③或④

說明

本題評量標的為「一般閱讀素養」之評價與省思，檢視學生「分析寫作手法」的能力。評量學生是否能辨識作者採用的寫作手法所呈現的閱讀效果。本篇文章主旨在於說明微物跡證在案件鑑定中所扮演的重要角色，屬於介紹性說明文，並以文言文故事舉例說明，以增強說明文的例子。

[5]問題8. 由於時代和科技的不斷演進，辦案過程也有所不同。

你認為以現今科學辦案的角度，李惠辦案過程，證據充足嗎？

作答範例：以現今科學辦案的角度，李惠判案採用的證據_____

（充足或不充足），因為_____。

計分

題型：建構反應題

滿分

代碼	評分規準與品質定義	作答實例
1	回答「不充足」，而且理由合乎邏輯得1分。	
1A	不充足，因為無具個性證據，或是證據無法個化（相同概念即可）	• “不充足，無法確認鹽是誰的” • 不充足，因為也有可能帶柴的那個人身上也沾上鹽 • 不充足，因為不一定只有負鹽者身上有鹽 • 不充足，因為不一定就是負鹽者的鹽 • 不充足，沒有決定性的證據
1B	不充足，因為需要更多證據	• 以現今科學辦案的角度，李惠判案採用的證據不充足，因為只看一個證據感覺太少了。 • 不充足，因為證物不足（理由似乎不足） • 不充足，還需要更多證據，說不定有例外

零分

代碼	評分規準與品質定義	作答實例
0	回答「不充足」但理由不合乎邏輯，或是回答「充足」，或沒有作答。	
0A	回答不充足，但理由無關個性證據、證據無法個化、需要更多證據	• 以現今科學辦案的角度，李惠判案採用的證據不充足，因為李惠忽略了纖維、毛髮、玻璃碎片、油漆碎片、火藥殘跡、血跡或花粉。 • 以現今科學判案角度，李惠判案採用的證據不足，因為她活在過去
0B	僅回答不充足，但沒寫理由	• 不充足

（續下頁）

代碼	評分規準與品質定義	作答實例
0C	回答充足	• 充足，因為有高等級的經驗 • 以現今科學辦案的角度，李惠判案採用的證據充足，因為那個證據就是鐵證了 • 以現今科學辦案的角度，李惠判案採用的證據充足，因為犯人自首了 • “1.充足；2.可能是重要的破案關鍵”
0D	其他錯誤答案	• 他偷羊皮
0E	與答案無關的作答	• ？？ • 他爽 他不爽
0X	未作答、空白	（作答區空白）

說明

本題評量標的為「一般閱讀素養」之評價與省思，檢視學生「針對文章內容或形式提出個人見解」的能力。評量學生是否能反思、評論文章內容。現今司法審判重視科學證據的強度，李惠指證的證據僅有鹽屑一項，也無法個化（individualization），亦即無法證實羊皮上的鹽屑，就是揹着鹽的人身上的鹽。故以現今科學辦案的角度，李惠判案採用的證據並不充足。

